DISCUSSION PAPER SERIES

IZA

IZA DP No. 9078

The Occupational Feminization of Wages

John T. Addison
Orgul D. Ozturk
Si Wang

May 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Occupational Feminization of Wages

John T. Addison
Durham University Business School,
University of South Carolina and IZA

Orgul D. Ozturk
University of South Carolina

Si Wang
Hunan University

Discussion Paper No. 9078
May 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9078
May 2015

ABSTRACT

The Occupational Feminization of Wages

This paper updates the major study by Macpherson and Hirsch (1995) of the effect of the gender composition of occupations on female (and male) earnings. Using large representative national samples of employees from the Current Population Survey, cross-sectional estimates of the impact of proportion female in an occupation (or feminization) on wages are first provided, paying close attention to the role of occupational characteristics. Specification differences in the effects of feminization across alternative subsamples are examined as well as the contribution of the feminization argument to the explanation of the gender wage gap. An updated longitudinal analysis using the CPS data is also provided. This examination of two-year panels of individuals is supplemented using information from the 1979 National Longitudinal Survey of Youth which has the advantage of offering a longer panel. Analysis of the former suggests the reduction in gender composition effects observed for females in cross section with the addition of controls for occupational characteristics becomes complete after accounting for unobserved individual heterogeneity. This is not the case for the latter dataset, most likely reflecting heritage effects of discrimination in what is an aging cohort.

JEL Classification: J31, J71

Keywords: occupational segregation, gender wage gap

Corresponding author:

John T. Addison
Department of Economics
Darla Moore School of Business
University of South Carolina
1014 Greene Street
Columbia, SC 29208
USA
E-mail: ecceaddi@moore.sc.edu

I. Introduction

The strong relationship between the gender composition of an occupation and the relative earnings of both females and males seemingly offers a clear rationale for measures geared to improving the lot of lowly paid workers via comparable worth policies and other anti-discrimination policies, in addition to more conventional instruments seeking to stiffen human capital endowments. Unfortunately, the empirical consensus does not extend much beyond agreement on the stylized facts of earnings disparities that are increasing in an occupation's proportion female. That is, there is disputation not only as regards magnitudes but also causation.

At root, the controversy has a basis in a literature that does not control for a number of variables that might reasonably be expected to influence earnings and earnings development. Further, in addition to often scant controls for observables, there is a dearth of studies using longitudinal analysis. In the latter case, the feminization argument may be correlated with unmeasured skill and taste differences among workers and in the former case with controls for occupational attributes that might reasonably be expected to influence earnings and earnings development.

The present paper is motivated by an important study of occupational sex segregation by Macpherson and Hirsch (1995) covering the interval 1973-93 that is notable in three principal respects: first, in its use of several large datasets; second, in its deployment of arguments not typically found in the literature; and, third, in offering a formal longitudinal analysis of wage change. In short, Hirsch and Macpherson investigate whether the material gender composition effects reported in the literature are a chimera – in large part the result of occupational characteristics, quality sorting on gender composition, taste differences, and other factors correlated with the proportion female in an occupation.

In examining CPS data over the interval 1993-2010, this paper updates Macpherson and Hirsch. As do these authors, it first presents cross-sectional estimates of the relation between proportion female in an occupation and wages, paying close attention to the role of occupational skills and job characteristics. Results are provided by year and also for the pooled data set to examine specification differences in the effects of feminization across alternative groups of workers, inter al. A decomposition of the gender wage gap by broad specification and year then assesses the contribution of feminization to the explained and unexplained gaps. The final stage of the analysis controls for unobserved fixed effects in measuring the relation between gender composition and wages. While also using the longitudinal capacity of the CPS for this purpose, since matched worker pairs are potentially available only for adjacent years the CPS panel analysis is supplemented using information from the National Longitudinal Survey of Youth. The goal in each case is of course to determine whether differences in unobserved skills and preferences are correlated with gender composition, and thereby facilitate our understanding of why predominantly female jobs pay lower wages to women and men.

II. Theoretical Considerations

There are two main explanations for the covariation of wages and the gender composition of occupations. One is human capital theory and the other is discrimination resulting in crowding and possibly to the undervaluation of women's work. Human capital theory is based on choice (Becker, 1985). Predominantly male occupations pay more than predominantly female occupations under a human capital interpretation because individuals in in the former have chosen to invest more in human capital. Similarly, by reason of their (historically) weaker labor force attachment, women choose occupations in which their skills will depreciate less rapidly during spells of absence from the labor market (Polachek, 1981, 1985).[1] According to the theory of occupational crowding, however, male jobs pay more because women excluded from them by discrimination are shunted into other occupations with no or lesser discrimination and the resulting increased supply of labor (or crowding) lowers their wages (Bergmann, 1974). The caveat is of course that where women are crowded into particular occupations by reason of their preferences, the negative effect of greater feminization may be a costly compensating differential. It may also be the case that persons employed in female-dominated occupations receive lower returns to occupational characteristics (e.g. specific vocational preparation) because their work – so-called "women's work" – is undervalued (Gerhart and El Cheikh, 1991) even though in principle their incumbents are equally well qualified. There is an extensive literature suggesting that wage inequality is socially constructed and that work in women's occupations is undervalued by reason of institutionalized bias against women (see, for example, Treiman and Hartman, 1981; Kilbourne et al., 1994; Magnusson, 2009) even if the skills required for lower-paid female dominated jobs are comparable to those in better-paid male-dominated jobs; one of

[1] This approach includes notions of firm-specific human capital formation that have been introduced into the occupational sex-segregation literature by Tam (1997) as part of a differential levels-of-specialization approach to pay differences as opposed to occupational sex segregation.

the more transparent aspects of which is the devaluation of caring and nurturing skills associated with females.[2]

Not surprisingly perhaps the standard models are thin on the details of allocation – in short, how individuals progress through a jobs hierarchy. By analogy with the above narrative, this would on the one hand involve consideration of how individuals control those prospects through the acquisition of knowledge and skills. On the other, it would also encompass the institutionalist challenge based on notions of *social technology* (Osterman, 1987), having to do with the manner in which jobs are structured, the selection of individuals into those jobs, and the valuation of jobs. The present treatment will eschew consideration of the promotion process, despite its potential importance in producing female-dominated and male-dominated jobs and thence the application of bureaucratic processes, customs, and notions of fairness that may lead to the systematic undervaluation of women's work (e.g. McArthur, 1985).

As was noted earlier, research by both economists and other social scientists has confirmed that the share of females in an occupation is negatively associated with the wages received by women (and men) in that occupation. Given the competing explanations for this phenomenon, it follows that measurement issues loom large. Much progress can be made using large data sets with detailed occupational controls, including importantly job amenities/disamenities. This may be seen as the central contribution of Macpherson and Hirsch (1995). But there are two remaining issues. One is selection and the other is unobserved individual heterogeneity. If inclusion in the wage regression sample 'favors' those with higher wage offers, the selection bias will be positive; that is, the mean of actual wages will be higher than the mean of wage offers. If, on the other hand, inclusion in the sample is selective of those with lower values of time in alternative uses (e.g. nonmarket activity or self-employment), then the bias will be negative; that is, the mean of actual wages will be less than the mean of wage offers. And if two groups – men and women – vary in the direction or magnitude of this selectivity bias, estimates of differentials between them based on observed wages will be biased. More importantly, the feminization effect might reflect unobserved productivity differences – in abilities, training, and occupational characteristics – *and other differences such as tastes* that may be expected to lessen the effect of occupational composition on earnings. Both biases tend to have been neglected in the literature for data reasons. However, as was also noted earlier, in its

[2] A separate although related theme is provided by socialization theories; see, for example, Clausen (1968).

supplemental analysis the present treatment seeks to control for unmeasured individual labor quality/taste differences and to offer a more thorough analysis than Macpherson and Hirsch (1995) in this regard. That said, although quality sorting on gender composition is taken into account in what follows, the focus on observed wages (rather than wage offers) means that correction for standard potential selection biases (into employment) will not be examined.

III. A Brief Literature Review

The large plurality of feminization studies focus on wages.[3] The focus here is also wage studies, and in particular those investigating the impact of occupational feminization on *individual* earnings.[4] In an early study using Current Population Survey (CPS) data that controls for selection into employment on the part of males and females, Blau and Beller (1988) examine earnings differentials by gender for 1971 and 1981. (The selection coefficients are negative (positive) for men (women), implying that nonparticipants had higher (lower) wage offers than those in employment.). The authors find that the female-male earnings differential increased over time, and seek to explain the trends. Abstracting from the influence of time inputs, both selection and gender composition emerge as key to this improvement. For its part, selection explains a large part of the improvement for (white) females since the increase in the selectivity of the wage regression was greater for men than for women. Women also earned modestly more than men with similar characteristics in 1981 than 1971 (selectivity-adjusted estimates). Blau and Beller deploy two inverse measures of feminization, namely situations in which the male share of an occupation is greater than 70 percent, termed 'male occupations' and those where it is between 41 and 69 percent, termed 'integrated occupations.' The coefficients for both are positive and well determined for males and females in 1981 and 1971, and are increasing over time. As a matter of fact, the percent of females in these two categories increased materially over time. However, the total effect of this increased penetration was to widen the differential. This was because the coefficient increased more for males than it did for females in both male and integrated occupations (see also Lewis, 1996), so that the increased entry of women was insufficient to turn the tide. The authors speculate that the increase in the returns to being in a

[3] For promotions within occupations and an analysis of the assignment of job points to occupations, see Paulin and Mellor (1996) and Schumann et al. (1994), respectively.

[4] A useful summary of seven early studies using either unweighted or weighted occupations as the unit of analysis is provided by Sorensen (1990: Table 1). See also Perales (2010: Table 1).

male occupation (for both genders) may have reflected an increase in crowding in the female sector while the greater increase in men's than women's earnings in male occupations may have occurred disproportionately in entry level positions. Overall, however, factors serving to widen the overall male-female differential were dominated by others serving to narrow it.

Sorensen (1990) offers a test of the crowding hypothesis using data from the 1984 Panel Study of Income Dynamics (PSID) and the May/June 1983 CPS. Three different earnings models are estimated. The first is a standard human capital model augmented by the gender composition of the worker's occupation (the proportion of women in that occupation). The second adds a wider array of explanatory variables (including union status), and the third adds detailed industry dummies. The coefficient estimates for the gender composition variable decline with each augmentation but remain statistically significant throughout. For the full model, female earnings are reduced by 23 percent in the PSID and by 15 percent in the CPS if they are employed in an exclusively female occupation rather than an exclusively male one. The corresponding values for male earnings are decreases of 24 percent and 25 percent, respectively. And, again for the full model, the proportion of the earnings gap explained by feminization is 23 percent for the PSID and 20 percent for the CPS. These are sizable estimates (cf. Johnson and Solon, 1996), although as the author cautions the variable might overstate the impact of crowding where it is correlated with unobserved productivity characteristics

The three remaining studies considered here return to the issue of biases in estimating the effect of feminization on earnings. One approach to the problem is that adopted by Groshen (1991), who first attempts to separate out the effect of segregation by occupation from that associated with firm and job cell (an individual's job cell is defined as all workers in the same job classification at the same establishment). Using cross section data for five industries from the BLS Industry Wage Surveys, 1974-78, and a regression of the log wage on proportion female in the occupation, proportion female in the establishment, and proportion female in a job-cell, together with an individual female dummy, Groshen estimates that the largest contribution (obtained by multiplying the coefficient on proportion female by the gap between the proportions of female and male employment in the occupation, establishment, and job cell) stems from occupation because occupations are highly segregated and their wages are strongly linked to proportion female. Even in integrated occupations, people work primarily with members of their own sex, and this segregation tends to raise men's wages and lower female wages. Focusing on occupation, Groshen seeks finally to determine which of the two main theories – human capital or discrimination – is most plausible by adding measures of union status, region, general education, vocational training, strength, as well as physical demands and quality of environment for each occupation to the wage regressions previously only containing gender variables. These job attributes had little effect on the estimated coefficients for occupational gender composition.

A more conventional approach to tackling unmeasured variables is to estimate a fixed effect model of earnings and feminization. Gerhart and El Cheikh (1991) use data from the National Longitudinal Study of Youth (NLSY) for the two years 1983 and 1986 when respondents were aged between 18 and 25 and 21 and 28 years, respectively. The authors provide both cross section and fixed effect wage estimates. The underlying earnings function includes the percentage of females in the individual's 3-digit 1970 occupation as well as a number of other occupational characteristics (such as specific vocational preparation, general educational development, and physical demands of the job), individual characteristics (education, weeks worked, collective bargaining coverage, marital status, inter al.), and industry and year dummies. In practice, the within-group model is estimated in first differences and separate regressions are run for males and females. Focusing on results from the longitudinal sample, the authors' between groups model suggest that a movement from a 100 male to a 100 percent female occupation is associated with a 21.6 percent decrease in earnings for males and a 5.8 percent decrease for women. But the within-group estimator reduces the percentage female coefficient by a third (and is not statistically significant) while the male coefficient is unchanged. The suggestion is that when fixed effects are added to models that control for occupation and industry, the impact of feminization in cross section may have more to do with (differences in) the types of people who choose to work in the more feminized occupations. Finally, when the authors decompose earnings differences into the components due to percentage female, individual characteristics, and the remaining variables (occupational characteristics, industry dummies and intercept) it is apparent that the individual and other characteristics and other variables dominate.

Despite its vintage, the final study considered here represents the most extensive evaluation to date of the role of gender composition in wage determination and is perhaps most representative of the current state of play in this area of research.[5] Macpherson and Hirsch (1995) use nationally representative national samples from the January 1983 through December 1993 monthly CPS Surveys, offering unusually large sample sizes (the total sample size is 1.84 million), in addition to various CPS supplements. The authors examine changes over time in the gender composition of jobs and its evolving effect on wages and the gender gap. The authors also estimate longitudinal wage change models for matched worker-year pairs from 1983/4 to 1992/3, now representing 25 percent of the size of the full sample.

Wage level results from the authors' standard model – familiarly containing individual characteristics, location, and broad occupation and industry – indicate that the gender composition (proportion female in the worker's 3-digit occupation) effect is large and of roughly the same absolute magnitude for both genders.[6] Expanded wage regressions containing job characteristics, such as mean years of required occupational training, computer usage, and indices of physical demands, produce much reduced gender composition coefficients (of roughly one-quarter (one-half) for women (men)), pointing to the influence of compensating differentials and/or quality sorting on the job characteristics associated with gender composition. In a final application of the wage level analysis, the authors examine the contribution of gender composition to the gender wage gap. For the standard model, gender composition accounts for more than half the explained portion of the gap, although this is reduced by about one-third for the expanded model.

Despite the importance of controlling for detailed job characteristics, there remains the issue of unmeasured skills and tastes. Here the authors' longitudinal analysis based on two-year panels of individuals seems to point to the decisive influence of person-specific labor quality and/or preference differences. Thus, for the standard model, estimates of the effects of gender composition are reduced by roughly one-half using longitudinal analysis. For the expanded model where the effects of gender composition are already much reduced, the coefficients of gender composition on wage change are just -0.055 for women and -0.034 for men. That is to say, unmeasured skills/tastes when added to job characteristics explain some two-thirds of the

[5] For a review of more recent other-country studies and an analysis of three nationally-representative British data sets – the British Household Panel Survey, the Labour Force Survey, and the Skills Survey – see Perales (2010).

[6] Suggesting that wages are 7 percent lower for each in a typical female occupation than in a typical male occupation, or, equivalently, that that a movement toward equality of gender composition would lead to a 3.6% increase in average female wages and a 3.4% decrease for a typical male.

standard gender composition effect among women and four-fifths of the effect among men. Expressed in terms of the wage gap, gender compositional differences explain just .02 log points of a wage gap that averaged 0.30 log points over the 1983-93 period. In short, the gender composition variable is "correlated with differences in job characteristics, worker-specific productivity differences among observationally-equivalent workers, and taste differences regarding job characteristics" (Macpherson and Hirsch, 1995: 455). The authors thus conclude that predominantly female jobs pay less to women (and men) mostly by reason of their skill-related characteristics and quality sorting with the unmeasured skills of both genders increasing in the proportion of males in an occupation.

The Macpherson-Hirsch study provides the motivation for the present paper because of its representativeness, use of an extensive set of variables (including importantly occupational skills and job disamenities), and complementary longitudinal analysis. Our goal is thus to update the analysis of CPS data to determine whether its (cross-section) findings continue to hold. Moreover, since its most optimistic results with respect to feminization have a basis in longitudinal analysis and given the limitations of the CPS in this regard – imprecise estimates because longitudinal data limited to two consecutive years are inadequate for significant mobility to occur – we shall follow Gerhart and El Cheikh (1991) in using the NLSY to provide both updated and better-suited data for this component of our analysis.

IV. Econometric Specification

In our econometric modelling, and in step with Macpherson and Hirsch (1995), we include many individual, job and occupation-related characteristics that contribute to productivity and human capital accumulation differences that may explain some of the wage disparity across genders. However, there is still room for unobserved taste and productivity differences, and we will need an econometric setup that accommodates these unobserved factors and the possible endogeneity of occupational choice and wage outcomes. Panel data methods are used to control for these unobserved time-invariant individual-specific effects.

We estimate for each gender:

$$\log\left(W_{ift}\right) = \theta_f FEM_{ift} + X'_{ift}\delta_f + Z'_{if}\beta_f + v_{ift}$$

$$\log(W_{imt}) = \theta_m FEM_{imt} + X'_{imt}\delta_m + Z'_{im}\beta_m + v_{imt} ,$$

where *f* and *m* are gender indicators; the *i* and *t* subscripts designate individuals and time, respectively; log(W) represents the natural log of hourly wages; FEM is an indicator of the proportion of females in each individual's occupation; X is a vector of observable time-varying individual-, job-, and occupation-level variables; Z is a vector of observable time-invariant characteristics; and θ, δ, and β (now excluding the gender indicators for simplicity) are the coefficients of interest. The error term v_{it} (for both genders) can be decomposed in the following way:

$$v_{it} = \mathrm{u}_i + \varepsilon_{it} ,$$

where u_i represents individual-specific time-constant unobservable effects; and ε_{it} is a stochastic error term.

Using CPS data, we first estimate the above models by year (and across years) and by gender, with and without human capital controls. We later add to these simple cross-sectional (yearly and pooled) models an extended set of job controls in an attempt to tease out the role of occupation and industry level characteristics. We employ a first-differenced wage regression as a complement to OLS regression to evaluate the extent to which the relationship between occupational feminization and wages is robust to the presence of unobserved individual heterogeneity which is potentially correlated with the observed factors.

We estimate the same set of cross sectional models with NLSY79 data. In utilizing the panel nature of this data set, in addition to first differenced (wage change) models we also estimate random effects and fixed effects models. The NLSY79 has a much greater number of time periods which enables us to achieve identification through a richer source of within-group variation. The standard errors in all models are adjusted to control for the clustering of observations within individuals whenever necessary.

We examine the cross sectional and longitudinal evidence on occupational feminization and wages in turn. Before turning to the longitudinal evidence, however, we investigate the sensitivity of the gender wage gap to the inclusion of the gender composition argument. Using standard procedures we decompose the gender wage gap by specification and year to throw further light on the contribution of occupational feminization to earnings.

V. Data Sources and Research Sample Construction

Given that the main interest of this paper is to extend and update the analysis in Macpherson and Hirsch (1995), we start where they left off and construct our sample from the CPS Merged Outgoing Rotation Groups (CPS-MORG) for the years 1993-2010.[7] In these data, each household is interviewed 8 times over 16 months. Specifically, households are interviewed for 4 consecutive months and then, after 8 months out of the sample, for a further 4 months. Since 1979 only households in their fourth and eighth interviews (the outgoing groups) have been asked the earnings question in the CPS-MORG survey. Approximately 60,000 households are interviewed monthly in this 4-8-4 rotation system. Sample size thus distinguishes the CPS-MORG data from other supplementary CPS files, yielding over 2.75 million observations for 1993-2010 of which 1.37 million are females.

Our CPS-MORG sample is restricted to workers aged 16 years or more. We do not consider the self-employed or those who work for no pay. The military sample is also excluded. The wage measure is hourly wages (viz. usual weekly earnings divided by usual hours worked), which are reflated by the monthly Consumer Price Index to December 2010 dollars. As in Macpherson and Hirsch, observations with real hourly wages lower than $1.00 are not used in this analysis. Moreover, adjusted mean earnings above the cap were assigned for the top-coded groups on the assumption that the upper tail of the earnings distribution follows a Pareto distribution.[8]

In addition to the CPS-MORG, we provide additional evidence using a long panel of individuals from the core cohort of the NLSY79 for the years 1993 to 2010. The NLSY79 provides a nationally representative panel of data for the cohort of individuals aged 14 to 22 years in 1979, and who have been interviewed regularly since that year. The core data exclude the oversample of Hispanic, black, and low income youth as well as the military. As for the CPS sample, we exclude individuals who are self-employed or who work for no pay. Having also excluded those with missing information on any of the variables used in the analysis, or having no data on hourly wages (or reporting hourly wages of less than $1. We also excluded observations where wage entries were clearly wrong[9]), we have about 32,000 person-year

[7] The data were downloaded from http://www.nber.org/morg/annual/ on August 11, 2014

[8] Top codes are contained in Hirsch and Macpherson (2011: 6)

[9] For example, we have a few observations where individuals experienced wage growth of more than 100%, followed by huge declines in the very next period with no material accompanying changes in job characteristics.

observations over the survey years analyzed. Use of the NLSY79 has a number of advantages. One is that it allows us to track workers' actual labor market experience, which corrects for the potential measurement error in the standard experience indicator based on age and education. Another is that it also allows us to control for ability through ASVAB test scores, unavailable in the CPS-MORG. Furthermore, we can make use of the *long* panel nature of NLSY79 to better model unobserved heterogeneity.[10]

Although labor market activity has been surveyed in CPS – and in great detail in the NLSY79 since its inception – the occupational and industry codes are not recorded consistently across each wave of either survey. Until 2002, occupations in the CPS were recorded using the 1990 Census Occupational Classification (COC)[11] while in the case of the NLSY79 the 1980 COC was used.[12] After this year, both datasets use only the 2000 COC to code occupations. Similarly, industries are described by their 3-digit 1990 Census Industry Classification (CIC) in the CPS[13] until 2002 and by 3-digit 1980 CIC in the NLSY until 2000. Thereafter, the industries are measured by 4-digit 2002 census code in both the CPS and the NLSY79**.** We mapped these occupation and industry codes so as to be able to study the full extent of the data panel available to us.[14] The occupations are divided into 6 separate, aggregated groups as in Macpherson and Hirsch as follows:

[1] Management/professional/technical/financial/sales/public security;

[10] As noted earlier, Macpherson and Hirsch also examine longitudinal evidence, constructed from the CPS ORG files, consisting of matched worker pairs (i.e. over adjacent years) for the period 1983/84 -1992/93

[11] The variable that captures the occupation codes is designated as occ80 in the data; however, this is a misnomer. The technical appendix shows that occupation codes were changed to COC 1990 codes in 1992 (viewed online at https://cps.ipums.org/cps/resources/earner/cpsxNBER.pdf, page 35).

[12] In the NLSY79 for 2000, occupations are measured by 1980 codes and for 2002, occupations are measured by 2000 census codes. In the CPS for 2000, 2001, 2002, the occupations are measured by both 1990 and 2000 census codes. The mapping for this paper uses the 1990 codes for these three years. In the CPS data there is a 0 appearing as the 4th digit in the 2000 codes. Dividing the latter by 10 gives the standard 3 digit 2000 census codes.

[13] Similar to the problem with occupation coding, in the case of industry codes the variable is designated as ind80 in the data; however, the technical appendix shows that industry codes are changed in 1992. https://cps.ipums.org/cps/resources/earner/cpsxNBER.pdf, page 34）

[14] We use do-files kindly provided by David Macpherson to create a program to map the 2000 occupation codes into 1990 and 1980 codes. For some of the missing occupations, we updated Macpherson's crosswalk using the distribution of COC1990 to COC2000 (using https://usa.ipums.org/usa/volii/occ_ind.shtml). Macpherson's do-files also map and group industry classifications. Some 14 industry groups were generated once all the crosswalks were completed using the guidelines provided in http://www.census.gov/people/io/methodology/. Blau et al. (2013) use gender-specific crosswalks and find that gender segregation is underestimated if aggregate mappings are used. For the period studied here, segregation indices calculated by either method are not very different and so we eschewed the use of gender specific crosswalks.

[2] Administrative support and retail sales;

[3] Low-skill service;

[4] Precision production and craft;

[5] Machine operators, assemblers and inspectors; and,

[6] Transportation/construction/mechanics/mining/agricultural.

And for industries we have the following 14 one-digit groups:

[1] Agriculture, forestry, fishing and hunting;

[2] Mining;

[3] Construction;

[4] Manufacturing (Durable Goods);

[5] Transportation and warehousing, and utilities and information;

[6] Wholesale trade;

[7] Retail trade;

[8] Finance and insurance, and real estate and rental and leasing;

[9] Professional, scientific, and management, and administrative and waste management services;

[10] Other services except public administration;

[11] Arts, Entertainment, and Recreation, and Accommodation and Food Services;

[12] Educational Services, and Health Care and Social Assistance;

[13] Public Administration; and,

[14] Manufacturing (Non-durable Goods).

We supplemented both datasets with occupational characteristics obtained from the Occupational Information Network (O*NET) and the Occupational Projections and Training Data (OPTD),[15] together with additional 3-digit industry and occupational level controls from the CPS supplements. From O*NET data we have information on strength and computer interaction requirements in each occupation, as well as occupational hazard levels and physical and environmental conditions. Besides working conditions and computer skills, we used

[15] The O*NET variables are those used in Hirsch and Schumacher (2012). The O*NET extract is from 2008 and the OPTD extract is from 2002. Details on the creation of the O*NET extract is described in Hirsch and Schumacher (2012). These data are for 2000 3-digit census occupational codes.

occupational education categories from the OPTD, capturing workers' levels of human capital accumulation including schooling and job training. The proportion of workers in big firms (having more than 1000 employees) was calculated from the 2003-2007 CPS Annual Social and Economic (ASEC) supplement and average job tenure for each occupation was generated using 2004-2010 CPS job tenure supplement. Annual levels of union membership for each industry were calculated from CPS data, as were occupational part-time employment shares.[16] Finally, our main control variable FEM measures the female intensity of an occupation, namely the share of female workers in the relevant 3-digit occupation.

VI. Descriptive Evidence

In Table 1 we report sample sizes and mean wages by gender, as well as the female-to-male wage ratio and Duncan segregation index for each year of the CPS. As expected, average male wages (ranging between \$20.73 in 1993 to \$24.87 in 2010) exceed those of females (\$15.87 in 1993 and \$19.44 in 2010). The Duncan segregation index is calculated as $\frac{1}{2}\sum|m_j - f_j|$, where m and f are the shares (in percent) of the male and female labor force in occupation j. Feminization (FEM) levels are reported separately by gender. The last two columns of the table give the estimates for θ_f and θ_m, namely the log wage regression coefficient on FEM without any other controls. These coefficients suggest a strong unconditional relationship between FEM and the average female wage, and a weaker one between FEM and the average male wage.

(Table 1 near here)

Figure 1 combines the data on segregation and the female-to-male wage ratio provided in this table with those in Macpherson and Hirsh (1995: Table 1) to illustrate the trends over the last four decades. We observe that the Duncan index has declined through time, perhaps suggesting some modest reduction in market segregation. But even if the market has become more integrated, as reported in the literature most occupations at the detailed level remain rather segregated by gender (see item A in the Appendix for some extreme examples). This tendency is captured in Figure 1 by an almost flat segregation line over the last decade of the sample period, plateauing at around 51 percent. Observe also that the female-to-male wage ratio broadly stabilized at around 0.80 over the last decade or so.

[16] Additional CPS data were downloaded from the Integrated Public Use Microdata Series (IPUMS) website.

Appendix Table 1 provides additional descriptive information, reporting means of selected variables by quartile of gender composition for occupations in 1994 and 2010. We observe that wages are lower for females on average, while for both genders they are lowest in predominantly female jobs. Even though both males and females have low wages in the predominantly male jobs, wages generally decline with rising female share in gender composition for the rest of the distribution. There is a clear U-shaped relationship in average levels of education and training. Female jobs also have a higher share of part-timers.

There are further differences in human capital and demographic characteristics among male and female employees by gender concentration of jobs. There are also differences in broad sector, and union status, as well as in occupational and industry characteristics across gender and FEM quartile. Our standard econometric specification uses the *individual* characteristics and general sector and unionization variables in a baseline Mincerian setup. We then expand this parsimonious model to incorporate the occupational- and industry-level differences of jobs. The next section reports estimates from cross-sectional models using the CPS-MORG data.

VII. Cross-Sectional Evidence

In Table 2 we provide estimates of θ_f and θ_m from "standard" and "expanded" models separately for each year of the sample period. The "standard" specification includes controls for years of schooling, potential experience (measured by age-schooling-5) and its square, and dummies for union coverage, public sector employment, large metropolitan area, full-time employment (usual hours worked are at least 35 hours), Hispanic heritage, race (2), marital status (2), region (8), industry (13), and occupation (5). The "expanded" specifications include all controls used in "standard" specifications and 10 additional occupational and industry controls to include controls for working conditions (environment, hazard, strength, physical), computer skills (computers), education and job training requirements (education & training), average job tenure for each occupation, the proportion of workers in big firms (having more than 1,000 employees), and the proportion who are union members for each industry. These controls are intended to capture the degree to which wage differences are compensating for job (dis)amenities and possible entry barriers that are likely to affect female occupational choice.

(Table 2 near here)

The third and seventh columns of Table 2 repeat the coefficient estimates from the summary “no controls” regression in Table 1 where, as we have seen, the relationship is very strongly negative for females and moderately so (especially towards the end of the study period) for males. When individual demographic and work history and current job characteristics are controlled for, along with occupation and industry dummies, the coefficients are 3 to 5 percentage points smaller in absolute magnitude for females. However, in the case of males the negative effect of gender composition doubles (or even triples) in the standard model relative to the regressions with no controls. According to these estimates, average wages are 5.3 percent [(0.678-0.327) * -0.151] lower for females and 6.3 percent [(0.678-0.327) * -0.181] lower for males in typical female jobs compared with typical male jobs in 2010. A non-segregated market (0.49 female presence in each job) would have increased female wages on average by 3 percent while decreasing male wages by the same rate.

Predominantly male and predominantly female occupations are very different and the FEM variable may be capturing these differences in job characteristics and thence their effect on wages. For example, there are more part-time jobs in predominantly female occupations. Among males more of the workers are in industries with larger shares of big firms in predominantly female jobs. Compared with predominantly male jobs they have lower strength requirements and lower exposure hazards. That said, they have slightly higher environmental scores meaning they experience more conflict, noise, and physical extremes. Also observe that, at a time when the rest of the occupations saw reduced unionization, predominantly female jobs preserved their share of union coverage.[17] By including occupational characteristics, the expanded model controls for such differences. And we see that the gender coefficients estimates are halved for females and reduced by about one-third for males. For example, by 2010 the FEM coefficient was -0.151 for females in standard model and -0.075 in expanded model; the corresponding values for males being -0.181 and -0.125. In the expanded specification and contrary to the standard specification the gender composition effect is no longer stronger for males. This means that most of the effect of feminization is explained away by human capital and job characteristics for females and there seems to be a negative quality sorting towards highly feminized jobs. For males, as the most

[17] For example, from Appendix Table 1 observe that predominantly male jobs had 19% (20%) unionization among females (males), which shares fell to 15% (16%) by the end of the sample period. On the other hand, in predominantly female jobs union coverage remained at 17% among females and increased by about 1 percentage point to 19% among males.

extreme male jobs are very low skill, poorly paying jobs the unconditional means are blurred by this non-linearity and do not capture the negative effect of FEM on the rest of the distribution.

(Table 3 near here)

In Table 3, we pool the data and treat feminization first as a continuous variable as before (referred to as ‘Model 1’) and then with dummies for each quartile (but excluding the predominantly male jobs to form the baseline group) in order to capture a possibly non-linear relationship between feminization and wages (‘Model 2’). Observe that the relationship while somewhat U-shaped in the model without controls is relatively linear for the standard and expanded models (as indeed it is for Macpherson and Hirsch). Once we control for human capital and occupational characteristics the impact of FEM is linear, implying that the highest wage penalties are experienced in highest feminization groups.

(Table 4 near here)

In order to capture those characteristics that contribute most to the FEM-wage relationship we next estimate models with alternative specifications. We see that the addition of broad industry categories to the base set of individual characteristics (line 3) does not explain much for females but reduces the FEM coefficient significantly in the case of males (from -0.131 to -0.054) at the same time as the inclusion of occupation dummies (line 4) strongly increases that coefficient. (For males, then, much of the negative correlation between gender composition and wages is accounted for by industry differences and occurs primarily within broad industry groups.) For females occupational education and training requirements (line 6) and share of part-timers (line 8) are most influential while physical (dis)amenities explain little once all other occupational and industry level characteristics are controlled for (comparing lines 13 and 14). This phenomenon may be capturing occupations such as nursing and teaching for where physical demands of the jobs are high and pay is quite low compared with the other jobs for highly educated individuals. Even though the role played by the physical demands of the job is similarly unimportant for males, differences in training requirements and part-time status do not explain as much of the negative FEM wage relationship in their case. For females neither occupational tenure nor computer use explain much of the variation (with very slight decreases in FEM coefficient) while for males inclusion of either increases the FEM coefficient significantly. These results indicate that a sizable portion of the negative wage-FEM relationship is due to occupational differences in skill requirements and job attachment for females. Another observation is that the gender specific nature of these relationships confirms the need for separate estimates for males and females.

(Table 5 near here)

In Table 5, our standard and expanded models are estimated for different educational, demographic, and occupational subgroups. With respect to age, the most negative effects obtain for 30-40 year olds. The overall relationship between age and the gender composition effect is somewhat U-shaped. Individuals may be sorting into female jobs when they need flexibility for fertility reasons or to care for elderly relatives, responsibilities that also reduce productivity. Timing of these events very likely overlap with the mid-career years when occupational investments such as training or longer work hours may yield the highest wage returns, resulting in this U-shaped relationship. With respect to marital status, among females the biggest effect is where the individual is married with a spouse present; for males, it is for the once married who are now separated, divorced or widowed. In the case of education, those with the highest education levels (16 or more years of schooling) are the most damaged by gender composition, and those with the next highest levels (13-15 years of schooling) in the case of males. As far as race is concerned, the harmful effect of feminization is lowest among blacks of both genders, although this outcome may of course be capturing the lack of opportunities confronting blacks in high wage markets. The negative gender composition effect is also larger in the non-union sector in the expanded model for both genders, although on this occasion not for the standard model in the case of males. Finally, the gender composition penalty applies generally to full-time work; indeed, the FEM coefficient is positive for females in part-time jobs.

(Table 6 near here)

Earlier we examined the sensitivity of FEM coefficient. In Table 6 we decompose the log wage gap between males and females, now exploring the sensitivity of the gender wage gap by specification and year to the inclusion of FEM. In the standard model without FEM, human capital attributes explain about 17 percent of the observed wage gap. With the inclusion of FEM it can be seen that the unexplained portion is reduced by about 0.02 to 0.05 log points – from 0.14 to 0.16 log points. For the standard model, some 90 percent of the explained difference in the wage gap is explained by gender compositional differences between men and women. This falls to about 43 percent in the expanded model (average across all years), with the highest contribution being in 2002 (approximately 61 percent) and the lowest in 1994 (28 percent). The share of part timers explains about 6 percent of the overall difference in earnings and 13 to 24 percent of the explained portion. While the education and training requirements of a job do not contribute to the explanation of the wage gap, computer usage and physical nature of jobs contribute to reducing it. However, remaining occupational characteristics (such as union presence and share of big firms) contribute to 15 to 45 percent of the wage gap. But the main impression conveyed by Table 6, however, is the very scale of the unexplained part of the gender gap. Even with a very full set of human capital and job controls, some 60 to 70 percent of the wage gap remains unexplained. This outcome may result from unobserved individual differences in tastes and productivity, or it may reflect discrimination. In modelling unobserved individual heterogeneity using longitudinal data, we will directly (indirectly) explore the relevance of the former (latter) explanation, while also returning to the discrimination issue in our concluding remarks.

VIII. Longitudinal Evidence

In our preceding cross-sectional analysis, the role of observed individual characteristics and industry and occupation level job attributes in explaining the negative relationship between feminization and wages has been established. However, even after controlling for a rich set of these characteristics, there remains a negative relationship between FEM and wages that is both economically and statistically significant. In Table 7, we probe the role of unobserved factors by utilizing the panel nature of the CPS-MORG data. In addition, we run a wider set of panel models using our NLSY79 sample, the main contribution of this paper being not only to extend and update the in-depth study of Macpherson and Hirsch (1995) but also to expand their analysis using a much longer panel data that surveys individual work histories more thoroughly. Here, we will firstly incorporate unobserved heterogeneity to our standard and expanded models and then substitute actual tenure and work experience for potential experience *and* control for unobserved ability by using age- and education-adjusted ASVAB scores in an additional ('expanded plus') specification. We note parenthetically that the NLSY79 sample closely resembles its CPS-MORG counterpart in descriptive statistics and with respect to the estimates derived from the pooled data and cross-sectional models. Appendix Tables 2, 3, and 4 contain the descriptive and cross-sectional evidence from the NLS, complementing our earlier CPS-MORG data results.

(Table 7 near here)

The top panel of Table 7 reports wage level (pooled OLS) and wage change equations (first differenced models) using the CPS-MORG data.[18] Pooled OLS level FEM coefficients are similar to those given in Table 3. Were we to obtain similar results from the wage change equations, we would conclude that unobserved heterogeneity cannot contribute anything beyond the explanation offered by human capital characteristics and job attributes, with the FEM coefficients likely capturing the discrimination and resulting crowding effects on wages. But this is not the case. The results for the expanded version of the wage change equations reported in the upper panel of the table yield FEM coefficient estimates that are significantly different from the estimates derived from the wage level equations for both females and males. The suggestion is then that preference/taste differences and possibly unobserved productivity play a very important role in explaining away the effects of gender composition on wages. Even though the estimates are much smaller and indeed just positive for females, they are still statistically significantly different from zero.[19] For males, statistically significant negative effects still persist, but their magnitude is only about 13 percent of the corresponding value in the level estimates (-0.018 as compared with -0.139). As the level results are very close to those reported in Table 3 for the whole CPS sample, we can characterize them as nationally representative even though we had to restrict the sample only to those individuals who were interviewed in both outgoing rounds (namely, the 4th and 8th interviews as described in the data section) and can be matched across interviews.

In the lower panel of Table 7 we report not only wage level (OLS) and wage change FEM coefficient estimates but also the corresponding estimates from random effects and fixed effects models for the NLSY79 sample.[20] Note that the NLSY79 data are from a much longer panel, so that identification is obtained through more rounds of within variation for the FE and FD models.[21] Moreover, as was noted above, in an additional model specification the measure of

[18] Not all observations can be matched across rounds in the CPS, which reduces our sample to 1,535,538 observations from 767,769 individuals.

[19] Macpherson and Hirsch restrict their sample to individuals who have changed jobs over the 16 month interval between two interviews. We cannot impose such restrictions in the NLSY79 data as most of the job changes occur in early career years and our sample (aged 28 to 36 years in 1993 and 45 to 53 years in 2010) is no longer very mobile. We can apply this restriction to CPS-MORG panel, however, and the results are no longer statistically significant for females. These results are available from the authors upon request.

[20] Hausman test statistics indicate that only the fixed effects model estimates are consistent and we cannot ignore the possibility of correlation between observed covariates and the unobserved fixed individual component.

[21] Observe that we ignore the 1993 round of the data and only use information from the 1994 and subsequent rounds (namely, that portion of the data for which the NLSY79 becomes biennial). As a result, estimates for the wage

potential experience is replaced by the actual labor market experience and controls for age and unobserved ability (through age- and education-adjusted ASVAB scores) are added. Males and females of the same age and education may differ significantly in their market experiences as females' labor market participation is frequently interrupted. For its part, the inclusion of age and education- adjusted ability scores helps distinguish between unobserved taste and unobserved ability explanations.

Comparing the wage level estimates from pooled data from the CPS-MORG and NLSY79 samples, we observe that the negative effects of female density in a job are stronger for both males and females in the NLSY79 data. In standard panel data models – that is, with only human capital and demographic controls – the magnitude of the FEM coefficient is reduced by 35 to 50 percent for females and by 50 to 65 percent for males, although it is still statistically significant for all specifications for both genders. In the expanded models, however, not only are the FEM coefficient estimates reduced more dramatically but they are also no longer statistically significant for males in the FE and FD models. For females, however, the significance of the negative effects persists and approximates 40 percent of the levels estimate**.** Comparing the expanded and expanded plus specifications, we observe that unobserved tastes and preferences explain away statistically significant negative FEM coefficients for males in their mid- and late-career years. However, for this older cohort of females, unlike the CPS-MORG cohort, even when unobserved tastes and abilities (which contribute very little to the explanation) are controlled for,[22] there remains a log point difference of about 0.02. This value is less than 10 percent of the overall difference of 0.250 log points in 2010, as obtained from log wage decompositions that are not reported here.

IX. Conclusions

Not only are females encouraged to enter (currently) male dominated and highly paid fields (e.g. Science, Technology, Engineering, and Mathematics, the so-called 'STEM' disciplines) but also technological advances now make it possible to perform many physical jobs without the exertion of physical power, thereby eliminating a male advantage. Rates of college graduation are now

change regression have log wage change over a two-year year period as the outcome. For the purposes of comparison with the CPS results, therefore, the estimates need to be adjusted, that is, roughly halved.

[22] Calculated by multiplying the FEM coefficient by the difference in feminization rates between typical female and male jobs in 2010 (see Appendix Table 2); that is, 0.048*(0.666-0.293) = 0.0179.

much higher among females, rendering them equally or more qualified. Also, female participation rates are almost as high as those for males, making them also equivalent in the accumulation of experience, at least for the early career years. That said, none of these changes seems to be enough to actually eliminate wage differences or make labor markets less segregated. Even after many decades of increasing female presence in the labor market and evolving gender roles we still have male and females jobs. Moreover, female jobs are lower paid jobs not only for females but also for males. Do these jobs require less education, less experience, and less overall human capital? Or they are crowded with an excess supply of female labor that is discriminated against or excluded everywhere else? In the present exercise, we have examined the extent to which a higher share of females in a job contributes to observed wage differences. In seeking to understand the role of feminization, we have explored explanations such as quality sorting, discrimination, and unobserved differences in tastes and abilities.

Our results, in common with those of Macpherson and Hirsch, indicate that only a portion of the wages of males and females are explained by gender composition. Thus, for females, when we control for worker and occupational characteristics, the effect of gender composition declines materially. The specifics are as follows. In cross section, our FEM coefficients remain significant and negative for both genders, although in the presence of the human capital and occupational controls they are reduced significantly for females. The panel estimates for females using the CPS-MORG data are in fact no longer negative, albeit only marginally significant. For males, on the other hand, gender composition effects for the pooled sample become more negative in the presence of occupational controls, while in the panel estimates the negative impact of gender composition persists. The suggestion is that women tend to sort into predominantly female jobs either because of their lower unobserved skills or because of their unobserved taste differences that are correlated with gender composition and unmeasured job characteristics.

The NLSY79 provides us with a longer panel for an aging cohort. These individuals were a nationally representative sample of 14 to 22 year olds in 1979, most of whom had started their working lives by 1985. The majority of them are therefore well into their careers at the beginning of the study period and most are approaching retirement age by its end in 2010. Our pooled data analysis for this cohort yields a much higher negative FEM coefficient estimate, about half of which is explained away by demographic and human capital controls and by occupation and industry related characteristics. In contrast to the CPS-MORG findings, panel data analysis of this dataset indicates that the negative FEM coefficient for females remains economically and statistically significant throughout, even if now only roughly half of that obtained for the pooled data case. This finding indicates first of all that unobserved factors play a role in gender sorting into jobs. Moreover, it also points to possible existence of heritage effects. Younger cohorts may be less subject to occupational crowding today and perhaps given more room to do men's jobs.

We would conclude along with Macpherson and Hirsch that policies directed towards increasing the 'female component' in male jobs through quotas and the like will not be enough to solve wage discrepancies as the wage penalties paid by females for working in female jobs seem to be compensation for non-wage job attributes such as flexible schedules. In other words, even if they had the skills to be employed in the male jobs, females may choose not to enter them, on the grounds that such jobs will not provide sufficient flexibility, inter al. In order to increase female presence in male jobs, then, policies need to be directed toward addressing the (dis)amenities of male and mixed gender jobs through such measures as paid parental leave and family sick leave. As it stands, females in these (male) jobs may be having to sacrifice more financially or are being expected to accept less when seeking similar levels of flexibility and benefits that female jobs are possibly offering.

References

Becker, Gary S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* 3(1): 33-58.

Bergmann, Barbara R. 1974. "Occupational Segregation, Wages and Profits When Employers Segregate by Race and Sex." *Eastern Economic Journal* 1(2): 103-110.

Blau, Francine D., and Andrea H. Beller. 1988. "Trends in Earnings Differentials by Gender 1971-1981." *Industrial and Labor Relations Review* 41(4): 513-529.

Blau, Francine, Peter Brummund, and Albert Liu, 2013. "Trends in Occupational Segregation by Gender 1970-2009: Adjusting for the Impact of Changes in the Occupational Coding System." *Demography*, 50(2): 471-492.

Clausen, John A. 1968. *Socialization and Society*. Boston, MA: Little Brown and Company.

Gerhart, Barry, and Nabil El Cheikh. 1991. "Earnings and Percentage Female: A Longitudinal Study." *Industrial Relations* 3(1): 62-78.

Groshen, Erica L. 1991. "The Structure of the Female/Male Differential: Is it Who You Are, What You Do, or Where You Work?" *Journal of Human Resources* 26(3): 457-472.

Hirsch, Barry T. and David A. Macpherson. 2011. *Union Membership and Earnings Data Book: Compilations from the Current Population Survey*. Washington, DC: Bureau of National Affairs. Available at http://unionstats.gsu.edu/Pareto Top-Code Earnings.xls .

Hirsch Barry T., and Edward J. Schumacher. 2012. "Underpaid or Overpaid: Wage Analysis for Nurses Using Job and Worker Attributes." *Southern Economic Journal* 78(4): 1096-1119.

Hirsch Barry T., and Julia Manzella. 2015. "Who Cares – And Does it Matter? Measuring the Wage Penalty for Caring Work." *Research in Labor Economics*: *Why Are Women Becoming More Like Men (and Men More Like Women) in the Labor Market?* Volume 40, pp. 213-275.

Johnson, George, and Gary Solon. 1986. “Estimates of the Direct Effects of Comparable Worth Policy.” *American Economic Review* 76(5): 1117-1125.

Kilbourne, Barbara S., Paula England, George Farkas, Kurt Beron, and Dorothea Weir. 1994. “Returns to Skills, Compensating Differentials, and Gender Bias: Effects of Occupational Characteristics on the Wages of White Men and Women.” *American Journal of Sociology* 100(3): 689-719.

Lewis, Gregory B. 1996. “Gender Integration of Occupations in the Federal Civil Service: Extent and Effects on Male-Female Earnings.” *Industrial and Labor Relations Review* 49(3): 472-483.

Macpherson, David A., and Barry T. Hirsch. 1995. “Wages and Gender Composition: Why Do Women’s Jobs Pay Less?” *Journal of Labor Economics* 13(3): 426-471.

McArthur, Leslie. 1985. “Social Judgment Biases in Comparable Worth Analysis.” In Heidi I. Hartmann (ed.), *Comparable Worth: New Directions for Research*, pp. 53-70. Washington, DC: National Academy Press.

Magnusson, Charlotta. 2009. “Gender, Occupational Prestige, and Wages: A Test of Devaluation Theory.” *European Sociological Review* 25(1): 87-101.

Meyer, P.B, and A. M. Osborne. 2005. “Proposed Category System for 1960-2000 Census Occupations.” BLS Working Paper No. 383, Washington, D.C.: U.S. Bureau of Labor Statistics.

Osterman, Paul. 1987. “Choice of Employment Systems in Internal Labor Markets.” *Industrial Relations* 26(1): 46-67.

Paulin, Elizabeth A., and Jennifer M. Mellor. 1996. “Gender, Race, and Promotion within a Private- Sector Firm.” *Industrial Relations* 35(2): 276-295.

Perales, Francisco. 2010. “Occupational Feminization, Specialized Human Capital and Wages: Evidence from the British Labour Market.” Institute for Social and Economic Research Working Paper No. 2010-31, University of Essex.

Pergamit, Michael R., and Jonathan R. Veum. 1999. "What Is a Promotion?" *Industrial & Labor Relations Review* 52(4): 581-601.

Polachek, Solomon W. 1981. “Occupational Self-selection: a Human Capital Approach to Sex Differences in Occupational Structure.” *Review of Economics and Statistics* 63(1): 60-69.

Polachek, Solomon W. 1985. “Human Capital: A Defense of Human Capital Predictions.” *Journal of Human Resources* 20(3): 437-440.

Schumann, Paul L., Dennis A. Ahlburg, and Christine Brown Mahoney. 1994. “The Effects of Human Capital and Job Characteristics on Pay.” *Journal of Human Resources* 29(2): 481-503.

Sorensen, Elaine. 1990. “The Crowding Hypothesis and Comparable Worth.” *Journal of Human Resources* 25(1): 55-89.

Tam, Tony. 1997. “Sex Segregation and Occupational Gender Inequality in the United States: Devaluation or Specialized Training.” *American Journal of Sociology* 102(6): 1652-1692.

Treiman Donald J., and Heidi I. Hartmann (eds.) 1981. *Women. Work, and Wages: Equal Pay for Jobs of Equal Value*. Washington, DC: National Academy Press.

Appendix: Examples of ‘Female’/‘Male Occupations, and Further information on OPTD and O*NET Variables

A) Extreme Occupations:

Examples of female jobs include the following (FEM given in parentheses):

- Kindergarten and earlier school teachers (98%);
- Dental hygienists (98%);
- Dental assistants (97%);
- Secretaries (97%);
- Child care workers (94%);
- Licensed practical nurses (94%).

Examples of male jobs include the following:

- Heavy equipment and farm equipment mechanics (1%);
- Drillers of oil wells (1%);
- Elevator installers and repairers (1%);
- Bus, truck, and stationary engine mechanics (1%);
- Plasterers (1%);
- Concrete and cement workers (1%).

B) The OPTD Education & Training Categories:

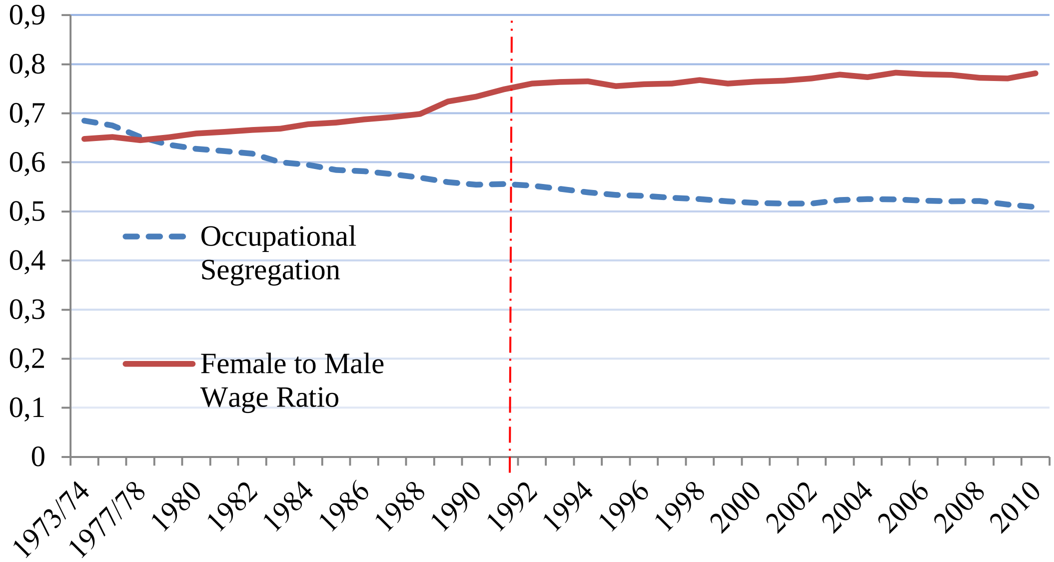
1	First professional degree.
2	Doctor's degree.
3	Master's degree.
4	Degree plus work experience.
5	Bachelor's degree.
6	Associate's degree.
7	Postsecondary vocational award.
8	Work experience in a related occupation.
9	Long-term on-the-job training.
10	Moderate-term on-the-job training.
11	Short-term on-the-job training.

C) *Content of O*NET Working Conditions Indices:*

38 of the 259 O*NET variables used by Hirsch and Schumacher (2012) and Hirsch and Manzella (2015) are as follows:

Static Strength Explosive Strength Dynamic Strength Trunk Strength Stamina	strength
Frequency of Conflict Situations Deal With Unpleasant or Angry People Deal With Physically Aggressive People Indoors, Environmentally Controlled Indoors, Not Environmentally Controlled Outdoors, Exposed to Weather Outdoors, Under Cover In an Open Vehicle or Equipment In an Enclosed Vehicle or Equipment Physical Proximity Sounds, Noise Levels Are Distracting or Uncomfortable Very Hot or Cold Temperatures Extremely Bright or Inadequate Lighting Cramped Work Space, Awkward Positions	environment
Exposed to Contaminants Exposed to Whole Body Vibration Exposed to Radiation Exposed to Disease or Infections Exposed to High Places Exposed to Hazardous Conditions Exposed to Hazardous Equipment Exposed to Minor Burns, Cuts, Bites, or Stings Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection	hazard
Spend Time Sitting Spend Time Standing Spend Time Climbing Ladders, Scaffolds, or Poles Spend Time Walking and Running Spend Time Kneeling, Crouching, Stooping, or Crawling Spend Time Keeping or Regaining Balance Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls Spend Time Bending or Twisting the Body Spend Time Making Repetitive Motions	physical

Figure 1: Occupational Segregation in the Labor Market and the Female to Male Wage Ratio



Note: 1973-1993 data are taken from Macpherson and Hirsh (1995), dashed vertical line indicating the end of their study period.

Table 1: Mean Wages, the Wage Gap, Gender Composition, and the Wage-Gender Composition Relationship by Year, 1993-2010

Year	Female N	Female Wage	Female FEM	Male N	Male Wage	Male FEM	Female to Male Wage Ratio	Duncan Segregation Index	θ_f	θ_m
1993	82,032	15.87	0.689	84,965	20.73	0.300	0.765	0.540	-0.198**	-0.072**
1994	76,963	16.25	0.689	78,914	21.23	0.303	0.765	0.539	-0.212**	-0.081**
1995[a]	57,510	16.13	0.687	59,007	21.34	0.306	0.756	0.534	-0.238**	-0.077**
1996	69,011	15.96	0.686	70,552	21.02	0.307	0.759	0.532	-0.227**	-0.066**
1997	70,175	16.27	0.682	72,088	21.39	0.309	0.761	0.528	-0.214**	-0.082**
1998	70,875	16.75	0.679	73,040	21.81	0.311	0.768	0.526	-0.233**	-0.073**
1999	71,683	17.07	0.677	74,234	22.45	0.312	0.760	0.521	-0.238**	-0.054**
2000	72,180	17.25	0.674	74,955	22.56	0.314	0.765	0.517	-0.235**	-0.053**
2001	76,656	17.73	0.675	79,142	23.12	0.315	0.767	0.516	-0.228**	-0.033**
2002	82,835	18.10	0.676	84,388	23.47	0.318	0.771	0.516	-0.208**	-0.003
2003	81,896	18.22	0.682	82,362	23.40	0.316	0.779	0.523	-0.181**	-0.033**
2004	79,866	18.30	0.681	81,231	23.64	0.314	0.774	0.526	-0.185**	-0.017+
2005	80,205	18.44	0.680	81,630	23.56	0.314	0.783	0.525	-0.199**	0.000
2006	79,986	18.47	0.678	82,185	23.69	0.313	0.780	0.522	-0.200**	-0.004
2007	80,093	18.68	0.678	81,378	24.00	0.317	0.778	0.521	-0.227**	0.010
2008	79,711	18.85	0.679	80,123	24.39	0.319	0.773	0.522	-0.201**	-0.009
2009	77,878	19.31	0.680	76,820	25.05	0.325	0.771	0.514	-0.193**	-0.039**
2010	77,031	19.44	0.678	75,958	24.87	0.327	0.782	0.509	-0.175**	-0.023*

Notes: Data are from the 1993-2010 CPS MORG files. Wages are calculated as usual weekly earnings divided by usual weekly hours and are converted into December 2010 dollars. The Duncan segregation index is calculated by $1/2\sum|m_j - f_j|$, where m_j and f_j are the percentage of the male and the female labor force in occupation j. θ_f and θ_m are the gender-composition or FEM coefficients from the regression of log wages on feminization (FEM) with no other controls. [a]Metropolitan area (smsastat) information is missing for June, July, and August. **,*, + denote statistical significance at 0.01, 0.05, and 0.1 levels, respectively.

Table 2: Gender Composition Coefficients from Unconditional, Standard, and Expanded Specifications, by Gender and Year, Wage Level Equations, 1993-2010

	Females				Males			
Year	*N*	No Controls	Standard	Expanded	*N*	No Controls	Standard	Expanded
1993	82,032	-0.198**	-0.177**	-0.070**	84,965	-0.072**	-0.217**	-0.144**
		[0.008]	[0.007]	[0.008]		[0.008]	[0.008]	[0.010]
1994	76,963	-0.212**	-0.163**	-0.059**	78,914	-0.081**	-0.198**	-0.103**
		[0.008]	[0.008]	[0.009]		[0.009]	[0.009]	[0.011]
1995	57,510	-0.238**	-0.183**	-0.077**	59,007	-0.077**	-0.227**	-0.138**
		[0.010]	[0.009]	[0.011]		[0.010]	[0.011]	[0.013]
1996	69,011	-0.227**	-0.180**	-0.087**	70,552	-0.066**	-0.223**	-0.129**
		[0.009]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
1997	70,175	-0.214**	-0.183**	-0.095**	72,088	-0.082**	-0.244**	-0.162**
		[0.009]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
1998	70,875	-0.233**	-0.181**	-0.082**	73,040	-0.073**	-0.245**	-0.152**
		[0.009]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
1999	71,683	-0.238**	-0.192**	-0.107**	74,234	-0.054**	-0.253**	-0.159**
		[0.009]	[0.008]	[0.010]		[0.009]	[0.009]	[0.012]
2000	72,180	-0.235**	-0.198**	-0.111**	74,955	-0.053**	-0.269**	-0.168**
		[0.009]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
2001	76,656	-0.228**	-0.188**	-0.094**	79,142	-0.033**	-0.275**	-0.174**
		[0.008]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
2002	82,835	-0.208**	-0.183**	-0.094**	84,388	-0.003	-0.266**	-0.167**
		[0.008]	[0.008]	[0.010]		[0.009]	[0.009]	[0.012]
2003	81,896	-0.181**	-0.143**	-0.054**	82,362	-0.033**	-0.186**	-0.128**
		[0.008]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
2004	79,866	-0.185**	-0.163**	-0.063**	81,231	-0.017+	-0.186**	-0.139**
		[0.008]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
2005	80,205	-0.199**	-0.171**	-0.069**	81,630	0	-0.172**	-0.108**
		[0.009]	[0.008]	[0.010]		[0.009]	[0.010]	[0.012]
2006	79,986	-0.200**	-0.168**	-0.069**	82,185	-0.004	-0.154**	-0.086**
		[0.009]	[0.009]	[0.010]		[0.009]	[0.010]	[0.012]
2007	80,093	-0.227**	-0.175**	-0.064**	81,378	0.01	-0.187**	-0.114**
		[0.009]	[0.009]	[0.010]		[0.009]	[0.010]	[0.013]
2008	79,711	-0.201**	-0.156**	-0.050**	80,123	-0.009	-0.176**	-0.101**
		[0.009]	[0.009]	[0.010]		[0.009]	[0.010]	[0.013]
2009	77,878	-0.193**	-0.167**	-0.068**	76,820	-0.039**	-0.179**	-0.098**
		[0.009]	[0.009]	[0.010]		[0.009]	[0.011]	[0.013]
2010	77,031	-0.175**	-0.151**	-0.075**	75,958	-0.023*	-0.181**	-0.125**
		[0.009]	[0.009]	[0.011]		[0.010]	[0.011]	[0.013]

Notes: The "no controls" specification reports FEM coefficients (θ_f and θ_m) from regressions with no other controls. The "standard" specification includes controls for years of schooling, potential experience (measured by age-schooling-5) and its square, and dummies for union coverage, public sector employment, large metropolitan area, full-time employment (usual hours worked are at least 35 hours), hispanic, race (2), marital status (2), region (8), industry (13), and occupation (5). "Expanded" specifications include all controls used in "standard" specifications and 10 additional occupational and industry controls. Standard errors are in brackets. **,* denote statistical significance at the 0.01 and 0.05 levels, respectively.

Table 3: Gender Composition Coefficients from Linear and Dummy Variable Models, Pooled Sample, 1993-2010

	Model 1	Model 2		
Specification	FEM	FEM25-49	FEM50-74	FEM75+
Females:				
No Controls	-0.210**	0.080**	0.031**	-0.067**
	[0.002]	[0.002]	[0.002]	[0.002]
Standard	-0.175**	-0.024**	-0.102**	-0.124**
	[0.002]	[0.002]	[0.002]	[0.002]
Expanded	-0.076**	0.009**	-0.027**	-0.027**
	[0.002]	[0.002]	[0.002]	[0.002]
N		1,366,586		
Males:				
No Controls	-0.037**	0.139**	0.070**	-0.165**
	[0.002]	[0.001]	[0.002]	[0.002]
Standard	-0.214**	0.006**	-0.085**	-0.138**
	[0.002]	[0.001]	*[0.001]*	[0.002]
Expanded	-0.134**	0.027**	-0.031**	-0.043**
	[0.003]	[0.001]	[0.002]	[0.002]
N		1,392,972		

Notes: In Model 1 the feminization variable is a continuous measure, while in Model 2 it is coded into three occupational female intensity dummies where the reference group is FEM< 25%. Year dummies are included in all models. The standard and expanded are defined as in Table 2. Standard errors are in brackets. ** denotes statistical significance at the 0.01 level.

Table 4: Gender Composition Coefficient Sensitivity to Specification, Pooled Data, 1993-2010

Specifications	Females	Males
1. No controls	-0.210**	-0.037**
2. Base (individual characteristics only)	-0.181**	-0.131**
3. Base + 13 industry dummies	-0.204**	-0.054**
4. Base + 5 occupation dummies	-0.167**	-0.278**
5. Standard model (base model + 5 occupation, 13 industry dummies)	-0.175**	-0.214**
6. Standard + OPTD education & training	-0.113**	-0.170**
7. Standard + Occupation tenure	-0.177**	-0.205**
8. Standard + Occupation part-time	-0.105**	-0.120**
9. Standard + O*NET computer	-0.160**	-0.195**
10. Standard + OPTD education & training, Occupation tenure, Occupation part-time, O*NET computer	-0.093**	-0.114**
11. Standard + O*NET environment, hazards, physical, strength	-0.180**	-0.226**
12. Standard + Industry big firm, union	-0.162**	-0.216**
13. Expanded (Standard + all job characteristics)	-0.076**	-0.134**
14. Expanded without O*NET physical	-0.071**	-0.135**
N	1,366,586	1,392,972

Notes: Coefficients shown are θ_f and θ_m. The “base” model excludes industry and occupation dummies. The standard and expanded specifications are described in Table 2. All models include year dummies. ** denotes statistical significance at the 0.01 level.

Table 5: Gender Composition Coefficients among Different Worker Groups, Wage Level Equations, Pooled Sample, 1993-2010

	Female			Male		
Group	*N*	Standard	Expanded	*N*	Standard	Expanded
All workers	1,366,586	-0.175**	-0.076**	1,392,972	-0.214**	-0.134**
Age:						
16-29	347,542	-0.104**	-0.032**	360,217	-0.159**	-0.127**
30-39	328,634	-0.222**	-0.106**	353,217	-0.255**	-0.152**
40-49	351,844	-0.208**	-0.101**	345,761	-0.240**	-0.130**
50-59	245,526	-0.166**	-0.067**	240,299	-0.213**	-0.108**
60+	93,040	-0.086**	-0.045**	93,478	-0.171**	-0.112**
Marital Status:						
Married spouse present	733,186	-0.206**	-0.108**	832,522	-0.239**	-0.135**
Married spouse not present/Divorced/Widowed	277,203	-0.168**	-0.075**	166,369	-0.252**	-0.171**
Never married	356,197	-0.123**	-0.037**	394,081	-0.162**	-0.105**
Education (in years):						
0-11	115,324	-0.088**	-0.052**	161,163	-0.146**	-0.097**
12	430,605	-0.117**	-0.032**	458,149	-0.148**	-0.106**
13-15	430,005	-0.094**	-0.006	378,128	-0.176**	-0.082**
16	268,418	-0.267**	-0.164**	261,758	-0.323**	-0.150**
>16	122,234	-0.298**	-0.152**	133,774	-0.188**	0.006
Race:						
White	1,134,952	-0.173**	-0.079**	1,194,765	-0.213**	-0.133**
Black	149,376	-0.155**	-0.043**	112,852	-0.142**	-0.078**
Other race	82,258	-0.205**	-0.067**	85,355	-0.266**	-0.151**

Continues next page

Continued from previous page

	Female			Male		
Group	*N*	Standard	Expanded	*N*	Standard	Expanded
Sector:						
Private	1,092,885	-0.157**	-0.089**	1,186,838	-0.182**	-0.130**
Public	273,701	-0.228**	-0.052**	206,134	-0.265**	-0.142**
Union status:						
Nonunion	1,181,767	-0.169**	-0.089**	1,160,244	-0.193**	-0.131**
Union	184,819	-0.172**	0.033**	232,728	-0.243**	-0.103**
Hours status:						
Part-time	341,063	0.018**	0.063**	145,973	-0.086**	-0.018+
Full-time	1,025,523	-0.229**	-0.110**	1,246,999	-0.231**	-0.140**
Occupation:						
Managerial and Professional	473,988	-0.123**	-0.112**	401,318	-0.177**	-0.109**
Administrative Support and Retail Sales	529,279	-0.239**	-0.117**	261,935	-0.305**	-0.172**
Services	246,575	-0.065**	-0.004	175,291	-0.061**	0.050**
Farming, Fishing, and Forestry	5,616	0.060*	0.167*	24,844	-0.236**	0.372**
Construction/Extraction/Maintenance/Repair	18,391	-0.287**	-0.078**	248,650	-0.128**	-0.160**
Production/Transportation/Material Moving	92,737	-0.175**	0.009	280,934	-0.158**	-0.120**

Notes: Coefficients shown are θ_f and θ_m. The standard and expanded specifications are described in Table 2. All models include year dummies. **,*, + denote statistical significance at 0.01, 0.05, and 0.1 levels, respectively.

Table 6: Decomposition of the Gender Wage Gap, by Specification and Year

Specification	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total log gap	0.228	0.240	0.248	0.241	0.238	0.234	0.239	0.231	0.229	0.220	0.214	0.212	0.203	0.205	0.208	0.210	0.209	0.198
Standard specification without FEM:																		
1a. Unexplained	0.185	0.189	0.197	0.193	0.197	0.193	0.202	0.197	0.197	0.194	0.172	0.179	0.175	0.172	0.175	0.174	0.169	0.164
1b. Total explained	0.043	0.052	0.050	0.048	0.041	0.042	0.037	0.034	0.032	0.026	0.042	0.033	0.029	0.033	0.033	0.035	0.039	0.034
Standard specification:																		
2a. Unexplained	0.139	0.146	0.152	0.149	0.152	0.148	0.157	0.149	0.153	0.151	0.142	0.147	0.142	0.142	0.143	0.144	0.138	0.135
2b. Total explained	0.089	0.094	0.096	0.092	0.086	0.086	0.082	0.081	0.076	0.069	0.072	0.065	0.061	0.064	0.065	0.066	0.070	0.063
2c. Explained due to FEM	0.078	0.072	0.079	0.077	0.080	0.079	0.082	0.085	0.082	0.080	0.057	0.062	0.063	0.059	0.063	0.058	0.059	0.056
Expanded specifications:																		
3a. Unexplained	0.142	0.149	0.155	0.152	0.154	0.150	0.159	0.151	0.155	0.153	0.145	0.151	0.146	0.145	0.146	0.147	0.141	0.138
3b. Total explained	0.086	0.091	0.093	0.089	0.084	0.084	0.080	0.079	0.074	0.067	0.068	0.061	0.058	0.060	0.062	0.063	0.067	0.060
3c. Explained due to FEM	0.031	0.025	0.034	0.034	0.040	0.036	0.042	0.047	0.041	0.041	0.023	0.028	0.026	0.021	0.024	0.020	0.021	0.026
3d. Explained due to all job characteristics	0.032	0.037	0.033	0.028	0.022	0.028	0.021	0.017	0.019	0.018	0.011	0.009	0.009	0.012	0.014	0.011	0.013	0.006
3e. Explained due to selected job characteristics																		
Education & Training	0.003	0.003	0.002	0.002	0.001	0.002	0.001	0.001	0.001	-0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000
Computer	-0.006	-0.006	-0.004	-0.006	-0.006	-0.006	-0.004	-0.005	-0.004	-0.005	-0.003	-0.004	-0.003	-0.003	-0.003	-0.002	-0.002	-0.002
Physical	-0.015	-0.010	-0.010	-0.012	-0.013	-0.008	-0.012	-0.012	-0.014	-0.012	-0.010	-0.009	-0.012	-0.011	-0.011	-0.009	-0.008	-0.009
Part-time	0.015	0.018	0.018	0.015	0.015	0.016	0.016	0.014	0.016	0.016	0.009	0.010	0.009	0.013	0.011	0.010	0.011	0.008

Notes: The standard and expanded specifications are described in Table 2. Decompositions are performed using the *oaxaca* command in Stata 12.

Table 7: Gender Composition Coefficients from Panel Data Estimates, Wage Level and Wage Change Equations, CPS-MORG and NLSY79, 1993-2010

	Females			Males		
Data Source/ Model	Standard	Expanded	Expanded Plus	Standard	Expanded	Expanded Plus
CPS - MORG						
OLS	-0.190**	-0.087**	-	-0.226**	-0.139**	-
	[0.003]	[0.003]		[0.003]	[0.004]	
FD	-0.029**	0.009+	-	-0.050**	-0.018**	-
	[0.005]	[0.006]		[0.005]	[0.007]	
NLSY79						
OLS	-0.242**	-0.114**	-0.110**	-0.285**	-0.199**	-0.215**
	[0.017]	[0.021]	[0.020]	[0.020]	[0.026]	[0.025]
RE	-0.157**	-0.063**	-0.060**	-0.145**	-0.087**	-0.090**
	[0.016]	[0.021]	[0.020]	[0.020]	[0.026]	[0.026]
FE	-0.131**	-0.046*	-0.042+	-0.098**	-0.031	-0.027
	[0.017]	[0.022]	[0.022]	[0.021]	[0.028]	[0.027]
FD	-0.117**	-0.045+	-0.045+	-0.105**	-0.012	-0.002
	[0.020]	[0.026]	[0.026]	[0.025]	[0.034]	[0.034]

Notes: The CPS-MORG data contain 1,535,538 observations from 767,769 individuals. In the NLSY79 there are 32,717 observations from 4,974 individuals. For the NLSY79 log wage change (FD) regressions there are 21,619 observations. For these FD regressions we dropped the 1993 round in order to have a consistent measure of wage change over two-year intervals. In the "expanded plus" specification potential experience variables are replaced with actual labor market experience, tenure, and age variables. This latter specification also controls for the age- and education-adjusted ASVAB score. **,*, + denote statistical significance at 0.01, 0.05, and 0.10 levels, respectively.

Appendix Table 1: Means of Selected Variables by Gender Composition, for Males and Females in 1994 and 2010

	Females								Males							
	1994				2010				1994				2010			
Variable / Value of FEM	0-.25	.25-.50	.50-.75	.75-1.0	0-.25	.25-.50	.50-.75	.75-1.0	0-.25	.25-.50	.50-.75	.75-1.0	0-.25	.25-.50	.50-.75	.75-1.0
Wage	16.11	18.18	16.71	15.28	19.79	22.84	18.78	18.40	19.48	24.02	21.97	17.31	21.96	29.28	24.48	21.17
Schooling	12.83	13.41	13.80	13.41	13.59	14.06	14.01	14.06	12.66	13.81	14.32	13.82	12.91	14.20	14.34	14.60
Experience	19.58	20.26	19.04	19.76	22.59	23.23	20.38	23.33	20.54	19.77	18.22	16.39	23.21	22.64	19.62	19.78
Full-time	0.84	0.79	0.75	0.70	0.85	0.82	0.70	0.75	0.93	0.89	0.86	0.75	0.92	0.88	0.81	0.81
Federal	0.04	0.06	0.02	0.02	0.04	0.04	0.02	0.02	0.03	0.06	0.03	0.04	0.03	0.04	0.04	0.05
State	0.05	0.05	0.06	0.06	0.06	0.06	0.05	0.08	0.03	0.04	0.07	0.08	0.03	0.05	0.05	0.10
Local	0.08	0.06	0.12	0.13	0.09	0.06	0.08	0.16	0.07	0.04	0.12	0.11	0.08	0.05	0.06	0.16
Black	0.14	0.10	0.10	0.11	0.15	0.11	0.12	0.14	0.08	0.08	0.08	0.12	0.10	0.09	0.10	0.16
Union member/covered	0.21	0.13	0.15	0.16	0.14	0.10	0.09	0.17	0.25	0.15	0.17	0.20	0.17	0.10	0.10	0.22
OPTD education & training	8.80	7.65	7.60	9.10	8.40	7.35	7.99	8.60	8.90	7.72	7.33	9.52	8.71	7.47	7.92	8.50
Occupation tenure	2.93	2.89	3.07	2.72	2.67	3.11	2.52	3.04	2.99	2.98	3.30	2.38	2.89	3.24	2.55	3.04
Occupation part-time	0.09	0.15	0.21	0.29	0.10	0.15	0.27	0.25	0.07	0.14	0.19	0.30	0.09	0.14	0.25	0.23
O*NET computer	2.44	2.64	2.80	2.72	2.77	2.73	2.88	2.57	2.17	2.69	2.88	2.71	2.29	2.77	2.85	2.59
O*NET environment	2.75	2.35	2.25	2.15	2.72	2.37	2.23	2.17	2.88	2.38	2.24	2.17	2.87	2.38	2.23	2.16
O*NET hazard	2.15	1.67	1.54	1.56	2.04	1.64	1.47	1.64	2.36	1.69	1.52	1.54	2.31	1.66	1.47	1.60
O*NET physical	2.63	2.44	2.37	2.43	2.55	2.40	2.42	2.37	2.77	2.45	2.35	2.49	2.73	2.42	2.40	2.35
O*NET strength	1.93	1.16	1.02	1.00	1.82	1.12	1.02	1.06	2.14	1.16	0.99	1.08	2.09	1.10	1.01	1.11
Industry union	0.19	0.17	0.16	0.17	0.15	0.12	0.11	0.17	0.20	0.18	0.17	0.19	0.14	0.12	0.11	0.19
Industry big firm	0.40	0.44	0.41	0.40	0.41	0.45	0.42	0.39	0.34	0.42	0.40	0.44	0.33	0.42	0.42	0.43
N	3,369	15,712	18,214	39,668	3,860	13,833	27,295	32,043	35,392	26,692	11,456	5,374	33,486	22,607	15,843	4,022

Notes: Weekly wages are expressed in December 2010 dollars. OPTD education & training is a 1 to 11 index of the education and training requirements of an occupation, 1 being the highest and reserved for jobs requiring professional degrees. It is obtained from the 2002 data in SOC 2000 codes that are then mapped into 2000 census occupation codes. The proportion of workers in big firms (having more than 1000 employees) was calculated from the 2003-2007 CPS Annual Social and Economic (ASEC) supplement and average job tenure for each occupation was generated using 2004-2010 CPS job tenure supplement.. Annual levels of union membership for each industry were calculated from CPS data, as were occupational part-time employment shares. The O*NET extract is from 2008 data and occupations are classified according to 2000 census occupational codes. Details on the creation of the O*NET extract are provided in Hirsch and Schumacher (2012).

Appendix Table 2: Mean Wages, the Wage Gap, Gender Composition, and the Wage-Gender Composition Relationship by Year, NLSY79, 1993-2010

	Female			Male			Female to Male Wage	Duncan Segregation		
Year	N	Wage	FEM	N	Wage	FEM	Ratio	Index	θ_f	θ_m
1993	1,869	15.85	0.678	1,901	20.764	0.310	0.763	0.540	-0.248**	0.109*
1994	1,734	16.70	0.667	1,802	20.798	0.311	0.803	0.539	-0.243**	0.010
1996	1,874	17.07	0.665	1,778	22.569	0.307	0.756	0.532	-0.203**	0.243**
1998	1,927	17.31	0.664	1,771	24.534	0.306	0.706	0.526	-0.298**	0.210**
2000	1,852	18.43	0.654	1,700	25.871	0.304	0.712	0.517	-0.223**	0.178**
2002	1,727	19.40	0.653	1,613	27.508	0.284	0.705	0.516	-0.138*	0.253**
2004	1,538	19.37	0.654	1,490	27.709	0.281	0.699	0.526	-0.210**	0.328**
2006	1,201	19.23	0.670	1,249	26.288	0.282	0.732	0.522	-0.186**	0.241**
2008	1,556	19.11	0.673	1,433	28.389	0.294	0.673	0.522	-0.139*	0.281**
2010	1,554	20.39	0.666	1,388	29.174	0.293	0.699	0.509	-0.194**	0.204**

Notes: See notes to Table 1. Values of the Duncan index of segregation are from the CPS-MORG samples. The corresponding index values constructed from the NLSY79 are generally a few percentage points higher, possibly reflecting the ages of our cohort. Standard errors are in brackets. **,* denote statistical significance at the 0.01 and 0.05 levels, respectively.

Appendix Table 3: Means of Selected Variables by Gender Composition, Males and Females, NLSY79, 1993-2010

	Females				Males			
Variable / Value of FEM	0-.25	.25-.50	.50-.75	.75-1.0	0-.25	.25-.50	.50-.75	.75-1.0
Wage	18.202	19.901	19.529	16.396	21.275	28.447	29.391	18.630
Schooling	13.417	13.749	14.254	14.048	12.750	14.302	15.082	14.837
Age	39.491	39.848	40.520	39.850	39.674	40.020	39.725	38.515
Tenure at current job	5.867	6.410	6.354	5.859	6.900	7.227	6.684	5.321
Prior Experience	9.925	10.138	11.035	10.474	11.392	11.556	11.885	11.466
Full-time	0.844	0.835	0.765	0.703	0.958	0.954	0.925	0.860
Public Sector	0.125	0.148	0.160	0.241	0.126	0.098	0.159	0.191
Black	0.120	0.136	0.115	0.136	0.115	0.122	0.094	0.167
Union member/covered	0.210	0.149	0.138	0.196	0.257	0.166	0.161	0.253
OPTD education & training	8.582	7.859	7.492	8.912	8.920	7.733	7.022	8.943
Occupation tenure	3.172	3.112	3.026	2.831	3.103	3.321	3.250	2.671
Occupation part-time	0.077	0.118	0.194	0.263	0.064	0.102	0.167	0.245
O*NET computer	2.387	2.807	2.898	2.693	2.148	2.902	3.011	2.645
O*NET environment	2.812	2.377	2.238	2.161	2.944	2.392	2.256	2.190
O*NET hazard	2.236	1.693	1.482	1.586	2.433	1.708	1.471	1.604
O*NET physical	2.637	2.468	2.332	2.403	2.791	2.439	2.283	2.426
O*NET strength	1.941	1.217	0.905	1.011	2.217	1.208	0.861	1.105
Industry union	0.160	0.143	0.141	0.173	0.184	0.141	0.141	0.178
Industry big firm	0.368	0.422	0.412	0.400	0.317	0.400	0.411	0.428
N	823	3,353	5,320	7,336	7,026	5,292	3,088	719

Notes: See notes to Appendix Table 1. As the NLSY79 has an aging cohort, variable means are taken over the entire panel.

Appendix Table 4: Gender Composition Coefficients by Specification, Gender, and Year, Wage Level Equations, NLSY79 1993-2010

Specification	1993	1994	1996	1998	2000	2002	2004	2006	2008	2010	ALL
Females:											
No Controls	-0.248**	-0.243**	-0.203**	-0.298**	-0.223**	-0.138*	-0.210**	-0.186**	-0.139*	-0.194**	-0.215**
	[0.051]	[0.054]	[0.054]	[0.052]	[0.054]	[0.058]	[0.060]	[0.072]	[0.061]	[0.064]	[0.018]
Standard	-0.242**	-0.283**	-0.231**	-0.238**	-0.261**	-0.179**	-0.234**	-0.242**	-0.192**	-0.191**	-0.242**
	[0.051]	[0.054]	[0.054]	[0.049]	[0.051]	[0.058]	[0.062]	[0.076]	[0.063]	[0.069]	[0.017]
Expanded	-0.089	-0.122+	-0.145*	-0.025	-0.097	-0.124+	-0.065	-0.083	-0.07	-0.146+	-0.114**
	[0.064]	[0.067]	[0.068]	[0.060]	[0.062]	[0.069]	[0.073]	[0.090]	[0.072]	[0.079]	[0.021]
Expanded Plus	-0.096	-0.153*	-0.148*	-0.036	-0.097+	-0.105	-0.086	-0.065	-0.106	-0.147*	-0.110**
	[0.060]	[0.064]	[0.064]	[0.056]	[0.059]	[0.066]	[0.069]	[0.086]	[0.067]	[0.074]	[0.020]
N	1,869	1,734	1,874	1,927	1,852	1,727	1,538	1,201	1,545	1,554	16,832
Males:											
No Controls	0.109*	0.01	0.243**	0.210**	0.178**	0.253**	0.328**	0.241**	0.281**	0.204**	0.180**
	[0.051]	[0.053]	[0.055]	[0.057]	[0.061]	[0.068]	[0.069]	[0.074]	[0.073]	[0.073]	[0.020]
Standard	-0.261**	-0.329**	-0.283**	-0.254**	-0.383**	-0.399**	-0.119	-0.171*	-0.341**	-0.472**	-0.285**
	[0.057]	[0.058]	[0.064]	[0.060]	[0.066]	[0.074]	[0.077]	[0.081]	[0.081]	[0.084]	[0.020]
Expanded	-0.151*	-0.303**	-0.164*	-0.111	-0.232**	-0.329**	-0.074	-0.128	-0.214*	-0.358**	-0.199**
	[0.073]	[0.074]	[0.080]	[0.075]	[0.083]	[0.089]	[0.091]	[0.097]	[0.095]	[0.098]	[0.026]
Expanded Plus	-0.163*	-0.317**	-0.183*	-0.145*	-0.256**	-0.326**	-0.086	-0.143	-0.242**	-0.372**	-0.215**
	[0.070]	[0.072]	[0.077]	[0.072]	[0.080]	[0.087]	[0.089]	[0.094]	[0.091]	[0.096]	[0.025]
N	1,901	1,802	1,778	1,771	1,700	1,613	1,490	1,249	1,433	1,388	16,125

Notes: See notes to Table 2. **,*, + denote statistical significance at 0.01, 0.05, and 0.1 levels, respectively.