Coworker Gender Preferences: Effects on Gender Gaps in Occupational Selection and Wages

Clint Harris

October 25, 2018

Abstract

This paper analyzes the effect of occupational gender composition on job-specific labor supply for workers of each gender. I construct a static model of job selection wherein preferences regarding coworker gender composition produce gender-specific compensating differentials. I estimate the model using maximum likelihood to identify the underlying coworker gender preference parameters. Based on estimated compensating differentials, men's preference is highest for occupations that are 60% female and lowest for female-dominated occupations. Women prefer jobs that are female-dominated, and are least satisfied with jobs that are 25% male, all else equal.

1 Introduction

Occupational selection is a major factor in a wide array of differential outcomes for men and women. Not only is it a major factor in gaps in pay between men and women, but it also has implications for productivity and job satisfaction. Given these important implications, it is valuable to identify the various factors that drive men and women to choose different jobs. I contribute to this effort by investigating the role that the gender composition of an occupation plays in attracting workers of each gender. I obtain nonlinear preferences for men and women over the gender composition of jobs by estimating gender-specific compensating differentials for the share of a job that is female.

This question is of particular interest not only because it considers another factor that explains differential job satisfaction and selection by gender but because this factor produces externalities by necessity and therefore has serious implications for welfare. When a person selects an occupation due in part to its gender composition, they affect its gender composition.

This necessarily affects the favorability of the job for themselves and all others in the job (except in the specific case where the job is already entirely dominated by their gender). The individual may take into account their own effect on the gender composition of the job when making their occupational decision, but will not generally internalize their impact on the attractiveness of the occupation to others.

Because I allow for preferences over the gender composition of jobs to vary nonlinearly and even nonmonotonically, my strategy allows for the possibility of multiple equilibria in the gender composition of jobs. For instance, I find that in male-dominated occupations, marginal increases in the female share of the occupation increase men's compensating differentials. This means that the firm's marginal cost of hiring a woman is not only given by her wage, but by the increase in men's equilibrium wages times the number of men currently working in the firm. Hiring a large numbe of women, however, reduces men's compensating differential while also reducing the number of men in the occupation, which causes a much smaller increase in firm costs per female hire. The policy implication of this is that firms may avoid marginal increases in female representation in jobs because they are at a local minimum of costs, and policies that shift large numbers of women into male-dominated jobs may achieve a different, and possibly welfare-improving, equilibrium. Any policy that induces such a shift would achieve its desired effect even if it is only tmeporary.

Table 1: Descriptive Statistics of Male and Female Wages

Share of Occupation Female	**variable**	**N**	**mean**	**sd**
0-25%	Share Female	49343	.1159596	.087984
	Mean Male Wage	44307	20.10564	11.25329
	Mean Female Wage	5036	20.99102	10.87163
25-50%	Share Female	143742	.357084	.0637713
	Mean Male Wage	97130	26.69123	11.55262
	Mean Female Wage	46612	19.7498	9.898433
50-75%	Share Female	27322	.6197304	.0769729
	Mean Male Wage	11534	21.92464	11.69915
	Mean Female Wage	15788	16.67704	8.212288
75-100%	Share Female	55772	.905058	.0705397
	Mean Male Wage	5892	19.23069	10.1875
	Mean Female Wage	49880	15.81967	6.902736
Total	Share Female	276179	.4506459	.2723732
	Mean Male Wage	158863	24.23173	11.85176
	Mean Female Wage	117316	17.71856	8.781421

Note: The unit of observation is an individual. Observations are on full-time, full-year workers with reported wages in job/state/year combinations with at least 50 observations from 1990-2015 in the Current Population Survey.

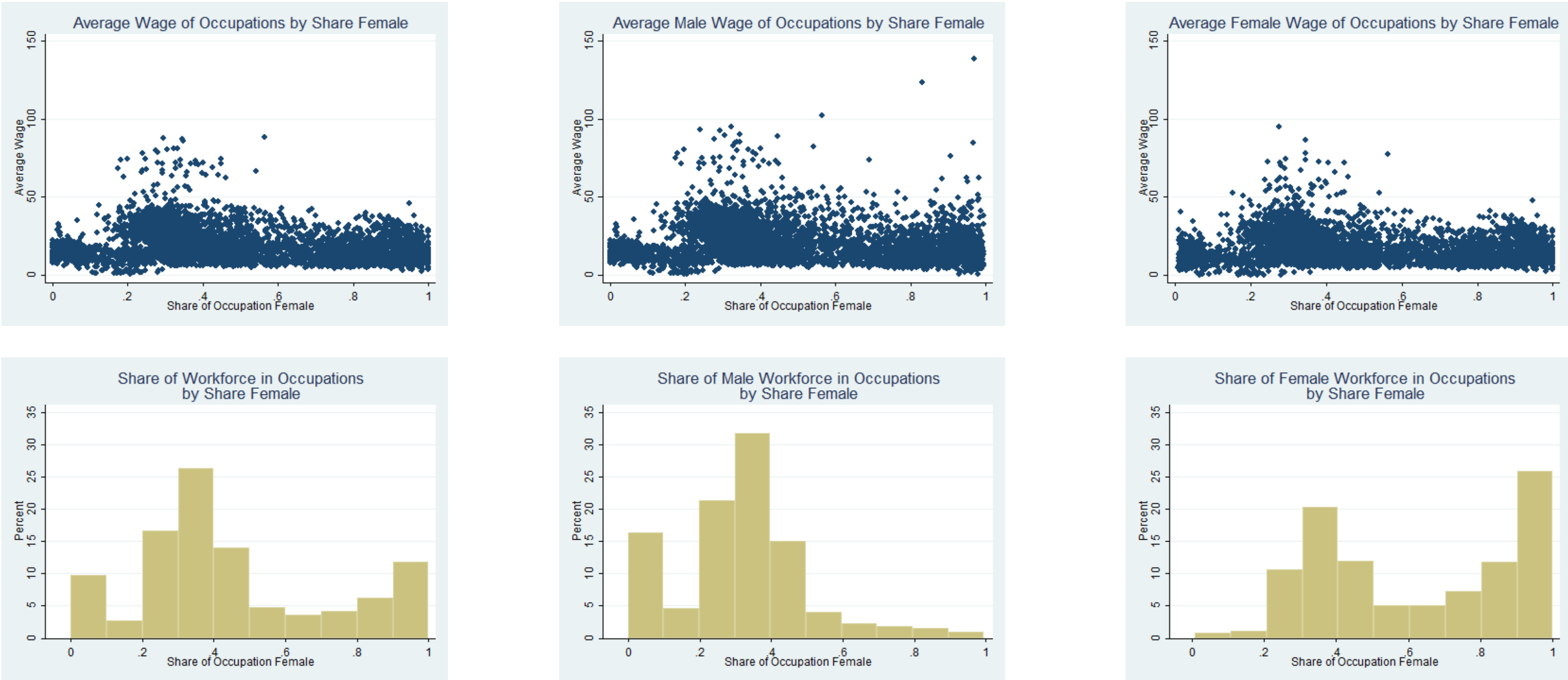

Figure 1: Average wages and population densities by gender according to share of women in an occupation in a given state and year. Occupations are defined at the 3-digit occ1990 occupational coding scheme available in the census. This figure plots average wages for full-time, full-year workers in occupations with at least 50 individuals by the share of the occupation that is female as well as the distribution of the workforce by the share of the occupation that is female. All data is from the CPS, 1990-2015.

Differences between minority and nonminority workers in terms of employment and wages has previously been explored through various mechanisms, including the preferences of employers, consumers, and coworkers. These works present models that are generally applicable but are overwhelmingly implemented in the investigation of differences based on either race or gender (which we consider here). The seminal work on the subject is that of Becker (1957). In Becker, firms receive utility from profits and lose utility from hiring black workers. In order to be indifferent between hiring whites and blacks, the firm provides a lower wage to blacks with the differential being exactly equal to the disutility received from hiring the black worker. The model predicts that profit is decreasing in the strength of the racial prejudice and due to entry and exit, the long-run equilibrium allows for only the most productive prejudiced firms to survive (they will hire only whites) and non-prejudiced firms will hire whites and blacks with a single wage. More broadly speaking, if there is prejudicial distaste by some employers, consumers, or coworkers, the market should allow prejudiced firms, consumers, and workers to interact exclusively with white-only versions of each other, while the rest of the market (blacks and non-prejudiced whites) interacts only internally as well. The essential outcome is one of highly segregated markets with no wage differentials unless segregation is impossible due to the numbers of minorities and firms in the market (in which case a wage gap can persist).

In our formulation, we allow for the possibility of individuals still interacting with other types if their idiosyncratic taste for a particular job is sufficiently high. This allows for less extreme outcomes than are predicted by the Becker model; segregation will be incomplete and we can maintain wage gaps with widespread weak own-type preferences. Additionally, we frame individual preferences in terms of own-group vs. out-group preference rather than a strict majority on minority prejudice. This allows us to consider the effects not only of the majority having distaste for working with minorities, but vice versa (in either case due either to actual prejudice or homophily).

Along very similar lines to worker preferences regarding their coworkers, we have Goldin (2014), which refers to the possibility of a job being "polluted" by the presence of a particular type of worker. Here, workers receive utility both from some wage that they earn at a job and from a sense of prestige that is associated with the perceived difficulty of the job by society. The model proceeds in 2 stages. In the first stage, men are alone in the market and work various jobs. In stage 2, men stay in the same job and women enter the job market. Each job requires some amount of some trait C, such as physical strength, in order to work at a given firm, and the worker's pay is proportionate to their level of C. Importantly, at the beginning of stage 2 there is a random technology shock to firms where the amount of the trait required to work in the firm either decreases or does not. Jobs that had a high C in stage 1 may now have a low C, but will retain the prestige associated with the stage 1 value if society is unable to determine that the amount of the trait needed to succeed in the job has diminished. If women are believed by society to have lower average values of C than was required to work such a job in stage 1, and this job hires a sufficient number of women, society will infer that the technological change has occurred in this industry and the job is no longer as difficult as it once was (even if this has not happened). This reduces the prestige of the job for the incumbent male workers. Because of this, these men will demand a wage premium to compensate them for their lost prestige over the women (who are modeled as not caring about prestige). Alternatively, firms can create a "different" occupation for women that is effectively the same as that of men, allowing the men to retain their prestige while still hiring women. The first case produces a wage gap and the second produces occupational segregation.

It is important to consider the difference between this notion of occupational pollution and the type of coworker preferences we are investigating. Here, the two types of workers do not mind working with one another, they simply dislike having their prestige reduced. In the scenario we investigate, the two groups actually get disutility from interacting with one another (perhaps due to outright harassment or more mild mechanisms such as restrictions in what types of opinions are considered acceptable to voice, which plausibly varies by how much one type dominates the social setting). It is plausible that both mechanisms are in effect, but discerning the magnitude of each effect may be difficult. It is plausible that the pollution effects may matter more for occupational choice, while coworker preferences may more strongly affect selection into industries. Comparing industrial segregation to occupational segregation (accounting for the expected correlation between the two) may provide insight into the relative potency of each of these mechanisms.

A related paper by **?** examines male and female preferences for jobs in a framework which also borrows from literature on the tipping phenomenon in housing markets. She investigates occupations in the U.S. economy and is able to identify tipping points wherein after a certain proportion of an occupation's workers become female, male growth in the occupation becomes negative and the job becomes heavily female-dominated. She uses IPUMS from 1940-1990 and is able to analyze changes in occupations in terms of gender representation. She shows that firms with 25%-45% female labor forces begin to have negative net male employment growth in white collar occupations, while the tipping point varies from 13%-30% for blue collar occupations. Additionally, she finds that tipping occurs sooner in regions where men hold more sexist attitudes toward women. This may also explain the lower tipping points in blue collar occupations vs. white collar occupations if blue collar men have higher rates of such opinions than white collar men. Such an analysis of changes in male and female representation in jobs is obviously related to our question, though we will attempt to identify compensating differentials instead of tipping points.

Sasaki (1999) produces a search model that is qualitatively similar to mine in that coworker gender preferences are directly modeled within the utility function. Broadly speaking, this model's qualitative predictions under an assumption of male distaste for female coworkers mirror our own in finding increased female participation produces higher female wages and lower unemployment, while increased distaste for female coworkers among males reduces female earnings and employment. Notable differences are present as well. Sasaki imposes that men have distaste for women without providing the opportunity for women to have symmetric preferences. Additionally, this model is somewhat more extreme in assuming that the nature of males' preferences is that if a firm hires any women at all, the male get a constant disutility from working in that firm. Our formulation allows for a flexible utility term that allows for the utility for each additional other-typed coworker to be convex, concave, and/or nonmonotonic. This allows us to more closely match the data to explain the tipping behavior observed empirically in the proportion of male workers in occupations, as well as selection and gender gaps in jobs.

I model labor market outcomes with preferences regarding the distribution of coworkers in a Roy model. If an individual has to choose between occupations in which to work, their reservation wage in each occupation will depend in part on disutility from working with each type of worker. Thus, if an individual is otherwise indifferent between two firms, but one has an even slightly higher proportion of their own type of worker (through random chance or through systematic correlation of type and idiosyncratic job preference), this individual will strictly prefer this firm over the other if they prefer their own type. In equilibrium, this can be expected to predict substantial segregation due to shifts in job selection by workers away from the hypothetical equilibrium without any homophilic preferences.

2 Model

2.1 Workers

Throughout the following, the index i will refer to types of individuals (male or female) while j will refer to jobs $(1, ..., N)$. Workers receive linear utility from wages w_{ij}, the proportion of female coworkers in a job ψ_j according to type-specific tastes $g_i(\psi_j)$, and an i.i.d. random additive job-specific nonpecuniary benefit $\alpha_{ij} \sim F_{ij}(\alpha)$. These are all common knowledge for all individuals and jobs. The utility function is thus given by:

$$u_{ij} = w_{ij} + g_i(\psi_j) + \alpha_{ij}, \quad i = m, f \quad \& \quad j = 1, 2, ..., N. \tag{1}$$

The condition for an individual of type i choosing job j over all other jobs is thus:

$$\begin{aligned} u_{ij} &\geq u_{ik}, \quad \forall k \neq j \\ w_{ij} + g_i(\psi_j) + \alpha_{ij} &\geq w_{ik} + g(\psi_k) + \alpha_{ik}, \quad \forall k \neq j \end{aligned} \tag{2}$$

Here we impose that each person of type i's draw of α_{ij} for each job comes from an i.i.d. Type I Extreme Value distribution with CDF:

$$F_{ij}(\alpha_1, ..., \alpha_N) = \exp\Big(-\sum_{j=1}^{N} e^{\mu_{ij} - \alpha_j} \Big). \tag{3}$$

The mode of the distribution, μ_{ij}, is allowed to vary both between firms and between types. This allows people to view a job more or less favorably based on their gender. For instance, if men have less distaste for particularly unpleasant jobs than women, $\mu_{mj} > \mu_{fj}$ will allow this to enter into the model. The Extreme Value distribution allows for an attractive analytical solution for the proportion of individuals of type i that will choose job j, which we will denote λ_{ij}. This proportion is given by:

$$\begin{aligned} \lambda_{ij} &= \Pr[w_{ij} + g_i(\psi_j) + \alpha_j > w_{ik} + g_i(\psi_k) + \alpha_k \quad \forall k \neq j] \\ &= \Pr[\alpha_k < w_{ij} - w_{ik} + g_i(\psi_j) - g_i(\psi_k) + \alpha_j \quad \forall k \neq j] \\ &= \int F_j[\alpha_j + \epsilon_1, ..., \alpha_j + \epsilon_N] d\alpha_j \end{aligned} \tag{4}$$

where $F_j(\cdot)$ is the derivative of the cdf with respect to it's jth term and $\epsilon_k \equiv w_{ij} - w_{ik} + g_i(\psi_j) - g_i(\psi_k)$. Evaluating this integral provides the following for the proportion of type i that will work in job j:

$$\lambda_{ij} = \frac{e^{w_{ij}+g_i(\psi_j)+\mu_{ij}}}{\sum_k e^{w_{ik}+g_i(\psi_k)+\mu_{ik}}} \tag{5}$$

Which we can rearrange to obtain the inverse labor supply:

$$w_{ij} = \ln\Big(\frac{\lambda_{ij}}{1-\lambda_{ij}}\Big) + \ln\Big[\sum_{k\neq j} e^{w_{ik}+g_i(\psi_k)+\mu_{ik}}\Big] - \Big[g_i(\psi_j) + \mu_{ij}\Big] \tag{6}$$

Looking at each term sequentially, we can see that the wage is a function of the proportion of type i that the firm hires (upward-sloping supply), the attractiveness of other firms (the functional form emphasizes the most attractive alternative), and the compensating variation both for the individual's preference for their coworkers and for the mode of their random preference for the job.

It is worth noting here that $g_i(\cdot)$ could be any function (individuals have knowledge of it, but we do not). In applications, we want to parameterize it flexibly enough that it allows for several possibilities. If $g_i(\cdot)$ is convex, individuals are largely unaffected by a few individuals like themselves, but will be strongly affected by many. Concavity naturally suggests the opposite. Nonmonotonic $g_i(\cdot)$ is consistent with either preference for diversity ($g_i(\cdot)$ increases then decreases with $\psi_j \in [0,1]$) or a preference to avoid cross-type tensions ($g_i(\cdot)$ decreases then increases with $\psi_j \in [0,1]$). A mixture of these types of preferences would naturally lead to a complex functional form, and I know of nothing in the literature that would make any strong prediction between the above possibilities.

Depending on the functional form of $g_i(\cdot)$, there could be multiple solutions for $\{\psi_j\}$ in the model. In order to obtain unique solutions, we use a sequential entry concept where men start out in the labor force (in their preferred job given $\psi_{mj} = 1$ for all jobs in the labor market) while women start out in the domestic sector. Arbitrarily setting the domestic sector to be j=1 provides initial conditions for female job choice that $\psi_{f1} = 1$, and $\psi_{fk} = 0\ \forall k \neq 1$. The infinitesimal mass of women with the highest utility for jobs will enter those jobs first, and the value of ψ_{fj} will update accordingly for all jobs. This process is repeated until an equilibrium value for ψ_{fj} is reached $\forall j$. Sequential entry implies that this first equilibrium will occur at the maximum value of the solution for ψ_{f1}, so this condition will designate the unique equilibrium. The potential for multiple equilibria provides a possible opportunity for welfare improvements in the event of a policy that can shift the allocation of labor to a new equilibrium with higher utility outcomes for workers.

2.2 Firms

Firms are perfectly competitive profit maximizers. Each firm j hires a mass η_{ij} of type i workers from a measure 1 mass of available workers, such that $\eta_j = \sum_i \eta_{ij}$ is the total mass of workers employed by firm j and $\sum_j \eta_j = 1$ is the total labor force. I assume that firms have CES production over male and female labor:

$$F_j(\eta_{mj}, \eta_{fj}) = \left(\gamma_{mj}\eta_{mj}^{\frac{\epsilon-1}{\epsilon}} + \gamma_{fj}\eta_{fj}^{\frac{\epsilon-1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}} \tag{7}$$

where γ_{ij} governs occupation-gender-specific productivity and ϵ is the elasticity of substitution between male and female labor which is common to all firms. Their problem is then to maximize profit:

$$\pi_j = \max_{\eta_{mj},\eta_{fj}} (\gamma_{mj}\eta_{mj}^{1-\epsilon} + \gamma_{fj}\eta_{fj}^{1-\epsilon})^{\frac{1}{1-\epsilon}} - w_{mj}\eta_{mj} - w_{fj}\eta_{fj} \tag{8}$$

The first-order conditions from this problem provide the wage equation for type i at firm j as follows:

$$w_{ij} = \gamma_{ij}\eta_{ij}^{-\frac{1}{\epsilon}}\left(\gamma_{mj}\eta_{mj}^{\frac{\epsilon-1}{\epsilon}} + \gamma_{fj}\eta_{fj}^{\frac{\epsilon-1}{\epsilon}}\right) \quad \forall i = m, f \tag{9}$$

This provides a channel whereby the firm can offer different wages to different genders, which is necessary for the emergence of a compensating differential. We could produce such a channel with perfect substitution if we were to assume monopolistic competition among firms (wherein firms would internalize workers' gender composition preferences) but the CES-perfect competition assumption has the advantage of being both more intuitive and more tractable.

2.3 Equilibrium

Given prices $\{w_{ij}\}$, an equilibrium is an allocation $\{\eta_{ij}, \psi_j, \lambda_{ij}\}$ in which workers choose their preferred job, firms maximize profit, and markets clear[1] such that we have:

$$w_{ij} = \ln\left(\frac{\lambda_{ij}}{1-\lambda_{ij}}\right) + \ln\left[\sum_{k\neq j} e^{w_{ik}+g_i(\psi_k)+\mu_{ik}}\right] - \left[g_i(\psi_j) + \mu_{ij}\right] \tag{10}$$

$$w_{ij} = \gamma_{ij}\eta_{ij}^{-\frac{1}{\epsilon}}\left(\gamma_{mj}\eta_{mj}^{\frac{\epsilon-1}{\epsilon}} + \gamma_{fj}\eta_{fj}^{\frac{\epsilon-1}{\epsilon}}\right) \tag{11}$$

[1] The market clearing equations are a simple application of Bayes' Rule where we use: $Pr(j|i) = \dfrac{Pr(i|j)Pr(j)}{Pr(i)}$.

$$\lambda_{fj} = \frac{\psi_j \eta_j}{S_f}, \quad \& \quad \lambda_{mj} = \frac{(1-\psi_j)\eta_j}{S_m} \tag{12}$$

where $S_i = \sum_j \eta_{ij}$ is the proportion of the population that is gender i. Furthermore, we make use of the following identities:

$$\eta_j \equiv \sum_i \eta_{ij}, \quad \sum_j \eta_j \equiv 1, \quad \& \quad \eta_{ij} = S_i \lambda_{ij} \tag{13}$$

which state, respectively, that the sum of workers of each gender in a job is equal to the total mass of labor in that job, that the sum of all workers in all jobs is equal to the total labor force, and that the mass of workers of type i in job j is equal to the proportion of type i in the labor force times the probability of a person of type i selecting job j.

We can solve the above system to obtain the percentage of each type that chooses each occupation $\{\lambda_{ij}\}$, the percentage of each occupation that is each type $\{\psi_j\}$, wages by type by occupation $\{w_{ij}\}$, the proportion of each type that is in each occupation $\{\eta_{ij}\}$, and the proportion of the total population that is in each occupation $\{\eta_j\}$.

3 Data

We use data from the 2010-2015 U.S. Census available through IPUMS to determine the magnitude of gender-specific compensating differentials predicted by the model. Census data is ideal because it provides data over time and regions, allowing us to effectively control for differences by state and year while still using state-year-job variation for identification. Additionally, the large number of observations available in this data set allow us to precisely estimate higher ordered polynomials of the gender preferences, which is necessary to pick up important information about the shape of the gender-preference function.

There is a question regarding the correct designation of “jobs” in the data, depending on the exact nature of gender preferences. Preferences may manifest differently at the firm, occupation, and industry levels. Workers may care about how masculine or feminine a job is perceived to be, which likely would occur less at the firm level and more at the occupation or industry level. However, workers may also care about interactions with coworkers, in which case firm-level data may provide more insight (though occupations within firms may also determine the duration, intensity, and frequency of interactions, so occupation remains a factor). Workers may choose occupations based on their expected interactions in firms, but may not know the gender composition of particular firms. My primary analysis focuses on occupations since I expect individuals to exercise substantially more control over their choice of occupation than the industry, and I expect them to have more information about occupations than firms. We make use of 3 occupational classifications provided through the Census (with 7, 14, and 46 job categories, specifically). A 3-digit designation is available with 389 occupational designations, but this many occupations is computationally infeasible given my empirical strategy.

To obtain a consistent job classification across years, I construct an occupational designation variable using the occly variable available through IPUMS (which provides an individual's 3-digit occupation for the prior year). This variable does not use a consistent coding scheme across years, so I make use of a crosswalk provided by IPUMS to convert the occly variable into a consistent coding scheme for last year occupations that is analogous to the current year occ1990 classification provided by IPUMS. I additionally construct variables governing the proportion of own-type workers who select a given job (λ_{ij}) and the proportion of workers in a given job who are female (ψ_j). Both variables are constructed using weighted observations of each state-year-job combinations at each level of occupational aggregation (1-digit, 2-digit, and 3-digit).

Because my identification relies on the assumption of a stable preference function $g(\cdot)$, it is important to only include years where this assumption is likely to hold. Because attitudes about gender roles regarding the workforce are constantly evolving, I balance the need for cross-year variation with the need for a stable preference function by using data from 2010 to 2015. In the future I will perform robustness checks on other year ranges.

4 Empirical Strategy

In order to identify the magnitude of gender preferences, I identify the compensating differentials obtained by each gender as a result of the gender composition of their job as predicted by the model. There are substantial challenges with this as the gender shares of occupations and industries are highly correlated with other aspects of the job such as physical, cognitive, and social demands, required education, hours worked, and so on. While prior research has documented that women make lower wages than men even within the same job (Blau and Kahn, 2016), there is room for serious consideration regarding the level of heterogeneity in the actual job content of the "same job" for different workers of that job, even for finely aggregated job classifications.

My empirical design avoids the pitfalls of comparing men and women in the same job category by splitting the sample by gender prior to the analysis. The relevant empirical distinction between this topic and much of the related literature is that here there is no need to compare women to men. I am only comparing people of each gender to other members of their own gender who work in the same job with different gender compositions. This means that if a particular job systematically pays women or men different wages for any reason (at any level of aggregation), I can control for this with a simple job fixed effect. This has the effect of disregarding a large amount of variation that is undoubtedly important for determining why men and women receive different wages in a given job, but is actually unrelated to the compensating differential for gender composition.

The equation I want to estimate is a parameterized version of equation (6) from the model:

$$w_{ij} = \ln\Big(\frac{\lambda_{ij}}{1-\lambda_{ij}}\Big) + \ln\Big[\sum_{k\neq j} \theta_k e^{w_{ik}+g_i(\psi_k)+\mu_{ik}\beta_k}\Big] - g_i(\psi_j) - \mu_{ij}\beta_j \tag{14}$$

I model $g_i(\psi)$ as a fourth-order polynomial, wherein ψ is defined at the state-year-job level. The error term μ is a set of controls including age, age-squared, full-time status, education, race, and fully-interacted state, year, and occupation fixed effects (with state-year-occupation fixed effects omitted to preserve identifying variation), and with β included as the vector of coefficients on these elements. For an individual's own job, their individual values are used in these controls while state-year specific occupational averages are used for other jobs. I vary the designation of jobs j and k by specification and weight outside option jobs by θ, which is consistent with the presence of multiple identical jobs in the more general model.[2]

The nonlinear nature of the equation of interest is problematic for direct estimation. We therefore estimate the following approximation[3]:

$$w_{ij} \approx \ln\Big(\frac{\lambda_{ij}}{1-\lambda_{ij}}\Big) + w_{i\ell} + g_i(\psi_\ell) + \mu_{i\ell}\beta_\ell + ln(\theta_\ell) - g_i(\psi_j) - \mu_{ij}\beta_j \tag{15}$$

where job ℓ is one other job. This approximation is exact when there are only two jobs in the economy and does well when job ℓ is the most relevant outside option, as measured by both

[2] This is easiest to see in equation 5. Imagine that among all jobs j, exactly θ_ℓ are type ℓ (and are identical). This would result in the denominator reading $\sum_{k\neq\ell} exp[w_{ik} + g_i(\psi_k) + \mu_{ik}] + \theta_\ell exp[w_{i\ell} + g_i(\psi_\ell) + \mu_{i\ell}]$.

[3] I use $ln(1+x) \approx x$ where $x = \frac{\sum_{k\neq j,\ell} \theta_k exp[w_{ik}+g_i(\psi_k)+\mu_{ik}\beta_k]}{\theta_\ell exp[w_{i\ell}+g_i(\psi_\ell)+\mu_{i\ell}\beta_\ell]} \approx 0$. It is clear that this term approaches zero when job ℓ is more attractive on average (the term in the exponent) and/or more common in the economy (θ_ℓ) than other jobs.

attractiveness and prevalence in the economy. In practice, I designate this outside option as the most common job other than an individual's own job by gender, state, and year. This approximation will only bias my estimates of $g(\cdot)$ if traits of other outside options (which are now left in the error term) are correlated with both wages and own-job gender composition in a way that isn't captured by the included outside option or other controls. Results from this specification are in table 2, and a visual representation is provided in figure 2.

Table 2: The Effect of Share of Occupation Female on Wages by Gender

	(1) Male	(2) Female
ψ_j	75.77	30.40
	(54.91)	(25.79)
ψ_j^2	-454.7.0	-93.06
	(315.4)	(84.12)
ψ_j^3	759.7	104.1
	(496.8)	(106.4)
ψ_j^4	-394.2	-39.05
	(247.0)	(46.19)
N	246219	219556
R^2	0.052	0.041

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: CPS 2010-2015, all employed adults. ψ is the share of an occupation that is female. See figure 2 for a graphical representation of the implied compensating differential from these results.

Table 2 shows that gender preferences have an insignificant effect on wages with economically relevant point estimates. It is possible that the inclusion of additional years in our dataset or the use of a finer occupational classification will produce findings that are significantly different from zero. I look forward to exploring these additional specifications in future versions of this paper. Because the coefficients on the quartic polynomial are difficult to interpret, I provide a visual representation of the implied compensating differentials in figure 2.

The results show that men make lower wages in female dominated occupations conditional on the controls included in the specification. This is consistent with men preferring jobs with many female coworkers. Because I include occupation-specific fixed effects, this finding cannot be explained by female-dominated jobs simply earnings lower wages. With the fixed-effects,

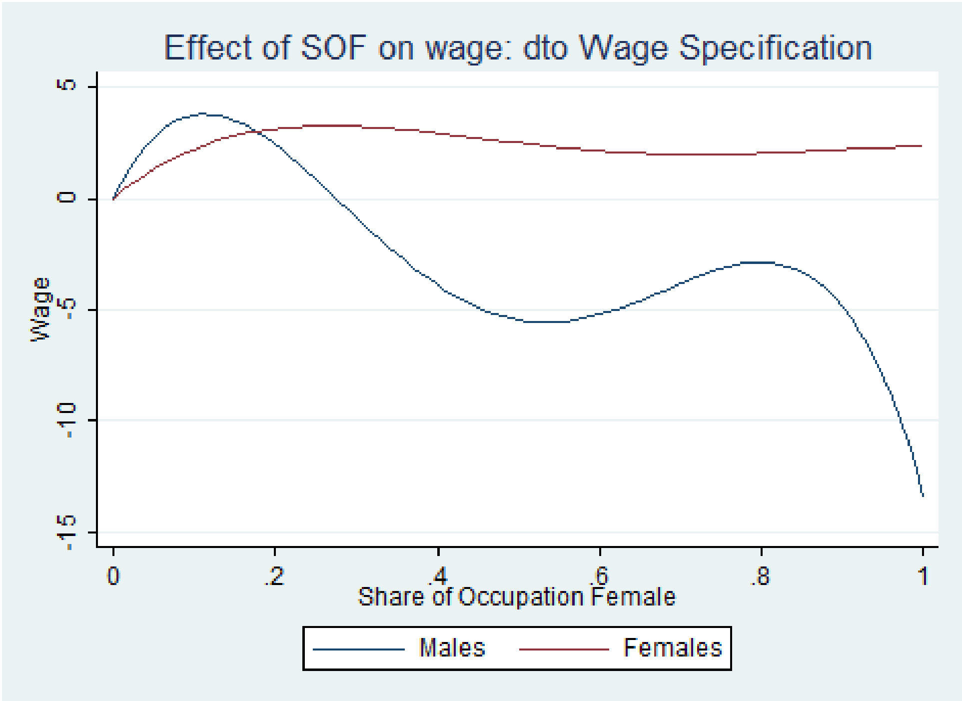

Figure 2: The demeaned effect of share of occupation female in a job (ψ_j) on wages for the detailed occupation specification (46 job categories). The curves give the magnitude of the compensating differentials for both men and women as a result of the share of their occupation that is female. Flipping the graph around the x-axis provides the gender composition preference $g_i(\psi_j)$.

these results show, for instance, that men in a given job in a state-time with high female job composition make lower wages than men in state-times with slightly lower female job composition. For example, these results suggest than male elementary education teachers make less money in states and times where a higher percentage of elementary education teachers are women than in states and times when a lower percentage are women.

5 Conclusion

Our analysis shows that workers may consider the gender composition of occupations when they select an occupation. This causes men and women's labor supplies to differ for jobs based on the proportion of the job that is filled by each gender. My model of occupational selection shows how differences in these preferences can drive both differential selection across occupations and differential wages within an occupation. Compensating differentials within jobs imply that men prefer jobs that are female-dominated while strongly disliking jobs that are roughly 25% female. Women prefer diversity, with their lowest compensating differential at jobs that are approximately 55% female. Women find either extreme less satisfactory, but are least happy with male-dominated jobs.

These results have significant implications for occupational segregation of males and females. Importantly, we can say that the gender composition of jobs are not only important as an outcome, but as an input. We must examine policies which seek to change occupational selection in equilibrium, as they will have direct welfare effects, wage effects, and effects on selection beyond the initial effect of a given policy. Of interest to policymakers, we can say that policies which increase the proportion of women in a job will produce a ripple effect which will increase the representation of the gender with the higher marginal job satisfaction due to increasing the share female. Our results suggest that for male-dominated jobs, increasing the share female will improve satisfaction for women while hurting it for men - thus driving an even higher share female. This effect does eventually reverse for jobs that are less male-dominated, but these jobs are potentially of less interest to policy-makers. Of related importance, the current equilibrium in the economy may not be welfare-maximizing if preferences allow for multiple equilibria with differing job satisfaction levels by gender.

References

Bayard, K., Hellerstein, J., Neumark, D., and Troske, K. (2003). New evidence on sex segregation and sex differences in wages from matched employee-employer data. *Journal of labor Economics*, 21(4):887–922.

Beaudry, P. and Lewis, E. (2014). Do male-female wage differentials reflect differences in the return to skill? cross-city evidence from 1980-2000. *American Economic Journal: Applied Economics*, 6(2):178–94.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago press.

Black, D. A. (1995). Discrimination in an equilibrium search model. *Journal of labor Economics*, 13(2):309–334.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants.

Borjas, G. J. and Bronars, S. G. (1989). Consumer discrimination and self-employment. *Journal of political economy*, 97(3):581–605.

Bowlus, A. J. and Eckstein, Z. (2002). Discrimination and skill differences in an equilibrium search model. *International Economic Review*, 43(4):1309–1345.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420.

Goldin, C. (2014). A pollution theory of discrimination: male and female differences in occupations and earnings. In *Human capital in history: The American record*, pages 313–348. University of Chicago Press.

Heckman, J. J. and Honore, B. E. (1990). The empirical content of the roy model. *Econometrica: Journal of the Econometric Society*, pages 1121–1149.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2013). The allocation of talent and us economic growth. Technical report, National Bureau of Economic Research.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature*, 50(4):959–1006.

Lang, K. and Manove, M. (2011). Education and labor market discrimination. *American Economic Review*, 101(4):1467–96.

Lang, K., Manove, M., and Dickens, W. T. (2005). Racial discrimination in labor markets with posted wage offers. *American Economic Review*, 95(4):1327–1340.

Lordan, G. and Pischke, J.-S. (2016). Does rosie like riveting? male and female occupational choices. Technical report, National Bureau of Economic Research.

Pan, J. (2015). Gender segregation in occupations: The role of tipping and social interactions. *Journal of Labor Economics*, 33(2):365–408.

Rosén, Å. (1997). An equilibrium search-matching model of discrimination. *European Economic Review*, 41(8):1589–1613.

Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics*, 1:641–692.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Sasaki, M. (1999). An equilibrium search model with coworker discrimination. *Journal of Labor Economics*, 17(2):377–407.