Gender Segregation in Occupations: Preferences or Homophily?

Miriam Larson-Koester*

05/23/2019

Abstract

Women remain concentrated in certain occupations despite the drastic increase in women's labor force participation in the U.S. since 1960. I examine whether occupations remain segregated because workers prefer to enter occupations that already employ more of their own gender. I build a model of labor supply and demand in which firms care about the gender and wages of their employees, and workers get utility from their occupation, wage, and the number of women in their occupation. Using a Bartik instrumental variables strategy, I find that women prefer to enter into female-dominated occupations, but men show no evidence of gender preference. In a world without women's gender preference, occupations would be 50 percent less segregated and the gender wage gap 18 percent smaller. However, women's preference for working with women would have to be almost twice as strong for past segregation to affect current segregation.

*Federal Trade Commission, mlarsonkoester@ftc.gov. The views expressed are those of the author and do not necessarily reflect those of the Federal Trade Commission. I am grateful to Francesca Molinari, Richard Mansfield and Victoria Prowse for their advice and support. For helpful discussions and insights I thank Francine Blau, Panle Barwick, Mallika Thomas, Larry Blume, JF Houde, Yang Zhang, Angela Cools, Jorgen Harris, Caroline Walker, and numerous seminar participants. All errors are mine.

1 Introduction

Women's labor force participation has skyrocketed in the U.S. since 1960, but men and women still go into very different occupations. As of 2009, approximately 50% of women would need change jobs in order to to achieve an equal number of men and women in every occupation, and this gap is unlikely to close soon (Blau, Brummund, & Liu, 2013). Recent literature has identified some contributing causes of segregation: preferences over amenities (Olivieri (2014), Wiswall and Zafar (2018))[1], productivity and skill differences (Baker and Cornelson (2018)), and occupational and educational barriers (Hsieh, Hurst, Jones, and Klenow (2016)). I identify another possible cause of the persistence of segregation: workers valuing the gender composition of occupations.

In particular, this paper examines whether current segregation depends on past segregation because workers care about the gender of their occupation in addition to intrinsic characteristics. This is important because the right policy to address segregation, if any, will depend on what mechanisms are at play. Policymakers have expressed concern about occupational segregation, for example the lack of women in STEM and men in nursing. If these occupations are male or female for historical reasons, then segregation itself may be a barrier to individual careers and the efficiency of the economy as a whole. If, on the other hand, occupations are male or female based on intrinsic characteristics of workers or jobs, then interventions that seek to change the gender composition of occupations alone are likely inappropriate, and policymakers concerned about gender segregation should examine policies to address more specific causes as they deem necessary.

Identifying a preference over occupation fraction female is a difficult empirical problem that involves disentangling many other causes of gender segregation. Workers likely care about intrinsic characteristics of occupations, such as skills and job amenities (see for example DeLeire and Levy (2004) and Reed and Dahlquist (1994)), as much as about the gender of their coworkers (see for example Lordan and Pischke (2015) and

[1]For overview see Bertrand (2011) and Cortes and Pan (2017).

Usui (2008)). At the same time, firms may care about which gender is perceived as more productive or valuable, as well as which gender of worker is cheaper to hire. Simply looking at which occupations contain more women will not allow me to distinguish firm and worker preferences, or worker preferences over the fraction female.

The key innovation of this paper, building on the insights of Choo and Siow (2006), Salanié (2014a), and Dupuy and Galichon (2017), is that the distribution of wages can be used to separately identify worker and firm preferences. The intuition is that the right tail of wages tells us more about what firms are willing to pay for workers of a given type, while the left tail tells us more about what workers are willing to accept for a given type of job. Assuming that wage offers are distributed lognormally, and that firms do not care about the productivity of individuals within gender,[2] one can use Maximum Likelihood to estimate firm preferences and the wage offer distribution for male and female workers by occupation. Although the estimation strategy in this paper was developed specifically for segregation in the labor market, it could be used in any context in which a price mechanism clears a two-sided market.

The second identification challenge is to disentangle preferences for the number of women in an occupation from other occupation attributes. If an occupation is becoming more attractive to women, we will observe the fraction female in the occupation going up, and more women entering the occupation, and may incorrectly infer that women are entering the occupation because there are more women, a classic problem of omitted variable bias. To solve this problem I find variation in the fraction female by occupation that is plausibly exogenous to changes in other occupation amenities using Bartik-style instrumental variables. The main identifying assumption is that changes to occupations that workers care about are not correlated across occupations.

For tractability I treat occupation choice as a static choice. Workers choose occupation once at the beginning of their career, based on the contemporary characteristics of the occupations, including fraction female. Therefore, the fraction female in each

[2]Unobserved productivity of individual workers could be included in theory, but would result in very weak identification and likely computational intractability because it would involve joint estimation of the model and integration over multiple sources of unobserved heterogeneity to interpret the wage distribution.

occupation only updates across cohorts of workers.[3] Lifetime wages are set to clear the market in static equilibrium for each cohort of men and women, which requires a large market assumption (continuum of workers and jobs of each type) to guarantee a unique wage equilibrium.

The data moments to identify the model are the shares of male and female workers by gender from the 1960-2000 U.S. Censuses and 2012 3-year ACS, and estimates of lifetime income by occupation, gender, and year constructed using income quantiles from the Census data combined with transition rates from the SIPP 2004 and 2008 panels. I take a two-step estimation approach, first estimating the firm side using maximum likelihood, then the worker side using instrumental variables regression, taking the wage offer distribution estimated on the firm side as data. The structure of the model allows me to simulate transition paths in the fraction female by cohort and determine there is path dependency, all while solving for equilibrium wages and fixing other firm and worker preferences.

I find that women care strongly about the number of women in an occupation, but no evidence that men care about the number of women. The point estimate in my preferred specification is very high, with an occupation change from 25% to 75% female being equivalent to an extra $3 million in lifetime income for a woman. The gender preference leads to more segregated outcomes. With no gender preference, the model predicts no occupations with fraction female greater than 70% or less than 10%, and a Duncan segregation index of 24% in 2012, meaning 24% of male or female workers would have to change occupations to make all occupations 50% female. With the gender preference, the Duncan index is predicted to be almost twice as high at 47%.[4]

Labor supply or demand shocks that affect the fraction female are reinforced by a feedback loop from the gender preference. An illustrative case study is insurance adjustors, which is one of many occupations observed to have moved from male to female by Pan (2015). I find that more women began to become insurance adjusters because

[3]Allowing workers to choose occupation myopically more than once during their lifetime, would lead to faster changes in occupation fraction females, but not otherwise change the dynamics of the model.

[4]The Duncan index in my observed 2012 data is 41%.

firms' demand for women in this occupation rose. Then as more women entered, the occupation became more attractive to women. This in turn made women cheaper to hire, which increased labor demand for female insurance adjusters, providing further reinforcement of the feminization of the occupation. My model predicts that if it were not for women preferring to work with women, insurance adjusters would have become only 50% female, rather than the observed over 70% female (from a starting point of 20% female in 1960). Thus my estimated model provides one possible mechanism that is consistent with the stylized fact that occupations tend to move rapidly from male to female. This phenomenon is referred to as "tipping" by Pan (2015).[5]

Although the gender preference I estimate is very strong, it is largely mitigated by compensating differentials. In a model without endogenous wages, all occupations eventually converge to either 0% or 100% female. Wage adjustment leads to more moderate outcomes. As women enter an occupation, firms can hire the same number of women at lower cost, and lower wages dampen the increase in female labor supply. Likewise as women exit an occupation, firms must pay the remaining women more, which slows the exit of women. I find that these compensating differentials explain 18% of the difference in lifetime income between men and women, with women in highly female dominated occupations suffering the largest loss in earnings. The result that as women enter an occupation, wages go down is consistent with previous literature (Levanon, England, & Allison, 2009; Harris, 2018).

The presence of a gender preference could make it most profitable for occupations to hire only men or majority women. In this case, the segregation today may have been selected due to initial conditions such as historical barriers or norms (Schelling, 1971; Pan, 2015). However the gender preference I estimate would have to be about twice as strong to cause current occupation sorting to depend on past occupation sorting.

There are two important takeaways from this paper for policymakers. First, short run shocks to the fraction female in occupations, such as temporarily pushing more

[5] "Tipping" is also a documented phenomenon in racial segregation by neighborhoods (Card, Mas, & Rothstein, 2008) and racial composition of schools (Caetano & Maheshri, 2017).

men into nursing or women into STEM, will not have long run consequences since there is no historical inertia to overcome. Policymakers seeking to reduce segregation should find specific supply and demand factors that concern them and address them directly, whether it be occupation amenities or discrimination. Second, policymakers should keep in mind that changes to labor supply and demand might have an outsized impact on segregation due to the feedback loop of women wanting to work with women. Making an occupation more attractive to women might have the unintended consequence of making the occupation highly female dominated, and thereby lowering the wages for those women over time through a compensating differential.

Lastly, future research should address the source of the gender preference to determine it is due to gender identity, or workplace environment and amenities. If workplace environment and amenities are the cause, then policymakers who care about reducing segregation and the gender wage gap might consider how to address these to make male-dominated occupations more welcoming to women.

Section 2 introduces the structure of the transferable utility matching model. Section 3 describes equilibrium wages. Section 4 describes the empirical specification and first stage of estimation, recovering the firm side parameters and reservation wages. Section 5 describes the second stage of estimation, the decomposition of worker utility taking reservation wages as given. Section 6 describes the data sources. Section 7 presents results, and Section 8 presents simulations of segregation dynamics, including counterfactuals.

2 Matching Model with Observed Transfers

In the model, career choice is a one-to-one match between a job and a worker. Workers choose an occupation at the beginning of their lives, based on their individual tastes for occupations and the wage offer for their gender in that occupation. The job choice, including any education costs that may be associated with it, is binding for life. I assume that workers do not anticipate future changes in the market that would lead them to switch occupations. With myopic workers, lowering switching cost would only lead to faster convergence to a fixed point in the fraction female by occupation. However, relaxing myopia would pose a challenge to tractability because the endogenous attribute (fraction female) would then depend on expectations over the future behavior of all agents, complicating today's choice set considerably.

Jobs choose the gender of worker that will maximize productivity per dollar spent, taking into account compensating differentials for job amenities, which may be differentially valued by men and women. This is consistent with a model of firms where jobs within a firm are filled independently, perhaps due to costly coordination or low complementarity. Individual jobs optimizing over rate of return does not imply a specific firm-level production function, but it produces similar behavior as firms minimizing cost with production quotas. Without data on firm size, I take this as an approximation of firm behavior.

The lifetime income distributions by gender and occupation needed to clear the market are determined in equilibrium to equate demand and supply, and are made up of two components: a job-specific component, and an aggregate component that varies by occupation and gender. The job-specific component reflects the idiosyncratic disutility of taking a specific job. In equilibrium firms must make workers indifferent across individual jobs within an occupation, and therefore firms perfectly compensate workers for idiosyncratic disutility. Wage heterogeneity within occupation and gender emerges from these job-specific compensating differentials.

In addition to an individual job level compensating differential, there is a component of wages which is common across individuals within gender and occupation, and serves to equate supply and demand. It is a function of all parameters in the model, specifically, both of the utility that workers receive from occupations as well as the value that jobs have for male and female workers. This component of wage can be thought of as the result of a market-wide open-ended ascending price auction where jobs make wage bids for workers, but each job can only "win" one worker, and there are no search frictions. The assumption of no search frictions is made more reasonable by the fact that I am considering lifetime occupation choices, not individual jobs. In order for equilibrium wages to not depend on sample size I must assume a large number of workers and jobs of each type (Galichon & Salanié, 2013a). In my model this means a large number of men and women and jobs in each occupation.[6]

The identification strategy in this paper builds off empirical applications of matching models to marriage markets (e.g. Choo and Siow (2006) and Chiappori, Salanié, and Weiss (2015)), but differs in several key respects. Because a key model parameter is how non-wage utility varies with occupation fraction female, it is critical that I separately identify non-wage utility from utility from wages. Previous literature only identifies the sum of wage and non-wage utility.

The first step to separate identification is to assume that observed wages are the transfers between job and worker that clear the market. Although some transfers in the labor market may be non-wage, such as enhanced benefits, wages are a natural first-order approximation. The difficulty is that observed transfers are the sum of an idiosyncratic and an aggregate component, as discussed above, and for counterfactuals these must be separately identified. The observed distribution of wages is also the result of optimization over these components and other model parameters to be estimated, specifically the willingness-to-pay for workers by gender and the variance of the distribution of unobserved heterogeneity in job amenities.

In the case that heterogeneity on both sides of the market is distributed extreme value, it is possible to solve for all parameters from a few key moments as suggested by Salanié (2014a). However in my application, such a strategy has some major drawbacks. First, logit heterogeneity does not fit the observed wage distribution. Wages are typically assumed to be lognormal since they are constrained to be greater than zero and have a long right tail. Unlike the logit case, the variance of the distribution of the maximum of lognormals depends on other model parameters to be estimated, so the variance needs to be jointly estimated unlike in Salanié (2014a).

[6]For reference, simulations of an ascending price auction with sample size of 20 already produce a wage vector similar to the equilibrium vector assumed by the large sample. I perform the simulations using a modified version of the auction mechanism outlined in Roth and Sotomayor (1990) page 209.

Second, unlike in the marriage market, what it means to be "unmatched" is not clearly defined on both sides of the market. For workers, it is clear that being unmatched means being unemployed or out of the labor force. On the job side, however, it is less clear. The number of unmatched jobs, as a primitive, would depend on an occupation production function and capacity constraints. Rather than estimate this, I choose to use data on posted vacancies that remain unfilled for some time. The decision to post a vacancy is endogenous to market conditions, so vacancy data is an imperfect proxy for unfilled jobs. In order to limit the importance of the vacancy data in my estimation, I choose a specification that in principle does not require it, by assuming non-negative profit on the job side. However in practice this relies very heavily on the functional form assumption on the tail of the distribution of job dis-amenities. As a result in most specifications I use data on vacancies by occupation imputed from the Job Openings and Labor Turnover Survey (U.S. Bureau of Labor Statistics, 2017).

Lastly, I estimate four variance parameters in total: variance of taste for occupation for men and women, and variance of job dis-amenities for men and women. These parameters are important because they govern the elasticity of labor supply to changes in wages or other non-wage amenities such as the gender ratio in the counterfactual, as well as the demand response to changes in equilibrium wage. Unlike Chiappori et al. (2015), who uses many markets to identify these parameters, I use the observed wage distribution as suggested in Salanié (2014a). However I estimate using maximum likelihood in order to use the information from the wage distribution efficiently, and because the lognormal assumption means that the scale parameters cannot be separately estimated from all other parameters, as noted above. For all these reasons maximum likelihood is the most natural strategy for my application.

2.1 General Model Structure

When a worker and a job match, total surplus is created from the match. In the worker's case the value of a match reflects the amenities of the job. A job might have a particularly collegial environment, or free child care for example. Amenities may be valued differently by gender. On the job side the payoff is the willingness-to-pay for a worker, which could reflect productivity, and differ by gender due to gender differences in turnover, differences in search cost by gender, differences in productivity, or devaluation, for example. The wage determines the split of the total surplus between the worker and the firm.

The most general payoff structure in a matching model would allow each possible match between a worker i and a job j to have its own unobserved match quality. To make the problem empirically tractable, I assume that no portion of the payoff depends on unobservable characteristics of both firm and worker, which is a standard assumption in empirical matching. So although the surplus may depend on i or j, it may not depend on i and j.

Assumption 1. *Additive Separability: No component of surplus depends on unobserved characteristics of both workers and firms.*

Formally, let g denote gender, which is observed as either male (M) or female (F) in this model. Let o denote occupation. We therefore have workers $i \in g \in G = \{M, F\}$ and jobs $j \in o \in O = \{1, 2, ..., 34\}$.[7] Under additive separability we have that the total surplus from a match between worker i and job j, S^i_j, can be decomposed:

$$S^i_j = S^g_o + \eta^i_o + \xi^g_j \tag{1}$$

Note that there are components that vary at the occupation*gender level (S^g_o), the occupation*worker level (η^i_o), and the gender*job level (ξ^g_j), but never the worker*job level. In other words, additive separability implies that there is no ξ^i_j or η^i_j. This assumption is important because it allows me to separate the matching problem into two separate discrete choice problems, one for each side of the market (Galichon & Salanié, 2013b).

The components of total surplus that depend on unobservables of either the worker

[7] Thirty-four occupations are chosen according to data constraints discussed below.

(η_o^i) or the job (ξ_j^g) can theoretically come from the worker's utility function, the job payoff function, or both. In order to gain identifying power from the observed wage distribution, and because my research question is focused the the role of worker utility in occupation choice, I assume all unobserved components of surplus originate from the worker's utility. This means that only workers have preferences over unobservables, and jobs care only about whether they chose to hire a male or female worker.

Assumption 2. *η_o^i and ξ_j^g are primitives in the worker's utility function.*

In other words, each worker has an individual taste for each occupation (η_o^i) and each job differs in how attractive it is to men and women (ξ_j^g). The job amenity heterogeneity can be thought of as any component of the attractiveness of a job that is orthogonal to the overall attractiveness of the occupation, which is included in S_o^g. For example child care offerings at a particular employer might differ relative to the average child care offerings in that occupation.

The fact that jobs care only about whether to hire a male or female worker and the wage that they must pay means that conditional on gender, all workers are equally productive. The counterfactual dynamics I am interested in are at the level of occupation and gender, and productivity and utility are allowed to vary freely by occupation and gender. Therefore including individual worker productivity heterogeneity should not have a major impact on my results. In terms of biasing parameter estimates, I may overestimate how undesirable small occupations are. High wages in small occupations may be due to selecting the most skilled workers rather than dis-amenities. Similarly I may overestimate the value of amenities in large occupations if wages are low due to a lower than average level of worker skill, rather than attractive amenities.

2.2 Payoff Functions

Jobs care only about the relative productivity of men and women and the wage that they will have to pay to an individual worker. Let WTP_o^g be the willingness-to-pay of a job in occupation o of hiring a worker of gender g. The total payoff to job j, π_j^g, is specified as the ratio of willingness-to-pay to the cost of hiring a worker.

$$\pi_j^g = \frac{WTP_o^g}{Wage_j^g} \tag{2}$$

The ratio in equation 2 is used as opposed to an additive profit function for two reasons. First, since firms in the model only fill one vacancy at a time, rate of return is a reasonable object for firms to care about that does not require making assumptions about the number of total vacancies at a firm. By allowing the firm to maximize return on a single vacancy at a time I make a reasonable job-level approximation of a firm level profit function. Second, it is common in labor economics to specify wages as lognormally distributed. The multiplicative specification allows model wages to be lognormal in the estimation stage, which in turn provides the best possible fit to the observed wage data.

$$\begin{aligned} log(\pi_j^g) =& log(WTP_o^g) - log(Wage_j^g) \\ \equiv& \overline{WTP}_o^g - \overline{Wage}_j^g \end{aligned} \tag{3}$$

The payoff to the firm of not hiring any worker, π_j^N, is normalized to one, so that the log payoff to not filling a vacancy is zero, ($\bar{\pi}_j^N = 0$).

The worker's taste for the occupation consists of two components: u_o^g which is common to all workers of gender g matching to jobs in occupation o, and η_o^i which is worker i's specific utility from occupation o. The dis-amenities of job j, which are denoted ξ_j^g, are the same for everyone conditional on gender.

The payoff to worker i is specified as:

$$u_j^i = \frac{u_o^g * Wage_j^g * \eta_o^i}{\xi_j^g} \tag{4}$$

Non-wage and wage utility thus enter the specification multiplicatively, including the taste heterogeneity term η^i_o. Define log utility to be $\bar{u}^i_j$ and re-parameterizing in terms of logs we have:

$$\begin{aligned} \bar{u}^i_j \equiv log(u^i_j) &= log(u^g_o) + log(Wage^g_j) + log(\eta^i_o) - log(\xi^g_j) \\ &\equiv \bar{u}^g_o + \overline{Wage}^g_j + \bar{\eta}^i_o - \bar{\xi}^g_j \end{aligned} \tag{5}$$

Workers choose a job j to maximize log utility $\bar{u}^i_j$. The log utility parameters for working, $\bar{u}^g_o$, include the disutility of working at all and are expected to be negative. However the underlying utility parameters $u^g_o = exp(\bar{u}^g_o)$ will always be positive leading to complementarity in wage and non-wage utility. Similarly, ξ^g_j will be assumed to be log-normally distributed and therefore always positive, so the higher the dis-amenity value of the job the lower the utility.

The common taste parameter for non-employment $\bar{u}^g_N$, where non-employment means either not in the labor force or unemployed, is normalized to zero. It is also assumed that in non-employment workers receive no wages or value from job amenities. The payoff to non-employment ($\bar{u}^i_N$) is therefore $\bar{\eta}^i_N$, the idiosyncratic taste for non-employment.

Furthermore worker utility will be decomposed into a common component and a dependence on the fraction female in the occupation, which is the endogenous amenity of interest. Let F_o be the fraction female in occupation o.

$$\begin{aligned} \bar{u}^i_j &= \bar{u}^g_o + \overline{Wage}^g_j + \bar{\eta}^i_o - \bar{\xi}^g_j \\ &= \alpha^g_o + \gamma^g F_o + \overline{Wage}^g_j + \bar{\eta}^i_o - \bar{\xi}^g_j \end{aligned}$$

So α^g_o and γ^g will be the parameters of interest, but $\bar{u}^g_o$ can be recovered from these parameters and the fraction female.

3 Equilibrium Wages and Stability

In the following section I outline conditions for a matching to be feasible and stable. I then introduce the equilibrium wage vector and show that it supports feasibility and stability.

3.1 Feasibility

A matching is feasible if every worker is matched to at most one job and every job matched to at most one worker. Formally, following Galichon and Salanié (2015), let μ_j^i be equal to either 0 or 1 where 1 indicates a match between worker i and job j. Then for every i and j a feasible matching has

$$\sum_{k \in \mathcal{J}} \mu_k^i \leqslant 1 \text{ and } \sum_{k \in \mathcal{I}} \mu_j^k \leqslant 1$$

Similarly following Galichon and Salanié (2015), the matching must be feasible given the number of men and women and jobs in each occupation available in the market, or

$$\sum_{j \in \mathcal{J}} \mu_j^g \leqslant n_g, \ \forall g \text{ and } \sum_{i \in \mathcal{I}} \mu_o^i \leqslant n_o, \ \forall o$$

3.2 Stability

Intuitively, pairwise stability implies that no worker and job that are not currently matched with each other, would prefer to match with each other. Let i and j be a so called "blocking pair", and let i be currently matched to $j(i)$ and j to $i(j)$. Then pairwise stability states that the sum of the individual surpluses from the existing matches (i with $j(i)$ and j with $i(j)$) must be greater than the surplus of the blocking pair (i and j). Therefore even with any possible transfer, i and j will not both prefer to match with each other, because the total possible surplus is lower.

Definition 1. *Pairwise Stability: In a matching where i is paired with $j(i)$ and j is* *paired with $i(j)$, it must be the case that $\bar{u}^i_{j(i)} + \bar{\pi}^{i(j)}_j \geqslant \bar{u}^i_j + \bar{\pi}^i_j, \quad \forall i, j$. In addition, each worker and job must attain higher surplus than their outside option, or $\bar{u}^i_{j(i)} \geqslant \bar{u}^i_N$ and $\bar{\pi}^{i(j)}_j \geqslant \bar{\pi}^N_j$, where N represents not working for the worker, and not hiring for the firm.*

Note that on the left hand side $\bar{u}^i_{j(i)} + \bar{\pi}^{i(j)}_j$ includes the wage paid out to the worker and by the job in their respective matches. On the right hand side $\bar{u}^i_j + \bar{\pi}^i_j$ the wage will cancel within the match leaving the underlying total surplus.

Following Shapley and Shubik (1972), the pairwise stable matching will be unique and the competitive equilibrium will coincide with the pairwise stable matching, but the competitive equilibrium wage vector may not be unique. I assume the observed wages are the equilibrium wages described in Galichon and Salanié (2015) and Salanié (2014b). These are the wages that make workers indifferent over jobs within each occupation, and jobs indifferent over workers within each gender. As the sample size of men and women goes to infinity, the equilibrium wages will be unique (Galichon & Salanié, 2015).

3.3 Market Clearing Wages

The equilibrium wage vector I propose, $\overline{Wage}^g_j$, equates supply and demand for workers by gender and jobs by occupation. It does so by compensating the worker for the idiosyncratic dis-amenities of a particular job within occupation, and through a common component W^g_o, which varies by gender and occupation.

$$Wage^g_j = W^g_o * \xi^g_j \tag{6}$$

A firm receives a draw of ξ^M_j and ξ^F_j, which is the amenity value of the job to men and women respectively. The firm then chooses to hire a male or female worker based on the overall productivity of men and women in that occupation (WTP^g_o) and the cost of hiring men and women which varies in order to compensate exactly for the utility or disutility experienced by the worker due to ξ_j^M and ξ_j^F. This is a compensating differential at the job level. Compensating differentials also emerge at the occupation level through the common component of wages, W_o^g.

The worker utility function therefore takes into account the job dis-amenities ξ_j^M and ξ_j^F as well as the workers taste for that occupation, and because $log(Wage_j^g) = log(W_o^g * \xi_j^g) = \bar{W_o^g} + \bar{\xi_j^g}$, workers are exactly compensated for ξ_j^g in the utility function. Plugging in and cancelling we get equilibrium utility for workers of

$$
\begin{aligned}
\bar{u}_o^i &= \bar{u}_o^g + \overline{Wage}_j^g + \bar{\eta}_o^i - \bar{\xi}_j^g \\
&= \bar{u}_o^g + \bar{W}_o^g + \bar{\xi}_j^g + \bar{\eta}_o^i - \bar{\xi}_j^g \\
&= \bar{u}_o^g + \bar{W}_o^g + \bar{\eta}_o^i \\
&= \alpha_o^g + \gamma^g F_o + \bar{W}_o^g + \bar{\eta}_o^i
\end{aligned} \tag{7}
$$

The equilibrium payoff to a firm is then

$$
\begin{aligned}
\bar{\pi}_j^g =& \overline{WTP}_o^g - \overline{Wage}_j^g \\
=& \overline{WTP}_o^g - \bar{W}_o^g - \bar{\xi}_j^g
\end{aligned} \tag{8}
$$

The fact that this wage vector supports the pairwise stable matching follows from the optimization problems of jobs and workers. Workers choose an occupation to maximize utility, and firms choose a worker to maximize rate of return, so the chosen job $j*$ and worker $i*$ respectively must satisfy

$$
\begin{aligned}
j^* \in o^* &= \arg\max_o (\bar{u}_o^g + \bar{W_o^g} + \eta_o^i) \\
i^* \in g^* &= \arg\max_g (\overline{WTP}_o^g - \bar{W_o^g} - \bar{\xi_j^g})
\end{aligned}
$$

From this it is clear than within an occupation, workers are indifferent to which job they are matched to, and likewise within gender, jobs are indifferent to which worker they are matched to.

This implies that if worker i were to match with a different job within the same occupation, we would have $\bar{u}^i_j = \bar{u}^i_{-j}$, and likewise for job j, $\bar{\pi}^i_j = \bar{\pi}^{-i}_j$, therefore the pairwise stability inequality holds trivially for observationally equivalent (same g and o) candidate matches:

$$\bar{u}^i_{j(i)} + \bar{\pi}^{i(j)}_j = \bar{u}^i_j + \bar{\pi}^i_j \quad \forall i, i(j) \in g \quad \forall j, j(i) \in o$$

Now consider matching worker i to a job in a different occupation. Both workers and jobs choose the occupation or gender that produces the highest payoff for them, given the wage vector. Let the optimal occupation be o^* and optimal gender g^*. Therefore we know that for worker i

$$\bar{u}^i_{j(i)} > \bar{u}^i_j \quad \forall j(i) \in o^* \text{ and } \forall j \in o \neq o^*$$

and for job j

$$\bar{\pi}^{i(j)}_j > \bar{\pi}^i_j \quad \forall i(j) \in g^* \text{ and } \forall i \in g \neq g^*$$

Therefore pairwise stability holds with strict inequality for all candidate matches that are not observationally equivalent (different g or o) to the competitive equilibrium.

The second part of pairwise stability is the requirement that the choice payoffs be greater than the outside option payoffs. Recall that the outside option for the worker is remaining unemployed is equal to the idiosyncratic taste for non-employment, $\bar{u}^i_N = \bar{\eta}^i_N$. The value to the firm of not hiring a worker is simply zero, $\bar{\pi}^j_N = 0$.

Another key aspect of the equilibrium wage vector is that it must be feasible, which in the case of this labor market is equivalent to equating supply and demand at the level of male and female workers and occupations. Crawford and Knoer (1981) and Roth and

Sotomayor (1990) prove the existence of such an equilibrium in a model with transfers. Intuitively, as long as the common component of wage, or W_o^g, is free to adjust, supply and demand can adjust until the market clears. The empirical implications of market clearing will be discussed in the empirical section below.

4 Empirical Strategy

The empirical strategy consists of two stages: first I use maximum likelihood to disentangle the selection effects of worker and firm optimization on observed matches and wages. I am able to separately identify the willingness-to-pay of the firm and the reservation wage distributions of the workers. This effectively separates the two sides of the market, unlike previous applications of transferable utility matching models, such as Choo and Siow (2006) and Chiappori et al. (2015), who were only able identify total match surplus because they lacked data on transfers.[8]

Second, the estimated reservation wage distributions from the maximum likelihood estimation are used to estimate the worker's utility. Workers care about non-wage amenities, wages, and the fraction female. In the second stage of estimation, a panel regression for each gender is run of shares of workers in each occupation on reservation wages, occupation intercepts, and fraction female. Instruments are used to control for omitted variables correlated with both fraction female and wages over time, and provide clean variation in the fraction female and reservation wage to trace out labor supply.

For clarity of identification, I estimate the two sides of the market, workers and jobs, separately. The job side of the model is a maximum likelihood estimation exploiting the full variation of the observed wage distribution at the individual match level. The worker side, by contrast, is an instrumental variables regression at the occupation-year level, where instruments provide clean variation in the fraction female and reservation wage to trace out labor supply. Combining the two estimation steps using joint GMM

[8]Fox (2010) is able to estimate payoffs from the interaction of characteristics of both sides of the market but not full payoff functions.

is theoretically feasible and would have the advantage that all data moments are used to identify all parameters. However the large number of parameters (914 parameters at once)[9] and the need to pool all years of data (over 7 million observations) makes joint estimation computationally unattractive.

In the following sections I lay out the assumptions that allow the ML estimation of the willingness-to-pay and reservation wages. The reservation wages are then taken as data in the estimation of worker utility parameters, which is described in the following section.

4.1 Job

Assumption 3. *Let the heterogeneity in job amenities, ξ_j^g, be distributed lognormal, and independently across j, such that $\bar{\xi}_j^g = ln(\xi_j^g)$ is distributed normally with location parameter normalized to zero and scale parameters for each gender, σ_ξ^F and σ_ξ^M.*

A firm will only hire a worker if the willingness-to-pay for that worker is higher than the wage the worker will accept in equilibrium, which means the payoff π_j^g must be greater than one.

$$\pi_j^g = \frac{WTP_o^g}{Wage_j^g} \geqslant 1$$

Assumption 4. *Jobs will remain unfilled if the wage is higher than the willingness-to-pay for both men and women.*

$$\pi_j^F = \frac{WTP_o^F}{Wage_j^F} < 1 \quad \text{and} \quad \pi_j^M = \frac{WTP_o^M}{Wage_j^M} < 1 \qquad \Longrightarrow \qquad j \text{ unfilled}$$

The above assumption implies that jobs are unfilled when both $\bar{\xi}_j^M$ and $\bar{\xi}_j^F$ are very high. High $\bar{\xi}_j^M$ and $\bar{\xi}_j^F$ means very high dis-amenity value to both men and women, requiring high wages to compensate. So if a job is unattractive enough, it will not be filled.

[9]There are 138 firm side parameters estimated in the first stage in each of 6 waves of data, for a total of 828 parameters. There are 86 worker parameters estimated in the second stage over all waves of data.

There is no equivalent on the firm side to the idiosyncratic preference for non-employment that exists on the worker side. Heterogeneity in the outside option for a firm could reflect heterogeneity in the cost of hiring for different jobs, or the substitutability of a particular job in the firm production function. I make the assumption of a fixed outside option because then the model is identified without use of data on vacancies,[10] allowing me to test robustness to the use of the vacancy data.

4.1.1 Likelihood Function

Parameters of the model are

$$\theta = \{WTP_o^g, \alpha_o^g, \gamma^g, \sigma_\eta^g, \sigma_\xi^g\}$$

Recall that WTP_o^g are the firms' willingness-to-pay parameters, α_o^g the workers' non-wage utility unrelated to the fraction female, γ^g the value of the fraction female to the worker, σ_η^g the scale of the worker taste heterogeneity, and σ_ξ^g the scale of the job dis-amenities heterogeneity. The unconditional centers of the reservation wage distributions, W_o^g, are reduced form outcomes determined in equilibrium as a function of the primitives θ. Nevertheless recovering the W_o^g from the observed conditional wage distributions is critical to recovering the fundamental parameters in the worker's utility function.

Recall the log wage specification,

$$\overline{Wage}_j^g = \bar{W}_o^g - \bar{\xi}_j^g$$

Since we observe log wages $\overline{Wage}_j^g$, the expectation of observed wages suggests itself as an estimator for $\bar{W}_o^g$. Unfortunately there is a wrinkle to this strategy. There is a selection problem in that the only wages that are observed are the wages that maximize the job's choice over male, female, or not hiring any worker. The choice to

[10] I can estimate the willingness-to-way of the firms with the maximum observed wage by gender and occupation. This strategy is less stable than using vacancy data.

not hire at all is especially problematic since it implies a truncation point that depends on W_o^g which is unknown. Likewise, the selection effects depend on the scale of the unobserved job heterogeneity, σ_ξ^g, which is unknown. The problem therefore suggests joint likelihood estimation with selection and truncation, where the selection takes the form of a Tobit Type 5 (Amemiya, 1985)[11], to estimate WTP_o^g, W_o^g and σ_ξ^g.

If the job is filled, its contribution to the likelihood function takes the form

$$\begin{aligned} LL_{j_{\text{filled}}} &= \prod_{j_{\text{filled}}} Pr(j \text{ hire g}, Wage_j^g) * Pr(j \text{ filled}) \\ &= \prod_{j_{\text{filled}}} Pr(j \text{ hire g}|Wage_j^g) * Pr(Wage_j^g) * Pr(j \text{ filled}) \end{aligned}$$

If the job is unfilled, its contribution is

$$LL_{j_{\text{unfilled}}} = \prod_{j_{\text{unfilled}}} Pr(j \text{ unfilled})$$

$\mathcal{I}($ j unfilled) be the indicator function equal to one if the job j is unfilled, and similarly for $\mathcal{I}(j$ filled). Then the total likelihood is given by

$$LL_j = \prod_j (Pr(j \text{ hire g}|Wage_j^g) * Pr(Wage_j^g) * Pr(j \text{ filled}))^{\mathcal{I}(j \text{ filled})} * (Pr(j \text{ unfilled}))^{\mathcal{I}(j \text{ unfilled})}$$

$Pr(j$ unfilled) is the probability that we do not observe a match, which I impute from the JOLTS vacancy data.[12] This occurs when hiring neither women nor men

[11] See appendix for the likelihood written in the notation of Amemiya (1985).
[12] Results do not appear sensitive to imputation method.

produces willingness-to-pay higher than cost.

$$\overline{WTP}_o^M - \bar{W}_o^M - \bar{\xi}_j^M < 0 \quad \text{and} \quad \overline{WTP}_o^F - \bar{W}_o^F - \bar{\xi}_j^F < 0$$

The $Pr(j \text{ filled}) = 1 - Pr(j \text{ unfilled})$ is then

$$\begin{aligned} & 1 - Pr(\xi_j^F < -(\overline{WTP}_o^F - \bar{W}_o^F), \quad \xi_j^M < -(\overline{WTP}_o^M - \bar{W}_o^M)) \\ &= 1 - \Phi_{0,\sigma_\xi^F}(-(\overline{WTP}_o^F - \bar{W}_o^F) * \Phi_{0,\sigma_\xi^M}(-(\overline{WTP}_o^M - \bar{W}_o^M)) \end{aligned}$$

Let Φ_{0,σ_ξ^g} and ϕ_{0,σ_ξ^g} are the cdf and pdf of the normal distribution with location zero and scale σ_ξ^g. Recall that

$$\overline{Wage}_j^g = \bar{W}_o^g + \bar{\xi}_j^g$$

Then other terms in the likelihood are as follows:

$$\begin{aligned} Pr(j \text{ hire M}|\overline{Wage}_j^M) &= \Phi_{0,\sigma_\xi^F}(\overline{WTP}_o^M - \overline{Wage}_j^M) - (\overline{WTP}_o^F - \bar{W}_o^F)) \\ Pr(\overline{Wage}_j^M) &= \phi_{0,\sigma_\xi^M}(\overline{Wage}_o^M - \bar{W}_o^M) \\ Pr(j \text{ hire F}|\overline{Wage}_j^F) &= \Phi_{0,\sigma_\xi^M}(\overline{WTP}_o^F - \overline{Wage}_j^F) - (\overline{WTP}_o^M - \bar{W}_o^M)) \\ Pr(\overline{Wage}_j^F) &= \phi_{0,\sigma_\xi^F}(\overline{Wage}_j^F - \bar{W}_o^F) \end{aligned}$$

4.1.2 Identification

Separate identification of the willingness-to-pay and reservation wage parameters (WTP_o^g and W_o^g) relies on the shape of the wage distribution, and in practice, the addition of vacancy data. We observe the share of jobs that are filled by men and women in each occupation, and the full distribution of $Wage_o^F$ and $Wage_o^M$ for those matches that do occur. Intuitively, the shape of the observed wage distribution will vary according to where the reservation wage distribution is centered, while what portion of the reservation wage distribution we see depends on the WTP_o^g, that is jobs relative valuation of male or female workers. To give intuition in terms of the likelihood function, W_o^F and W_o^M are pinned down by the $Pr(Wage_j^g)$ component, while the relative difference between WTP_o^F and WTP_o^M is pinned down by the $Pr(j \text{ hire g}|Wage_j^g)$ component, and the level of both WTP_o^F and WTP_o^M is pinned down by the denominator $Pr(j \text{ unfilled})$.

To illustrate identification consider the plots in Figures 1 and 2. These figures compare the reservation wage distributions implied by the model, with the wage distribution of matches, also implied by the model. The vertical lines denote the willingness-to-pay for male and female workers respectively, $\overline{WTP}_o^M$ and $\overline{WTP}_o^F$. In both example occupations, the wage distribution predicted for matches (top panel) is lower for women than men. In the first example in Figure 1, Sales Representatives, Finance, and Business Services, this is the result of firm preferences. In the second example in Figure 2, Health Service occupations, this is the result of worker preferences.

In the bottom panel of Figure 1, we see that in Sales Representatives, Finance, and Business Services, the reservation wages of women and men are centered at approximately the same location. This implies that men and women value the amenities of the occupation similarly and have similar outside options. Differences in the wage distribution of predicted matches are driven instead by firms being willing to pay higher wages for male workers. This manifests itself in the wage distribution by a long right tail male wages, beyond the support of the female wage distribution.

By contrast in Figure 2, in Health Service occupations, we see that the reservation wage distribution for men is much higher than for women. This means that men do not value this occupation, or have higher outside options. The high reservation wages for men drives the wage gap in this occupation. The estimated willingness-to-pay gap is very small because the support of the right tails of the male and female wage distributions broadly overlaps.

In summary using the model I am able to distinguish between the scenario in Figure 1, where female workers are less valued than male workers, and the scenario in Figure 2, where female workers are cheaper to hire than male workers. In both cases the observed wages for women are lower, but the mechanisms are very different. Separately identifying reservation wages from willingness-to-pay is important because taking reservation wages as given is what allows me to estimate the worker side of the market.

4.2 Worker

Assumption 5. *Let the worker taste heterogeneity for occupations, η_o^i, be independently distributed extreme value type 1.*[13]

The location parameter of η_o^i is normalized to zero, and the scale parameter is estimated separately for each gender, resulting in two scale parameters σ_η^M and σ_η^F for men and women respectively. In Chiappori et al. (2015) the scale parameters require many markets to estimate but with observed transfers we can rely on variation in lifetime income across occupations to trace out the scale.

I leverage well known properties of extreme value distributions to obtain choice probabilities, or the probability that the utility from one option is higher than the utility of all other options. Recall that $log(u_g^i) = log(u_o^g) + log(W_o^g) + \eta_o^i \equiv \bar{u}_o^g + \bar{W}_o^g + \eta_o^i$.

$$Pr(i \in g \text{ chooses } \forall j \in o) = \frac{exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})}{\sum_o exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})}$$

Assumption 6. *Let $u_N^i = \eta_N^i$. This implies that the $u_N^g = 0$ and no wages are received, leaving only idiosyncratic taste for non-employment η_N^i.*

We can then define the probability of choosing occupation o relative to the probability of choosing non-employment.

[13] Also known as gumbel for maxima or logit.

$$\frac{Pr(i \in g \text{ chooses } \forall j \in o)}{Pr(i \in g \text{ chooses } N)} = \frac{\frac{exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})}{\sum_Y exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})}}{\frac{1}{\sum_Y exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})}} = exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})$$

Let s_o^g be the share of workers of gender g who match to occupation o, and s_N^g the share that choose non-employment. These shares are observed. Rearranging from above we have

$$\sigma_\eta^g ln(s_o^g) - \sigma_\eta^g ln(s_N^g) = \bar{u}_o^g + \bar{W}_o^g \tag{9}$$

$$ln(s_o^g) - ln(s_N^g) = \frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g} \tag{10}$$

$$= \frac{\alpha_o^g + \gamma^g F_o + \bar{W}_o^g}{\sigma_\eta^g} \tag{11}$$

5 Preference for Gender Ratio

Recall that in a logit model the mean utility in occupation o is equal to the log of the ratio of the share employed (s_o^g) in that occupation to the share in non-employment (s_N^g), or $ln(s_o^g) - ln(s_N^g)$. The mean utility can then be decomposed into wage and non-wage utility which are common to all worker and job matches of a given occupation and gender. The non-wage component is $\bar{u}_o^g$ and the wage component $\bar{W}_o^g$. The impact of both on mean utility is mediated by the scale of the taste heterogeneity (σ_η^g).

$$ln(s_o^g) - ln(s_N^g) = \frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g} = \frac{\bar{\alpha}_o^g + \gamma^g F_o + \bar{W}_o^g}{\sigma_\eta^g}$$

This equation cannot be estimated in the cross section because we have one α_o^g parameter for each occupation, in addition to γ_o^g and σ_η^g, and only n_o moments.

5.1 Overview of Endogeneity and Instruments

In the sections below I introduce several strategies to identify the parameters $\frac{\gamma^g}{\sigma^g_\eta}$ and $\frac{1}{\sigma^g_\eta}$. First, I pool all six cross sections of data (1960-2012), allowing me to estimate the occupation-specific intercepts, α^g_o and use time variation in fraction female and reservation wages to identify coefficients $\frac{\gamma^g}{\sigma^g_\eta}$ and $\frac{1}{\sigma^g_\eta}$. Unfortunately using time variation means that estimates of $\frac{\gamma^g}{\sigma^g_\eta}$ and $\frac{1}{\sigma^g_\eta}$ likely suffer from omitted variable bias. Changes over time in both the gender ratio and the reservation wage are likely correlated with changes in unobserved occupation attributes.

For example if an occupation is becoming more family friendly over time, and this causes more women to enter the occupation, the coefficient on fraction female will be biased upward for women. Similarly compensating differentials would lead us to expect a downward bias on the coefficient on the reservation wage W^g_o. For both men and women if occupation amenities deteriorate over time, this may be correlated with increases in wages.

Therefore in my second estimation strategy I introduce Bartik-style instruments that exploit variation in the industry composition of occupations to isolate changes over time in the fraction female and wage by occupation that are caused by labor demand shifters. The use of instruments should isolate labor demand shocks, thus tracing out the labor supply curve conditional on fraction female and wage level.

The identifying assumptions for the first set of instruments are that changes in industry wage levels and gender ratios over time are uncorrelated with changes in how workers value industries, and that changes in occupation attributes are independent across occupations. The identifying assumptions for the second pair of instruments is that changes in industry size over time that are correlated with initial wage and gender ratios, are uncorrelated with changes in the worker valuation of the industry. Lastly I introduce an instrument that interacts the initial gender ratio by occupation and the relative growth rates of men's and women's labor force participation over time. The identifying assumption is that changes in unobserved occupation attributes are independent across occupations.

5.2 Fixed Effects Specification

Recall from equation 9 that worker utility can be written as a function of the share of workers by gender in each occupation and in non-employment:

$$ln(s_o^g) - ln(s_N^g) = \frac{\alpha_o^g + \gamma^g F_o + \bar{W}_o^g}{\sigma_\eta^g} \tag{12}$$

In order to better match time variation in shares, a time effect β_t^g is added. In addition, in the spirit of Berry (1994), $\epsilon_{o,t}^g$ represents changes over time in the utility of workers due to changes in unobserved occupation attributes, so changes not due to movement in the fraction female or the reservation wages.

$$ln(s_{o,t}^g) - ln(s_{N,t}^g) = \frac{\beta_t^g + \alpha_o^g + \gamma^g F_{o,t} + \bar{W}_{o,t}^g + \epsilon_{o,t}^g}{\sigma_\eta^g}$$

One concern with estimating this equation is that, leaving aside the structural interpretation of the discrete choice model, it is unclear if the coefficients reflect labor demand or labor supply. Skill requirements of occupations for example is one factor, unmodeled on the firm side, that could lead to a spurious negative correlation between the wage and the share of workers in an occupation. Simultaneously, there is the problem of omitted variable bias due to correlation between time varying occupation amenities $\epsilon_{o,t}$ and $F_{o,t}$ and $W_{o,t}^g$ discussed above. Ideally instruments driven by labor demand will also be uncorrelated with $\epsilon_{o,t}$ and therefore solve the endogeneity problem.

5.3 Instrumental Variables

5.3.1 Changes Over Time in Industry Wage and Gender Ratio

The first set of instruments uses industry variation in wages and gender ratios, and variation in the presence of industries within occupations, to predict occupation level wages and gender ratios. For example, the reservation wage for administrative assistants will be the weighted sum of the wages by industry for all industries that exist in that occupation.[14]

The predicted fraction female in occupation o in time t is as follows:

$$\hat{F}_{o,t} = \sum_I p_{Io} * \hat{F}_{Io,t}$$

Where p_{Io} is the fraction of occupation o in industry I, and $\hat{F}_{Io}$ the fraction female in industry I excluding workers in occupation o.[15] The predicted wage is similarly the sum over industry of the industry wage by the industry composition. Fraction female is measured as the fraction female in the older generations (ages 36-65) in the Census cross section. Industry wages are measured by gender as lifetime incomes of those beginning their careers in a given industry as simulated using the Census and SIPP data described above.

Pooling all six waves of data (1960-2010) is critical to the quality of the instruments in two ways. First, the industry composition can be fixed in 1950, prior to the sample data. This alleviates any concern that changes in industry composition are correlated with changes in unobserved occupation amenities, which is critical for the validity of the instrument (Goldsmith-Pinkham, Sorkin, & Swift, 2017). Second, occupation fixed effects absorb the mean utility coming from the industry composition. Therefore we

[14] Industries used are aggregates of the harmonized IPUMS codes of ind1990. Industries are as follows: Agriculture, Forestry and Fisheries; Mining; Construction; Manufacturing; Transportation, Communications, and other public utilities; Wholesale Trade; Retail Trade; Finance, Insurance, and Real Estate; Business and Repair Service; Personal Services; Entertainment and Recreation Services; Professional and Related Services; Public Administration.

[15] Note that in order for industry level fraction female or wage level to be estimated excluding occupation o, it is necessary that each industry contain multiple occupations. In fact at the level of aggregation I use, 14 major industries, every industry has employment in almost all occupations.

must only assume that *changes* in how workers value industries are independent across occupation, but changes in industry demand are correlated across occupations. For example, if wages in manufacturing are going up over time because the production technology in manufacturing has become more efficient, then we would expect this to have an impact on the wages of workers in occupations employed in manufacturing that is due to labor demand and independent of occupation amenities.[16] To avoid contaminating the estimate of industry wages and fraction female with changes in occupation amenities, they are calculated excluding the occupation that is being instrumented for.

The instrument will be invalidated if changes in industry amenities are correlated across occupation. For example, if the working conditions for all workers in manufacturing are getting worse over time, and this is causing the working conditions for administrative assistants in the manufacturing sector to get worse over time, then the instrument will predict an increase in wages among administrative assistants that is correlated with the change in working conditions.

5.3.2 Growth in Industry Employment

I also use a more standard Bartik IV exploiting change over time in the the size of industries to instrument for the reservation wage and the fraction female. The growth rate of industries since 1950 is the treatment and the initial wages and industry fraction female[17] are the exposure to the treatment. Changes in employment by industry is a fundamentally different source of variation. The instrument predicts changes in occupation wage and gender ratio over time due to the prominence of certain industries due to production technology or demand side factors, not including the impact of these factors on changes in wage or gender ratio since these are fixed in the initial period, and therefore controlled by the occupation fixed effects.

[16] The instrument is similar in style to Autor, Dorn, and Hanson (2013) where import demand is assumed uncorrelated across countries but import supply correlated across countries. It is also similar to Hausman instruments where product cost shocks are assumed correlated across markets but demand shocks are independent (Hausman, 1996; Nevo, 2001).

[17] Industry composition by gender is set in 1950. Wages are set in 1960 because the sample size in the 1950 Census is too small to calculate lifetime wages

The predicted occupation fraction female is

$$\hat{F}_{o,t} = \sum_{I} p_{Io,initial} * F_{Io,initial} * \frac{size_{Io,t}}{size_{Io,initial}}$$

where $p_{Io,initial}$ and $F_{Io,initial}$ are the occupation*industry composition and fraction female respectively, which are fixed at the initial period, and the $size_{Io,t}$ is the total employment level in industry I excluding occupation o in year t. The changes over time in the predicted $\hat{F}_{o,t}$ are driven by changes in industry size. I exclude own occupation in the estimation of industry size.

One requirement for validity of the instrument is that the growth in the size of industries, and therefore the exposure of occupations to the wages and gender ratios predicted by those industries, is due to product demand or productive efficiency changes over time. So for example say that demand for manufacturing is growing over time and therefore more jobs are available in the occupations in manufacturing. We would expect the wages and gender ratios in occupations found in manufacturing to be more and more reflective of the growing prevalence of manufacturing.

The main threat to validity of the instrument is if the wages and gender ratio in manufacturing are reflective of underlying amenity values that are common across all occupations in manufacturing. In that case, even if the growth of an industry is exogenous, its increased prevalence in an occupation will be correlated with a change in the amenities in that occupation. This is less important to the extent that my paper is concerned with occupation amenities not industry amenities.

5.3.3 Changes Over Time in Employment by Gender

Another source of variation made possible by the panel is variation in the relative value of employment and non-employment for men vs. women over time. Fixing occupation gender ratios in the initial period, the instrument predicts changes in occupation gender ratios due only to the gender ratio of those employed in all other occupations. Assuming changes in occupation attributes are independent across occupation, growth in the ratio of employed to non-employed in all other occupations will not be correlated with growth in amenities in the given occupation.

The identifying variation is changes in the relative value of home vs. work by gender, projected onto current occupation gender ratios based on previous gender ratios. This is similar to instrumenting for immigration patterns based on overall flows of immigrants and initial shares (Altonji & Card, 1991). One might expect initial fraction female to impact flows into an occupation through occupation attributes that are fixed over time. Since fixed attributes are controlled by fixed effects, variation over time is assumed to be due only to changes in labor force attachment of men relative to women relative to the initial period. For recent evidence of changes in labor force attachment by gender see Albanesi and Sahin (2017).

The instrument may be invalid if the changes in labor force attachment are driven by changes in occupation amenities, and these changes are correlated with initial amenities. So for example we see a high fraction female for teaching non-postsecondary in 1960 and also growth in the number of female teachers over time. Under the assumption of the instrument the growth in the number of female teachers is due to an overall increase in female labor force participation and the fixed amenities component of teaching being attractive for women. However, if the increase in female labor force participation is driven by growth in the amenity value of teaching, and the amenity value growth is correlated with the initial amenity value, then the instrument is endogenous.

The instrument is constructed as follows. Let $F_{o,initial}$ be the fraction female in the occupation in the initial period. Let the number of men and women employed in all occupations except o in time period t be $\#M_t$ and $\#W_t$.

Define the relative growth in female vs. male employment r_t as

$$r_t = \frac{\frac{\#F_t}{\#M_t}}{\frac{\#F_{initial}}{\#M_{initial}}}$$

Then the fraction female in occupation o predicted by the instrument in time t, $\hat{F}_{o,t}$ is as follows:

$$\hat{F}_{o,t} = F_{o,initial} * r_t$$

5.3.4 Firm Willingness-to-Pay Parameters

Since the first stage F statistics are low, I additionally include the willingness-to-pay estimates WTP_o^g from the job side of the model, as instruments. The WTP_o^g are a measure of how much firms value workers should be uncorrelated with unobserved amenities, assuming amenities are fixed, and therefore are a good proxy for labor demand side factors that will shift the reservation wage. The ratio of $\frac{WTP_o^M}{WTP_o^F}$ is also used as a proxy for labor demand.

6 Data

The primary data elements needed to estimate the model described above are: expectations of lifetime labor income by occupation and gender, shares of workers by gender and age cohort choosing each occupation and non-employment, a measure of unfilled jobs by occupation (for some specifications), and occupation attributes (for some specifications).

Lifetime wages and shares of workers by gender and occupation are estimated using a combination of Census data and SIPP data.[18] Occupation codes are constructed by aggregation of the IPUMS harmonized codes (occ1990). Aggregation is done to achieve sufficient sample size. See appendix 10.4 for the full list of occupation codes.

While the Census and ACS provide cross sectional wage and occupation, the SIPP is needed to provide a panel for the construction of lifetime labor income. While using cross sectional data may be sufficient to create estimates of lifetime income when only a few moments are needed, for my identification strategy the shape of the wage distribution is critical, and I therefore cannot assume that every worker

[18] Public use Census 1960, 1970, 1980, 1990, 2000, and 2012 three-year ACS data obtained from IPUMS (Ruggles, Genadek, Goeken, Grover, & Sobek, 2015). SIPP data from the 2004 and 2008 panels are constructed using the NBER files (U.S. Census Bureau, 2017).

obtains, for example, the median income for their chosen occupation at each age. The resulting distribution is mechanically quite lumpy. One reason that the lifetime income distribution in reality might be much smoother than a cross-sectional approximation is that workers transition stochastically over time between wage levels and occupations.

To model worker transitions over time as probabilistic, panel data from the SIPP is essential. Using pooled data from the 2004 and 2008 panels, which are four and five years long respectively, I construct transition rates through five quantiles of earnings and occupations by worker age and gender. Because of limited sample size, I assume that transition rates are the same for each five year period starting at age 25, and depend only on the current state not the past history (first order markov assumption). The five quantile cut offs are estimated in the Census data where the larger overall sample size allows for more accuracy.

Income quantiles from Censuse, by age, gender, and occupation, are denoted *quan* below. Occupation is denoted *occ*, and five year age bracket *age*. Transition rates from year to year, where x' indicates the value of x in the following year, are then estimated non-parametrically as follows:

$$Pr(quan', occ'|quan, occ, age) = \frac{\sum_i \mathcal{I}(i \in quan', occ', quan, occ, age)}{\sum_i \mathcal{I}(i \in quan, occ, age)}$$

This transition matrix is then used to simulate worker career paths from the starting point of workers aged 25-35 in the Census and ACS.[19] The lifetime income assigned to each observation in the Census or ACS is the sum of their simulated career path through the SIPP transition matrix. For a comparison of estimated lifetime income paths to observed lifetime income paths in the PSID, see appendix **??**. Their assigned choice of occupation is taken as the occupation they start out in at ages 25-35 as observed in the Census. This avoids measurement error in occupation that could come

[19]CPS or CPS MORG data could be added to get an earlier estimate of these transition rates.

from simulated occupation choice, and allows for the interpretation that the occupation choice early in life includes the expectation of all future transitions.

The Job openings and Labor Turnover Survey (U.S. Bureau of Labor Statistics, 2017) is used to construct a measure of unfilled jobs by occupation. Since JOLTS does not directly contain occupation, only NAICS industry codes, industries are projected into occupations using contemporaneous occupation industry shares estimated in CPS (Flood, King, Ruggles, & Robert, 2015). The estimated openings by occupation is then divided by the total number of people employed in the occupation to get the ratio of openings to employed.

I also make an estimate of the number of openings that remain open at the end of the year by assuming that the probability that a job is filled is uniform across time and jobs. Then the daily rate of hiring is equal to the total number of hires that month divided by the number of days in the month. Then the probability a job is not filled on a given day is one minus this daily hire rate, and the probability a job is unfilled for the year is this probability to the power of 365. This measure of unfilled jobs will be used for robustness.

To assess the impact of my assumption that workers make lifetime occupation choices, I examine the extent to which workers switch occupations during their working lifetime. In my 1960 data, simulated from the Survey of Income and Program Participation (SIPP), 14% of workers change occupation each year on average.[20] However the statistic that matters is how many workers spend many or most of their working years in the same occupation, mirroring the lifetime occupation choice dictated by the model. In the PSID, the average worker who spends most years working spends 80% of working years in the same occupation[21] In my simulated 1960 SIPP sample this number is only 64%. In reality dependence on past occupation likely extends beyond the immediate previous period, so this number is likely a lower bound due to my first-order

[20] Kambourov and Manovskii (2008) find that on average 13% of workers change occupation each year, by comparing retrospective and concurrent occupation data in the Panel Study of Income Dynamics (PSID). In my unadjusted PSID sample this number is 21%.

[21] I observations from 1968-2011 that appear in the PSID for at least 25 years beginning before age 30 and ending after age 55, and are not missing occupation data, which results in a sample of 764 workers.

Markov assumption in simulating the data. Furthermore, for 85% of workers in the PSID and 57% of workers in the simulated SIPP, the modal occupation for ages 25-35 is also the modal occupation for ages 25-55.

7 Model Estimates

7.1 Reservation Wages and Willingness-to-pay

Unlike the mean wage by occupation, the latent reservation wage $\bar{W}_o^g$ controls for the fact that in observed wages, we see only the most attractive jobs filled in each occupation. $\bar{W}_o^g$ is unconditional on selection effects, and reflects the common component of the reservation wage to every worker of gender g at any job in occupation o.

In Table 1 we see the "Reservation Wage Gap", which is the female to male ratio of reservation wage common components, "WTP Gap", which is the ratio of willingness-to-pay parameters, and "Lifetime Income Gap", which is the ratio of estimated lifetime income. All values are averaged across year by occupation. The full set of job side model parameter values and histograms of the model fit of the lifetime income distributions and shares can be found in Appendix Section 10.7.

In general women have lower reservation wages than men ($\frac{\bar{W}_o^F}{\bar{W}_o^M} < 1$) and are less valued by jobs ($\frac{\overline{WTP}_o^F}{\overline{WTP}_o^M} < 1$). In general the higher the fraction female, the larger the gap between male and female reservation wages. On the other hand occupations with a high fraction female tend to have a smaller gap between male willingness-to-pay parameters. This is consistent with a Roy model with women sorting into occupations in which they are more valued by jobs.

It has been noted that both men and women have lower wages the higher the female share in an occupation (see eg. Macpherson and Hirsch (1995); Levanon et al. (2009); Addison, Wang, and Ozturk (2017); Harris (2018)). I do not find a statistically significant correlation between my estimates of lifetime income and average fraction female in an occupation. However, I do find statistically significant correlations between estimated reservation wages and the occupation fraction female. I find that the centers of the reservation wage distribution for women ($\bar{W}_o^F$) are negatively correlated with the fraction female (correlation coefficient -0.77), but for men $\bar{W}_o^M$ are positively correlated with the fraction female (correlation coefficient 0.6).

Reservation wages are more closely related to the value that the worker has for an occupation than observed wages, since observed wages reflect additional selection by the firm. This explains why reservation wages could be so strongly correlated with the fraction female, but lifetime income not so much. Although reservation wages for men in female-dominated occupations are quite high, in the observed wage data we see only those jobs within the occupation that are so attractive to men that the men can be hired an comparable wages to women. In contrast, the higher the fraction female in an occupation, the lower the reservation wage to women $\bar{W}_o^F$, but since women are cheap to hire relative to men, we see even the least attractive jobs being filled by women. The lower common component of reservation wage for women in female-dominated occupations is consistent with a female preference for working with women, and the higher reservation wage for men in female occupations is consistent with a male preference against working with women.

7.2 Gender Preference Results

I find a strong and robust preference on the part of women for entering into more female-dominated occupations, which is around four times the preference for log wages, meaning that if the log reservation wage in an occupation went up by 10% this would have an equivalent effect on log utility ($\bar{u}_o^F$) of an increase to the percent female in the occupation of around 5%. I do not find evidence for a preference on the part of men against entering into more female-dominated occupations.

Graphical representation of the preference structure under the assumption of linear, quadratic or cubic form is given in Figure 3. Women have increasing utility in the fraction female, and the increase is steeper the fewer women there are in the occupation. By contrast for men, utility is relatively flat in the fraction female and not statistically different from zero.[22]

Results for the industry and employment growth panel instruments are in Tables 2 and 3. All tables report fixed effects regressions on the pooled data. Estimation is done using limited information maximum likelihood for robustness to weak instruments, but two stage GMM results are similar. Standard errors clustered at the occupation level.[23] I report only the linear specification for men and the cubic specification for women because these seem to fully capture the functional form and obtain the highest first stage F statistics in the preferred IV specification.

The first column is an un-instrumented fixed effects regression. The second column (IV1) includes the instruments using variation in industry wage and fraction female over time. The third column (IV2) includes these instruments and subsequently includes instruments from the variation in occupation size over time. The fourth column (IV3) also includes all previous instruments and also adds in the instrument exploiting variation in the size of the male and female labor force relative to the initial period. The last column (IV4) additionally includes the firm willingness-to-pay parameters WTP_o^g for men and women respectively and the ratio of firm willingness-to-pay parameters $\frac{WTP_o^M}{WTP_o^F}$. This final column is the preferred specification because it has the strongest first stage.

In the un-instrumented regression in the first column, both men and women have a negative wage coefficient. This implies that even though the fixed effects have controlled for any time invariant omitted variables, the changes over time may still be picking up unobserved labor demand factors, changes in occupation attributes, or other omitted variables. The instrumented specifications should avoid this endogeneity by identifying off of labor demand shocks. The wage coefficient is positive for men and women in all instrumented specifications. The last column, using all instruments discussed above, achieves an Kleibergen-Papp F statistic of around 10 for women and 8.6 for men.[24]

[22] Results from a beta distribution specification follow a similar pattern.

[23] I expect errors correlated within occupation due to occupation fixed effects and possible differences in model fit across occupations.

[24] Denoted “KP rk F” in the tables.

In no specification do men have a significant preference over the fraction female. The instrumented point estimates are close to zero in the preferred specification, but the standard errors are high, so I cannot rule out moderate effects in either direction. The lack of precision could be due to a lack of variation over time in men's labor market outcomes, which is corroborated by the much higher total sum of squares in the female regression.

For women the estimated preference is economically significant in terms of the impact on occupation choice. Moving the fraction female in a single occupation from 20% ($F = .2$) to 80% ($F = .8$) would have an average marginal effect of moving 1352% more women into that occupation under the $IV4$ specification (see Table 4). However this effect is exaggerated because it is partial equilibrium. Holding equilibrium wages fixed means that this result only reflects the decisions of workers without firm response. When firms are allowed to adjust wages in equilibrium, the effect of moving from $F = .2$ to $F = .8$ is only 124% on average. So if an occupation moved from relatively male dominated to relatively female dominated that would just over double the number of women who would enter that occupation in equilibrium.

To check the sensibility of the model it is useful to look at the estimated fixed effects from the regressions in Tables 2 and 3. These correspond to the estimated non-wage utility of men and women for each occupation. The occupation fixed effects by gender can be found in Table 10.2. The excluded category is "Teachers, Postsecondary". In combination with the estimated willingness-to-pay from Table 1, they track relatively closely with the observed fraction female by occupation. Occupations with higher female utility and higher willingness-to-pay for women have more women in them.

8 Counterfactuals

The dynamics of the model result from updating of the fraction female in each occupation with each successive cohort of workers. In each ten year period, only young workers ages 25-34 choose a new occupation, while older cohorts in age brackets 35-44, 45-54, 55-65 are fixed in their occupation. When the current cohort of workers, ages 25-34, make occupational decisions, they face the gender ratio produced by older cohorts of workers. The occupation choice of each worker is fixed for the rest of their working lifetime (assumed to be 4 periods, or 40 years), and workers do not take into account predicted future evolution of the fraction female.

Let the fraction female in an occupation observed by the young cohort in time t before making their decisions be F_t, where the occupation subscript $_o$ is omitted for convenience. Then we can write F as a function of the number of men and women (n^F and n^M) choosing that occupation in the previous three periods.

$$F_t = \frac{n^F_{t-1} + n^F_{t-2} + n^F_{t-3}}{(n^F_{t-1} + n^M_{t-1}) + (n^F_{t-2} + n^M_{t-2}) + (n^F_{t-3} + n^M_{t-3})}$$

Given this observation of F_t in each occupation, the young cohort makes their own occupation decisions. Their decisions determine the number of men and women in the occupation in time t, or n^F_t and n^M_t, which will influence the fraction female observed by the next three cohorts of workers.

8.1 Future Cohorts under Status Quo

In Figures 4-7 below, I simulate future generations of workers to determine if the current level of sorting is stable, and if not, to what fraction females the occupations will converge in the future. For counterfactual simulations I assume there are no changes to parameters, which isolates the impact of the endogenous movement of the fraction female across generations. In particular, I assume that the willingness-to-pay parameters WTP^g_o on the firm side are fixed, and that worker utility only changes due to changes in the fraction female F_o, and changes to the equilibrium wage. All other inputs are fixed at the 2012 values, the last year of data.

In the top right panels of Figures 4-7, I assume that the wage levels by occupation are fixed over time, simulating the outcome in a one-sided model of occupation choice. In this case, all occupations eventually converge to fully male or female due to the preference for fraction female.

In the bottom panel I allow the wages to update through the market clearing condition of the matching model described above. I solve for the new $\bar{W}_o^{*g}$ the equates the supply and demand for workers and jobs, given the changes to worker utility due to the evolution of fraction female.

In equilibrium, there can be no workers or jobs that want to match but cannot. Recall that n_o is the number of jobs in occupation o, and n^g the number of workers of gender g. Then for all occupations o and gender g the $\bar{W}_o^{*g}$ must solve

$$n^g * Pr(i \in g \text{ chooses } o) = n_o * Pr(j \in o \text{ chooses } g)$$

which in terms of the choice probabilities defined above is

$$\begin{aligned} n^g \frac{exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})}{\sum_o exp(\frac{\bar{u}_o^g + \bar{W}_o^g}{\sigma_\eta^g})} =& n_o Pr(\overline{WTP}_o^g - \bar{W}_o^g + \bar{\xi}_j^g \geqslant \overline{WTP}_o^{-g} - \bar{W}_o^{-g} + \bar{\xi}_j^{-g}) \\ & * Pr(\bar{\xi}_j^F \geqslant -\overline{WTP}_o^F + \bar{W}_o^F) * Pr(\bar{\xi}_j^M \geqslant -\overline{WTP}_o^M + \bar{W}_o^M) \end{aligned}$$

In with equilibrium wages clearing the market, there is much less movement in segregation patterns[25] because most occupations are very close to their unique equilibrium. However the model does predict changes in the gender composition of some occupations. Occupations predicted to become more than 10% more female in the long run include Health diagnosing occupations (38% to 60%), engineers, architects and surveyors (17% to 30%), math computer and natural science (32% to 43%), and precision production occupations (25% to 38%). Male occupations predicted to become less female in the future include various machine operators fabricators assemblers testers

[25]Although the male preference for working with men is weakly identified (and not statistically significant) these predicted future patterns hold broadly even if men are given a coefficient on fraction female at the lower bound of the confidence interval. This is the case in which we would assume men would be most likely to flee female fields and reinforce tipping.

(20% to 5%) and metal wood plastic print textile (30% to 10%).

The fit of the model can be assessed by comparing the top left panel, which is observed fraction female by occupation in the Census data,[26] to the right and bottom panels, which are model simulations. Note that the model simulations do not exactly match the Census data patterns in the top left because the dynamic updating of the fraction female across cohorts was not a moment targeted by the model. In the data workers also do not necessarily stay in their starting occupation for their lifetime.

8.2 Initial Parity

If preferences over an endogenous amenity like the fraction female are strong enough, there could be multiple equilibria. In this case long run sorting patterns could depend on the initial conditions, or historical segregation, which would determine which equilibrium is selected. As a first pass at testing this hypothesis, I begin my simulation with all occupations at 50% female in the initial year of 1960. The figures 8 through 11 show that this does not affect the long run outcome. The patterns are shockingly similar today as if occupations had begun from the observed segregated position.

This result suggests that we are currently in a stable equilibrium, and that parity is close enough to this stable equilibrium that it leads to convergence. It could also suggest that each occupation has in fact only one stable equilibrium in the fraction female, and would converge to this equilibrium regardless of any starting point. In the next section I further explore this question by searching for all equilibria in the fraction for each occupation separately.

8.3 Equilibria by Occupation

I search for all equilibria in the fraction female for each occupation individually. To do so I graph the mapping between fraction female in the current period and the next period. Recall that fraction female chosen today affects the choices in the next period

[26] Both observed and simulated fraction females are for age ranges 35-64

through women's preference for higher fraction female. By examining these graphs it is easy to find fixed points, where the fraction female this period is the same as the fraction female in next period, by observing intersections with the 45 degree line. I fix all attributes of other occupations, including the fraction female, to focus only on the occupation at hand. I graph ten equidistant starting points between 0% female and 100% female using 2012 parameter values, and show a fitted line through these points on the graphs.

I find that every occupation has one unique equilibrium in the fraction female. The equilibria are close to the observed 2012 values of fraction female, and where there is deviation from the observed fraction female they are close to the long run equilibria shown in Figures 4 through 7 where the fraction female is allowed to evolve in all occupations at once. The transition graphs for some selected occupations are shown in Figures 13 through 16.

If the preference for working with women is strong enough, as women leave a male occupation female wages will rise enough to price women out of the occupation and produce a 0% female equilibrium. Because I do not estimate a preference on the part of men to work with men, they will not be similarly priced out of female occupations, so a 100% female equilibrium is unlikely. As a result with a female preference, I would expect occupations to have equilibria at either 0% female, somewhere in between 0% and 100% female, or both. Figure 29 illustrates the stylized model for the case in which the gender preference produces both an all male and a mixed equilibrium (mixed meaning between 0% and 100% female).

I estimate that all occupations have only the mixed equilibrium. This is because the gender preference is not strong enough to cause female labor supply to actually cross male labor supply, meaning that employers are always able to hire a few women who really love the job for cheap, making a 0% female occupation unsustainable.

Furthermore, only a few occupations are characterized by very male or female dominated equilibria. At greater than 80% female we have only Health Services Occupations with a fixed point at around 87% female, and Health Assessment and Treating and

Therapists at around 85% female. At under 20% female we have Agriculture Forestry and Fishing at around 12% female, Machine Operators Fabricators Assemblers and Testers at around 15% female, Road Rail and Water Transportation at around 10%, and both Mechanics and Repairers, and Construction and Extraction, at around 5% female.

8.3.1 Conditions for Multiple Equilibria

Although the estimated parameter values produce only one equilibrium in the fraction female for each occupation, the model does allow for multiple equilibria to emerge. Below I explore two scenarios that could lead to multiple equilibria in the fraction female. First, doubling the magnitude of the preference for women to work with women, and second, fixing equilibrium wages so they are not allowed to adjust and form compensating differentials.

I find that doubling the preference over the fraction female produces multiple equilibria in some occupations, such as Postsecondary Teachers, and Engineers Architects and Surveyors. These occupations have one equilibrium at close to 0% female and another at majority female. Intuitively, a very strong gender preference makes it cheapest for an occupation to hire either all men or majority women. The initial fraction female would determine to which equilibrium the occupation converges.

Next I fix the wage so that the gender composition reflects only labor supply, and the fixed wage will no longer be able to clear the market as the fraction female changes. In the case of fixed wages, 27 of the 34 occupations have two stable equilibria as opposed to one. Figures 23 through 26 show the same set of occupations as Figures 13 through 16 but with fixed wages. In all of these examples we have an equilibrium that is majority female, and an equilibrium that is close to zero percent female, implying that allowing wages to vary freely to equate supply and demand plays a key role in eliminating multiple equilibria..

The wage moderates the impact of the preference of women to enter more female occupations, as it would over any endogenous amenity. As more women enter an occupation it becomes more attractive, but at the same time wages go down as employers are able to attract more women at lower cost, thus ultimately dampening the supply of women to the occupation. Likewise as women leave an occupation, the wage offered to women in that occupation goes up, which increase female labor supply to that occupation and dampens the movement towards 0% female.

8.4 Impulse Response Examples

To take a closer look at how wages adjust to compensate for the preference over fraction female, I plot the response of wage to a change in the fraction female in two case studies. Specifically I set a female dominated occupation to be 0% female, and separately, a male dominated occupation to be 100% female, and observe how wage adjustment facilitates convergence back to the unique stable equilibria, which occurs after about eight cohorts of workers.

In the first simulation I set nursing ("Health Technologists and Technicians") to be 0% female in 1960. In reality nursing was close to 100% female in 1960. If wages were fixed, 0% female would be a stable equilibrium for nursing. However, the gender preference is not strong enough to completely price women out of nursing once wages adjust to reflect labor demand. It is still cost-effective for firms to hire some women who really love nursing jobs, and so the female wage offer rate must be set high to equate supply and demand by compensating women for the disutility of the low fraction female.

The wage adjustment process for the simulation is shown in Figure 27. Solid lines are simulated counterfactual wage offer levels, while dotted lines are actual estimated wage offers. We can see that although simulated female wages in nursing start out way higher than reality due to the compensating differential, they very quickly drop as nursing feminizes.

Women with particularly high utility from nursing are enticed to enter the occupation by the high wages, which in turn makes nursing more attractive for the next cohort. This in turn lowers the simulated reservation wages of the next cohort of workers, which also makes women cheaper for firms to hire. This process continues until nursing is female-dominated and simulated wage offers have dropped to the levels estimated in the data.

In the second example, I set “Mechanics and Repairers”, a male dominated occupation, to be a 100% female in 1960. If wages were fixed, the occupation would then converge to its majority female stable equilibrium at around 80% female. However, with wage adjustment, which can be seen in Figure 28, female wages quickly skyrocket as more and more men start to become Mechanics and Repairers. After about eight cohorts, women are no longer affordable to hire and the occupation has converged to its unique stable equilibrium at around 0% female.

9 Conclusion

“We’ve got half the population that is way underrepresented in those fields [math, science, and engineering] and that means that we’ve got a whole bunch of talent...not being encouraged the way they need to.” (President Obama 2013)[27]

It is an open policy question as to what sort of encouragement would lead more women to enter male occupations and vice versa. Ideally we might see that a few men or women entering a field might lead to a flood of followers. I find women do prefer to go into occupations that already have more women. However, womens’ preference is not strong enough that simply putting more women in a field will lead more women to enter in the long run. I also find no evidence that men prefer to enter occupations that already have more men.

The reason that putting more men or women in an occupation has no long-run effect on segregation is that wages are free to adjust. As the fraction female goes up, wages for women go down, which slows the entry of women into the occupation. The level of gender preference would have to be two times as large as I estimate in order to overcome the tendency to converge back to the original sorting pattern through wage adjustment.

[27] https://obamawhitehouse.archives.gov/administration/eop/ostp/women

The preference on the part of women to work with women increases gender segregation by creating a feedback loop that amplifies the impact of gender differences in labor supply and demand. According to my simulations, the"tipping" patterns documented by Pan (2015) might be the result of changes in the perceived productivity of men and women, compounded by a feedback loop from the preference of women to work with women. This feedback loop mechanism also exacerbates the gender wage gap by lowering wages in female dominated occupations through compensating differentials.

The estimates of this model are likely imperfectly predictive of the future of occupation gender segregation, but the model does prove useful for learning about the role of wages in two-sided matching with an endogenous amenity. This paper focused on the fraction female in occupations, but future work could use this model to look at race, age, or any other group preference or endogenous amenity, and the moderating effect of price adjustment.

In addition, although I do not find evidence of tipping between multiple equilibria given my estimated parameter values, it is clear that this could occur under different circumstances, such as a stronger gender preference or stickier wages (for example equal pay for equal work laws). Future work is needed to fully understand occupation gender segregation, and could benefit from considering how tipping is mitigated by compensating differentials.

10 Appendix

10.1 Figures

Figure 1: Sales Representatives, Finance, and Business Services: Observed vs. Model Reservation Wages

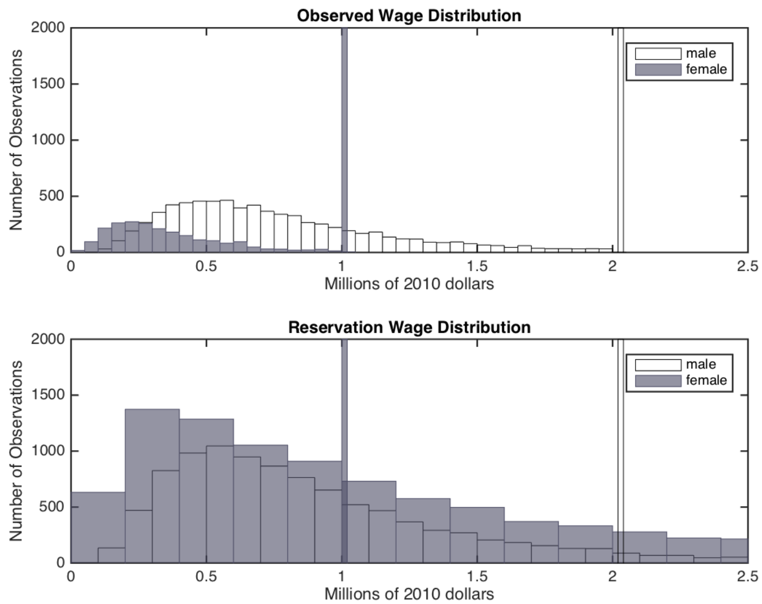


Figure 2: Health Service occupations: Observed vs. Model Reservation Wages

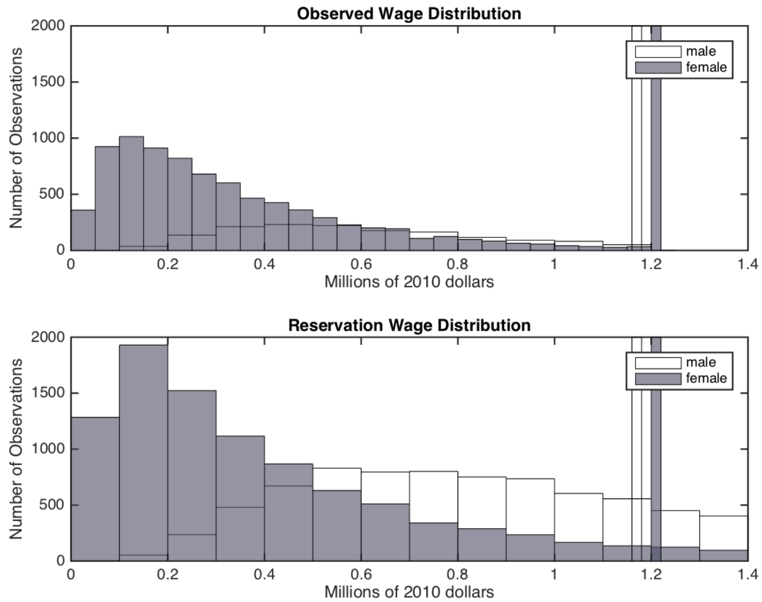
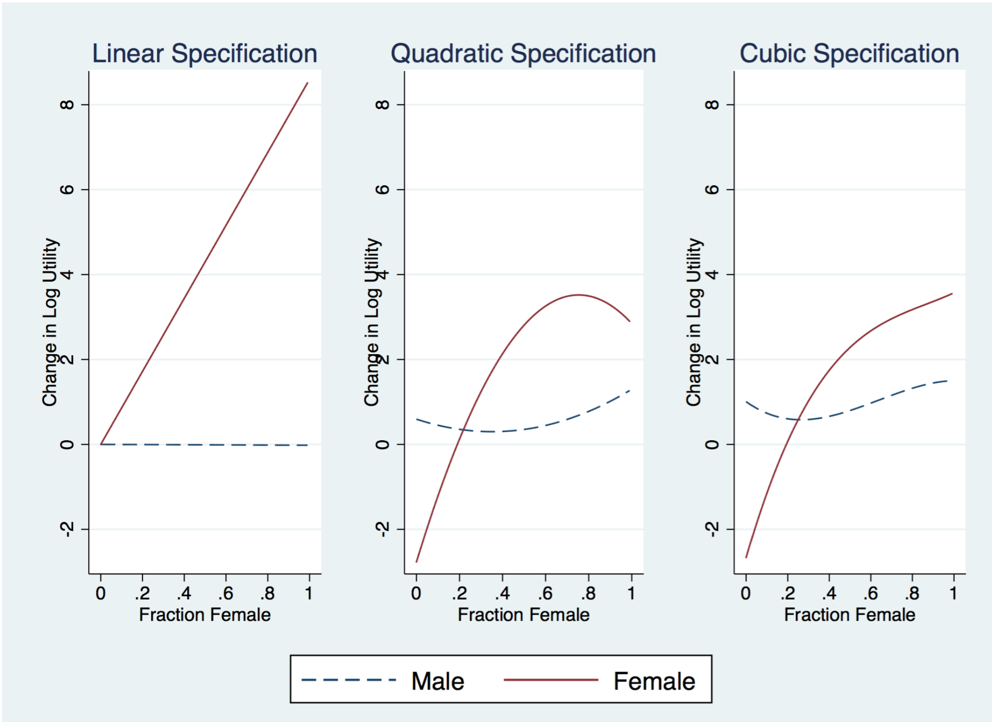


Figure 3: Male and Female Log Utility by Fraction Female in Occupation



Results of instrumental variables regressions with occupation fixed effects on 6 waves of Census and ACS data (1960-2012), 34 occupations, using reservation wages estimated earlier using MLE.

Figure 4: Status Quo: Simulated Occupation Segregation Patterns

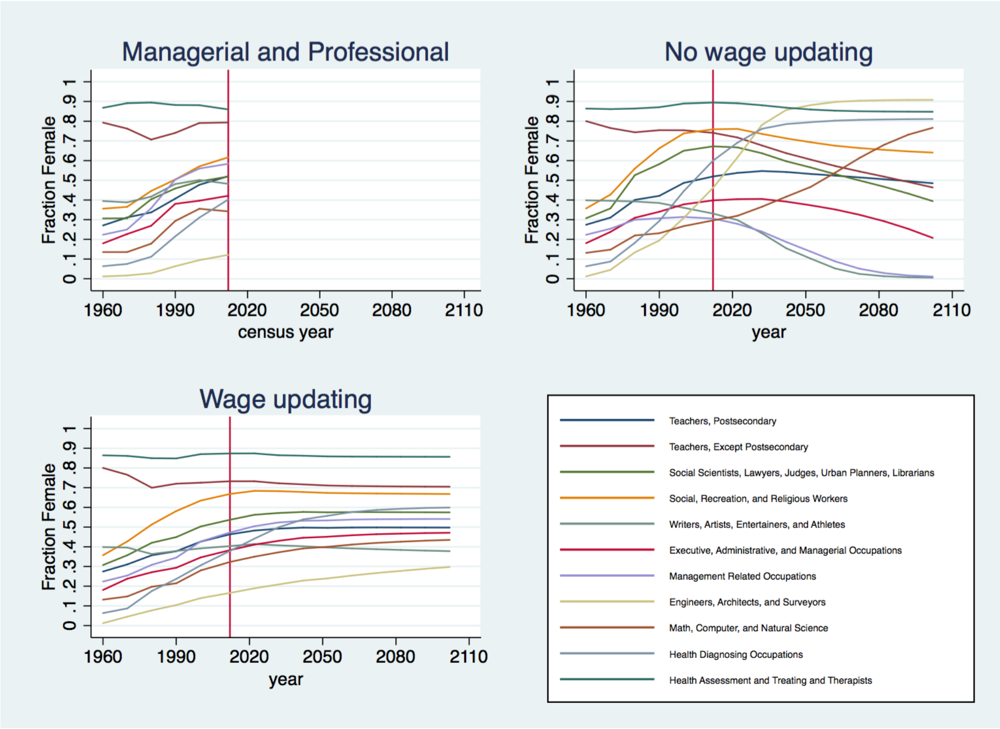


Figure 5: Status Quo: Simulated Occupation Segregation Patterns

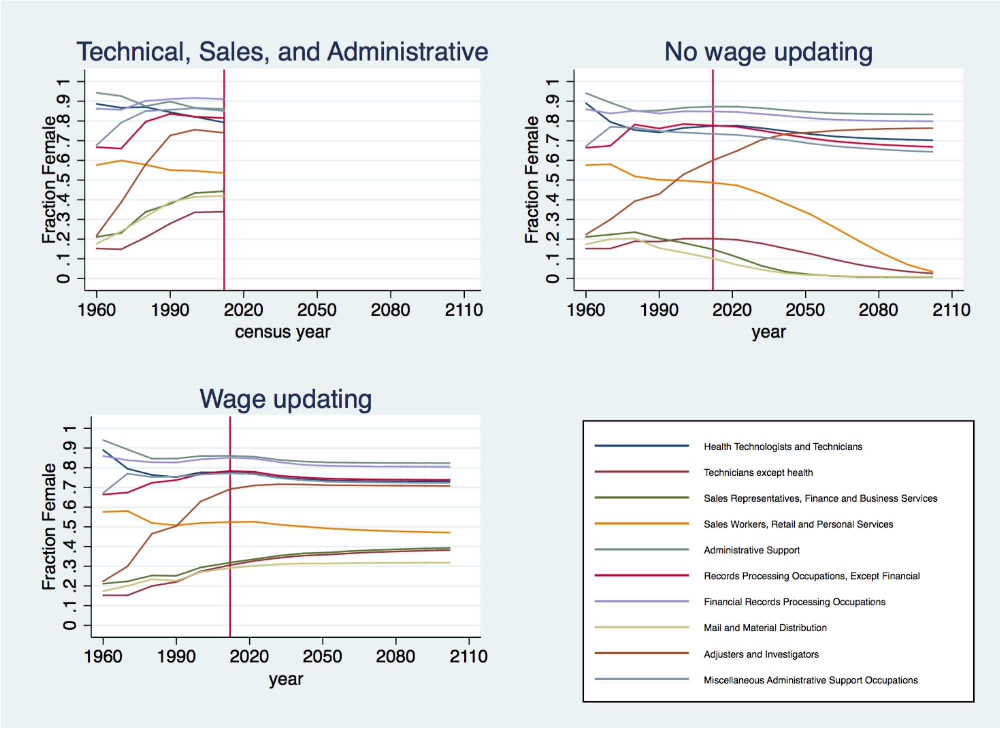


Figure 6: Status Quo: Simulated Occupation Segregation Patterns

Figure 7: Status Quo: Simulated Occupation Segregation Patterns

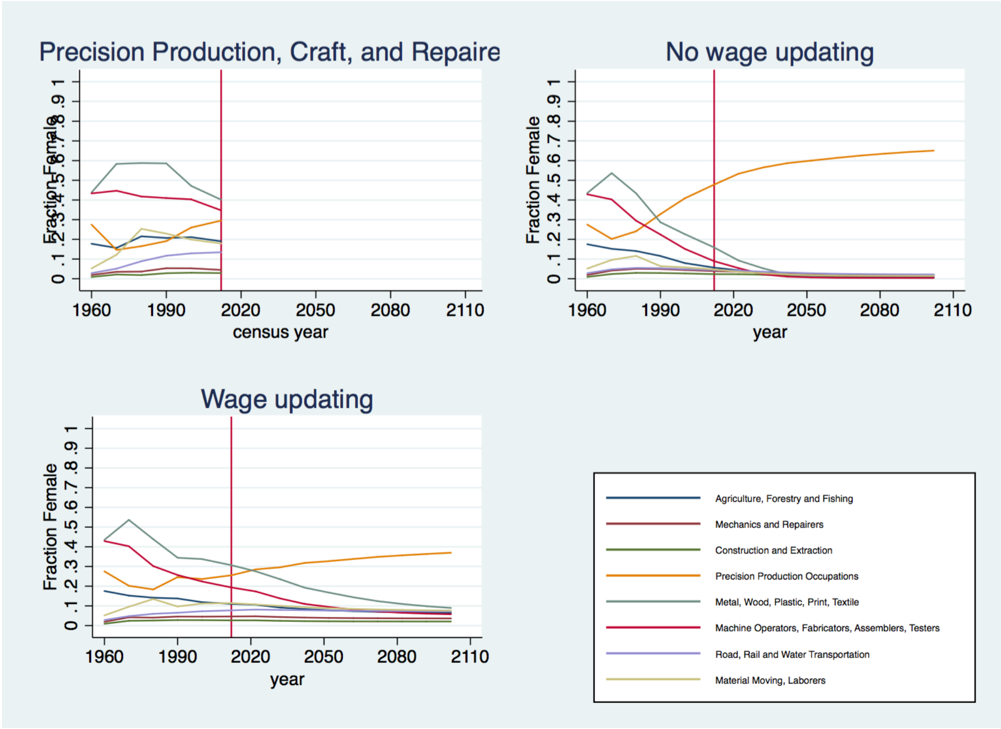


Figure 8: Initial Parity: Simulated Occupation Segregation Patterns

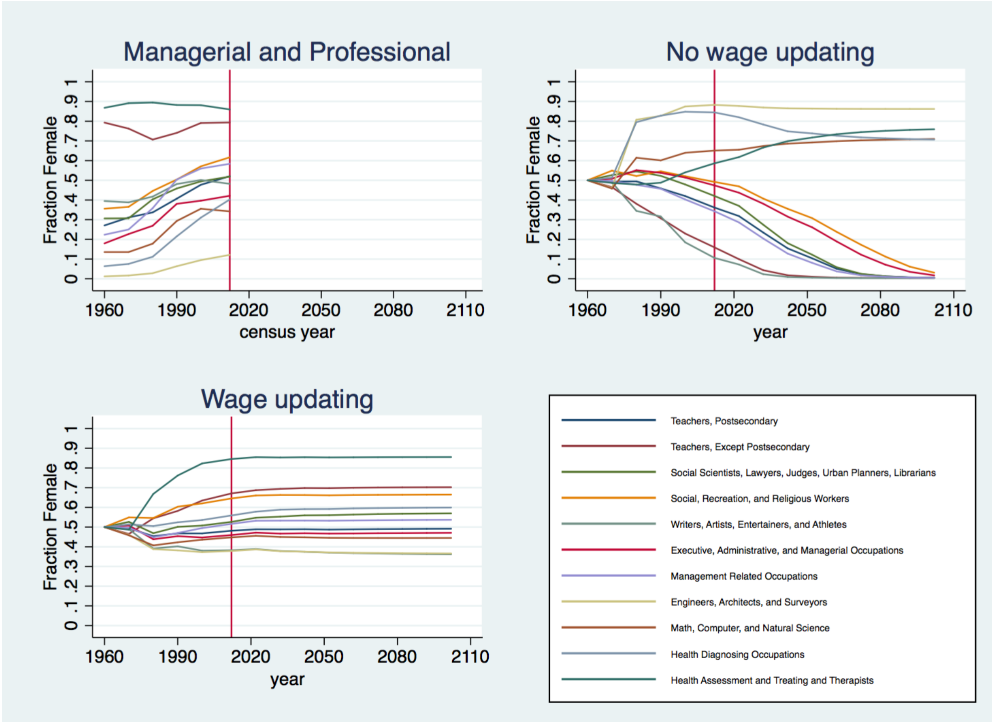


Figure 9: Initial Parity: Simulated Occupation Segregation Patterns

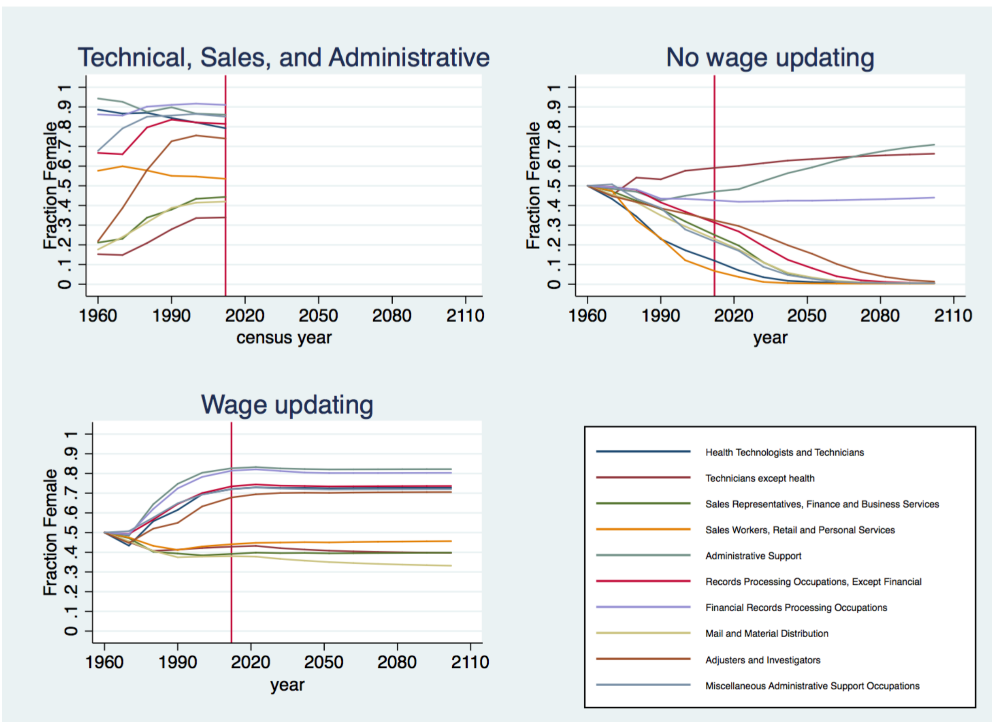


Figure 10: Initial Parity: Simulated Occupation Segregation Patterns

Figure 11: Initial Parity: Simulated Occupation Segregation Patterns

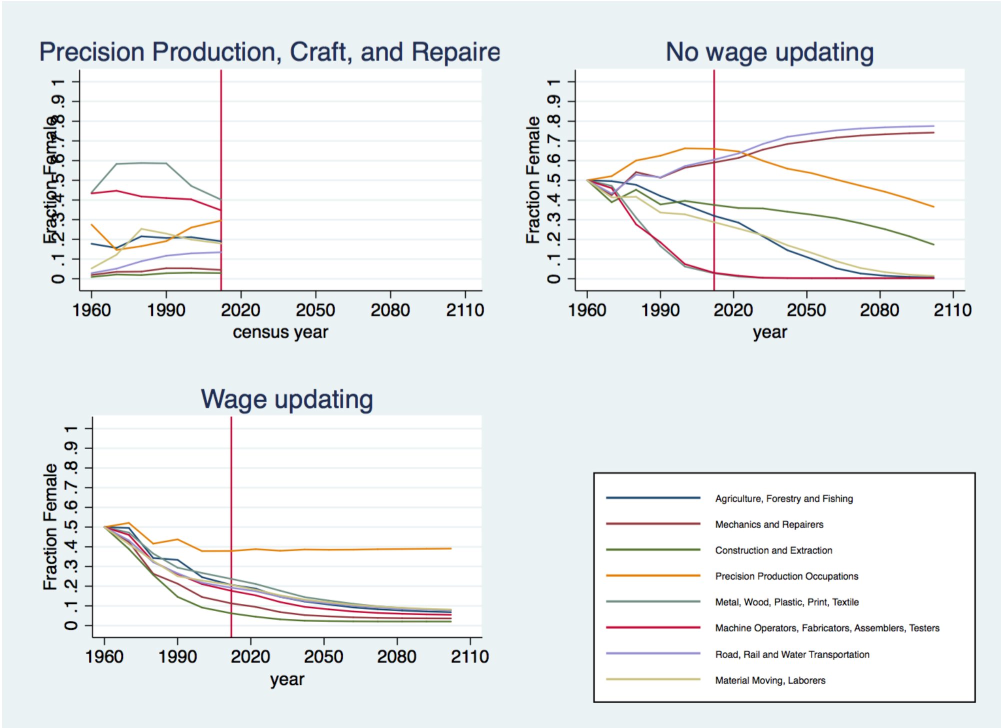


Figure 12: Transitions in Fraction Female Across Periods

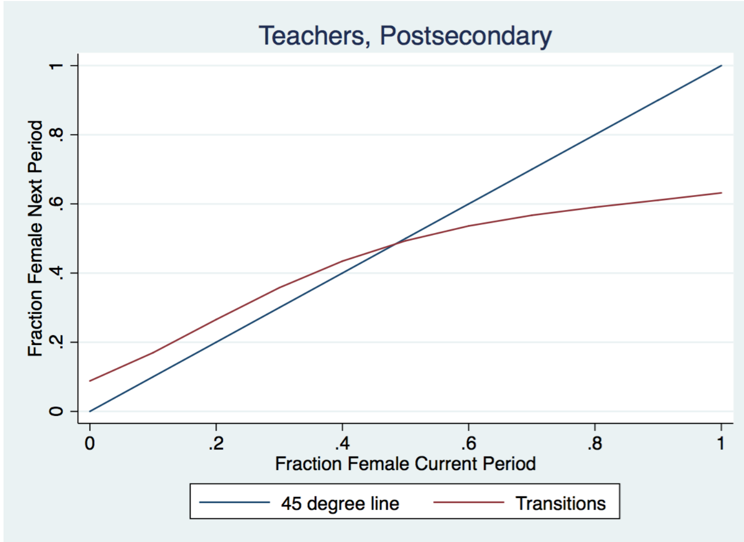


Figure 13: Transitions in Fraction Female Across Periods

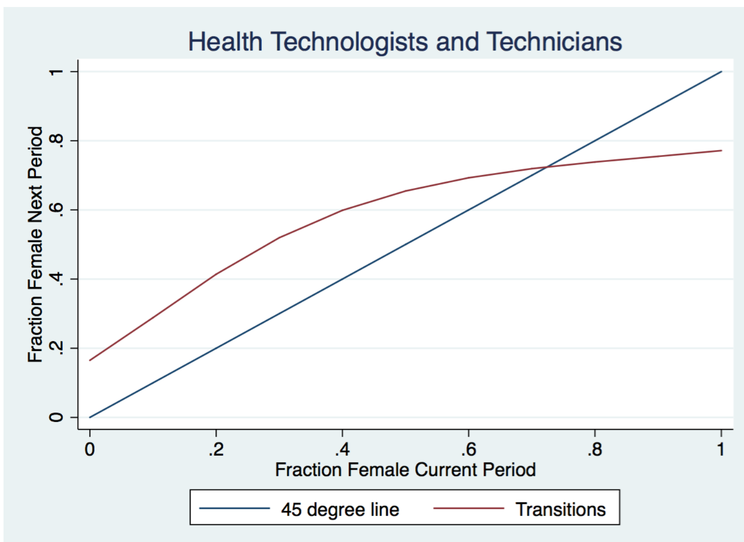


Figure 14: Transitions in Fraction Female Across Periods

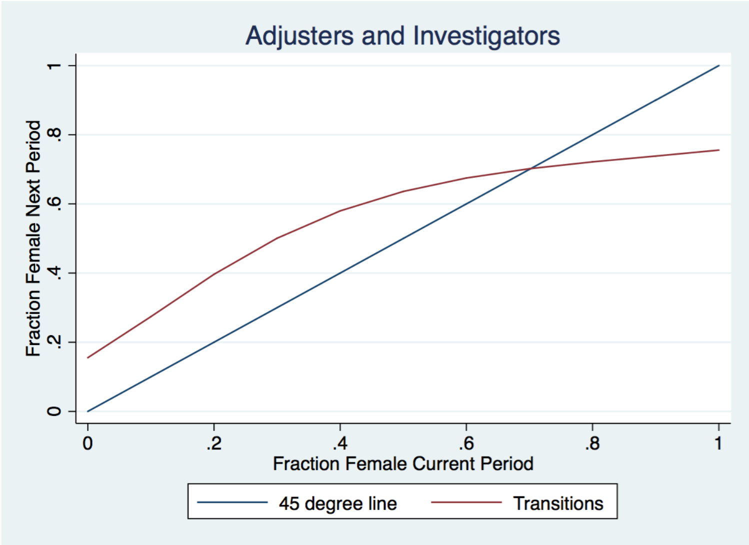


Figure 15: Transitions in Fraction Female Across Periods

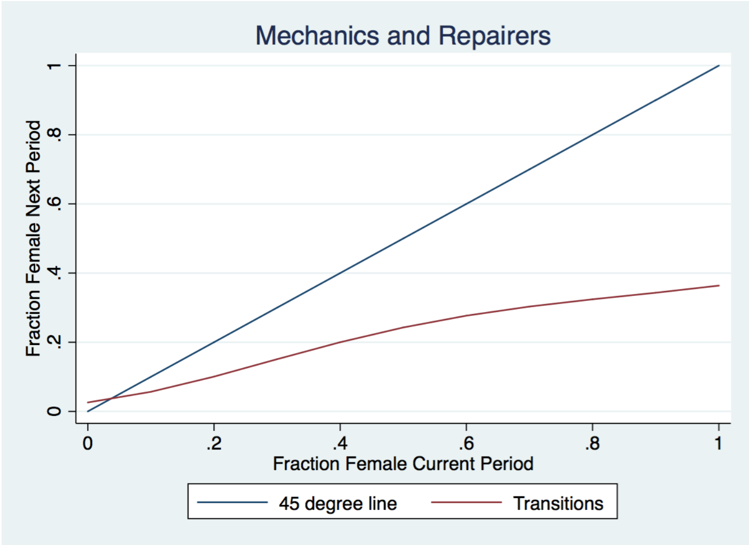


Figure 16: Transitions in Fraction Female Across Periods

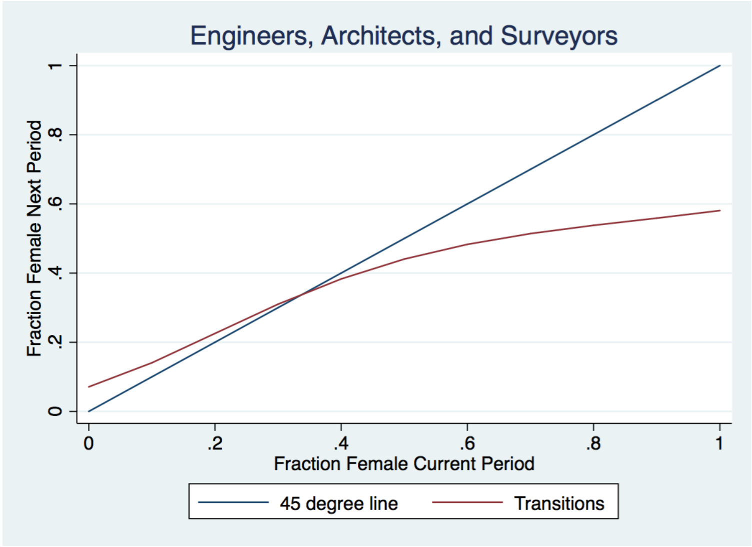


Figure 17: Transitions in Fraction Female Across Periods: Fixed Wages

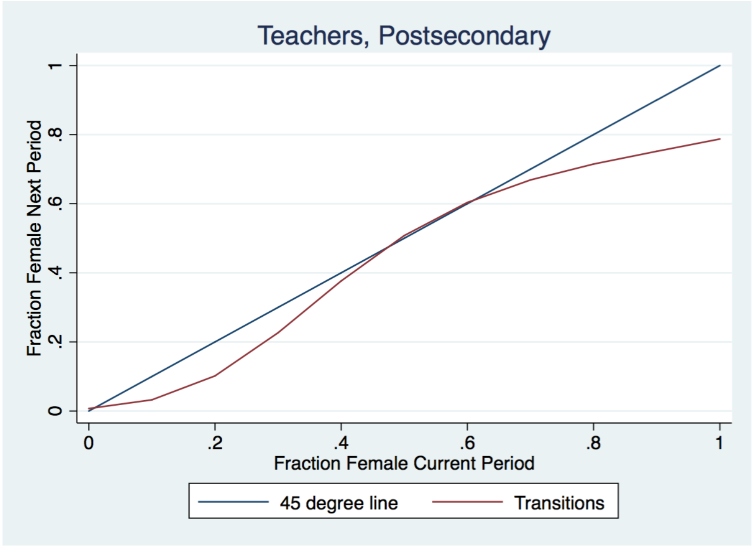


Figure 18: Transitions in Fraction Female Across Periods: Fixed Wages

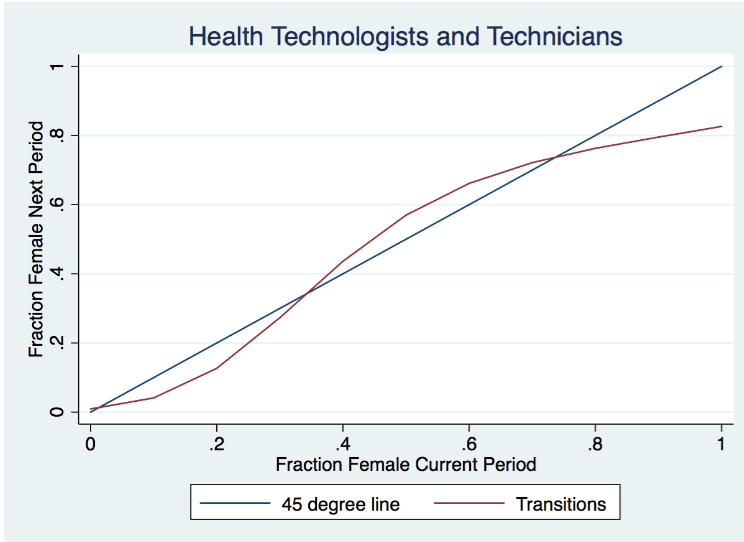


Figure 19: Transitions in Fraction Female Across Periods: Fixed Wages

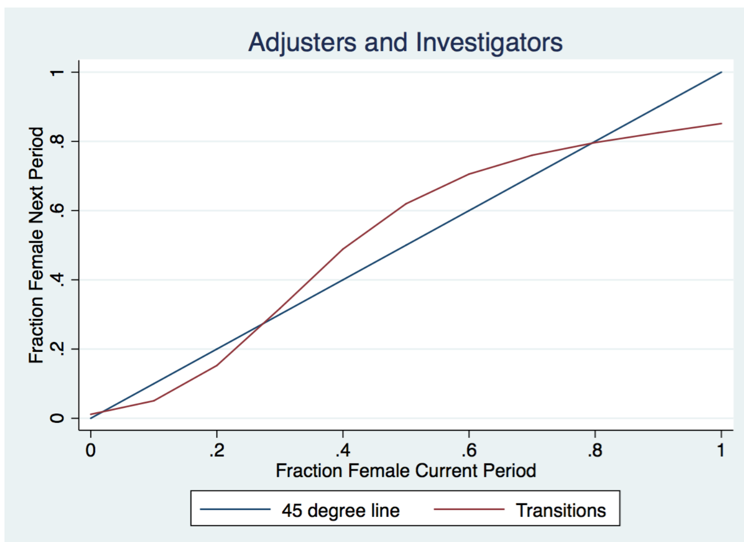


Figure 20: Transitions in Fraction Female Across Periods: Fixed Wages

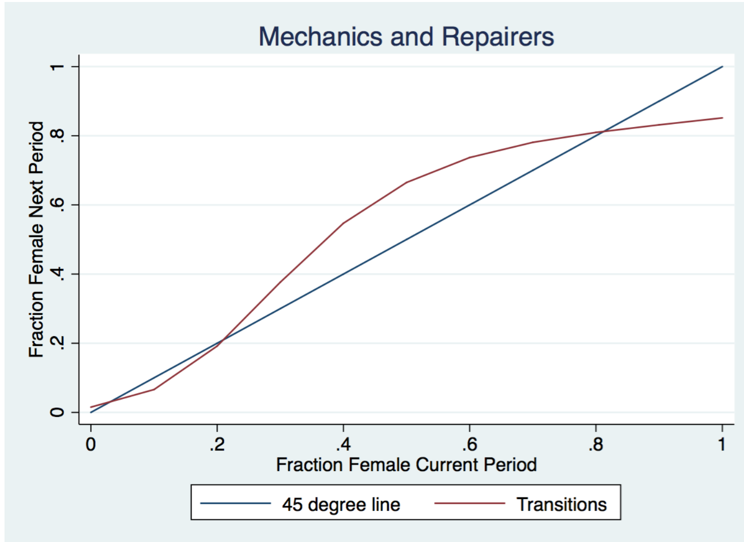


Figure 21: Transitions in Fraction Female Across Periods: Fixed Wages

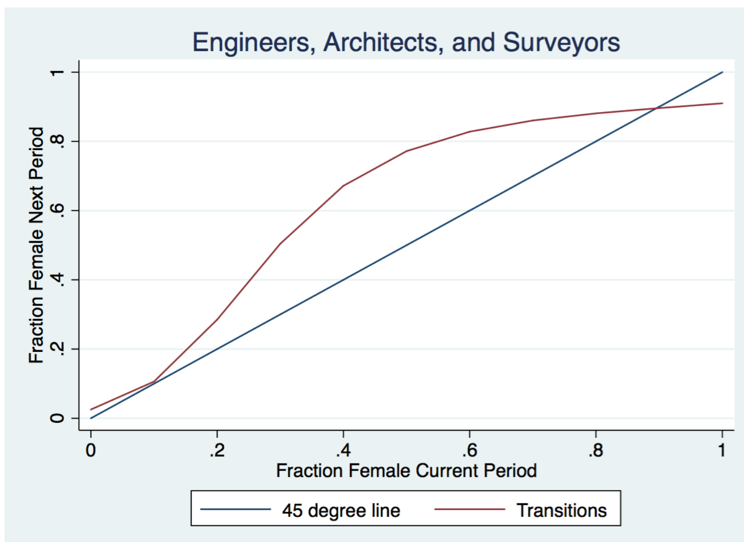


Figure 22: Transitions in Fraction Female Across Periods: Gender Preference Doubled

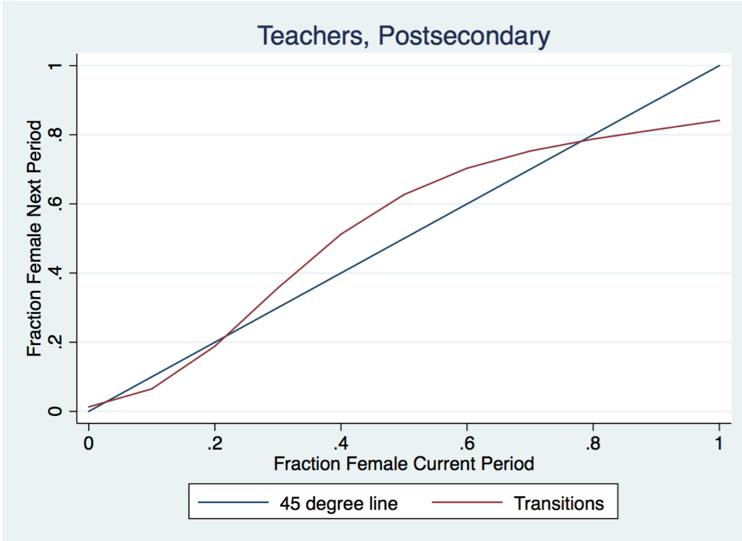


Figure 23: Transitions in Fraction Female Across Periods: Gender Preference Doubled

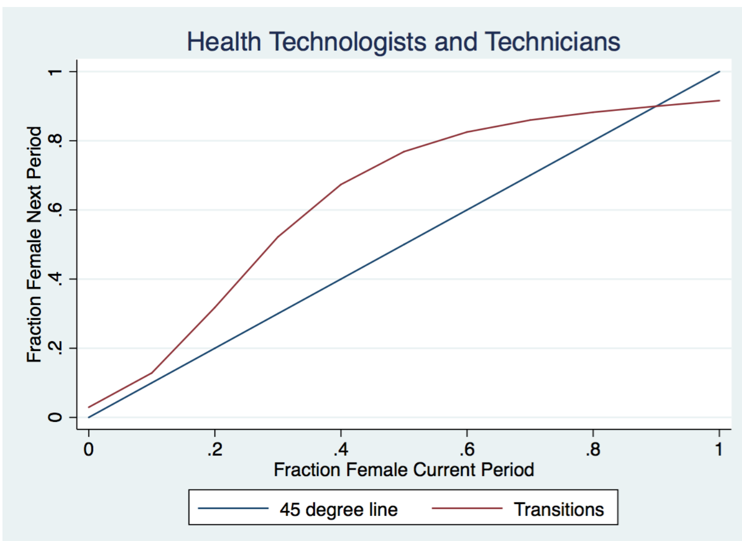


Figure 24: Transitions in Fraction Female Across Periods: Gender Preference Doubled

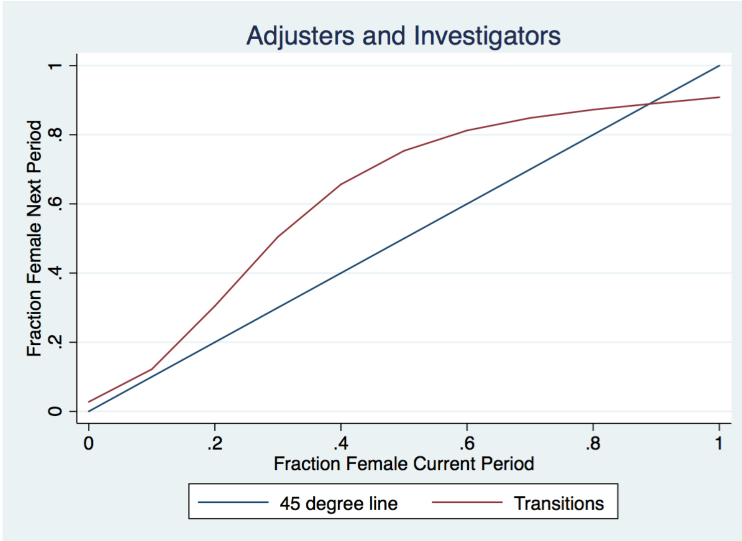


Figure 25: Transitions in Fraction Female Across Periods: Gender Preference Doubled

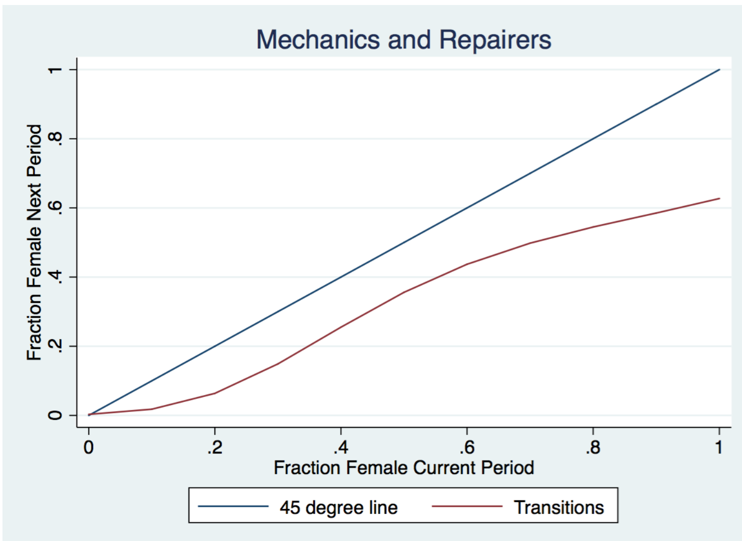


Figure 26: Transitions in Fraction Female Across Periods: Gender Preference Doubled

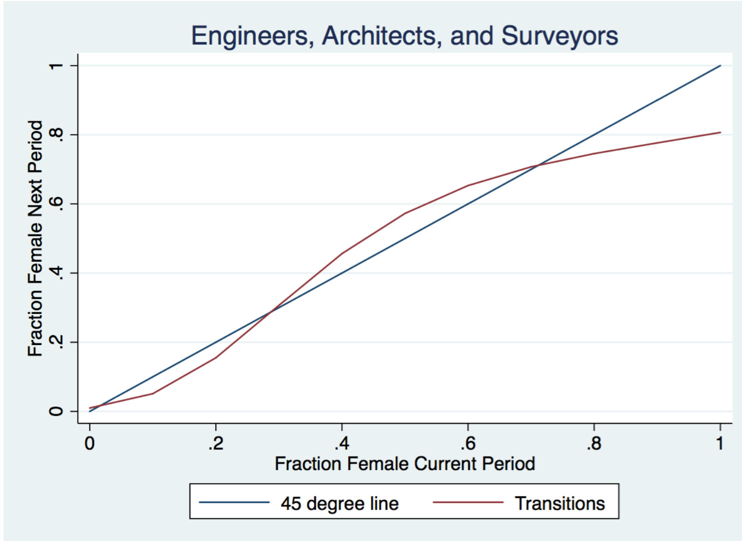


Figure 27: Status Quo: Simulated Occupation Segregation Patterns

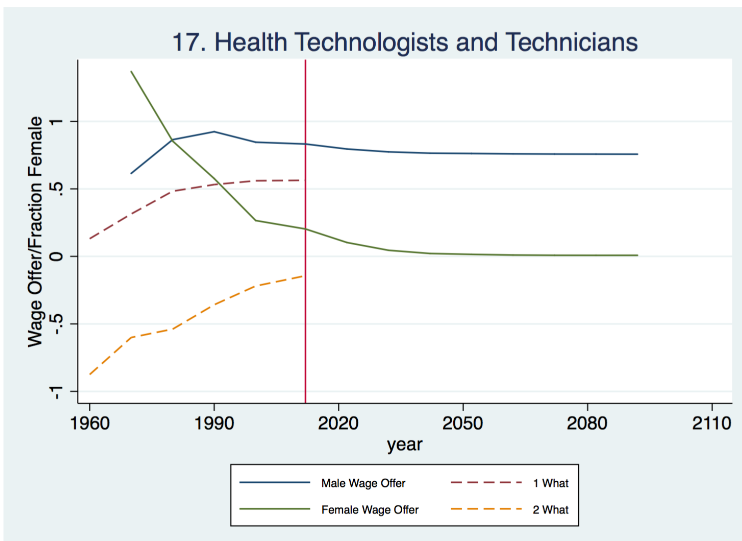
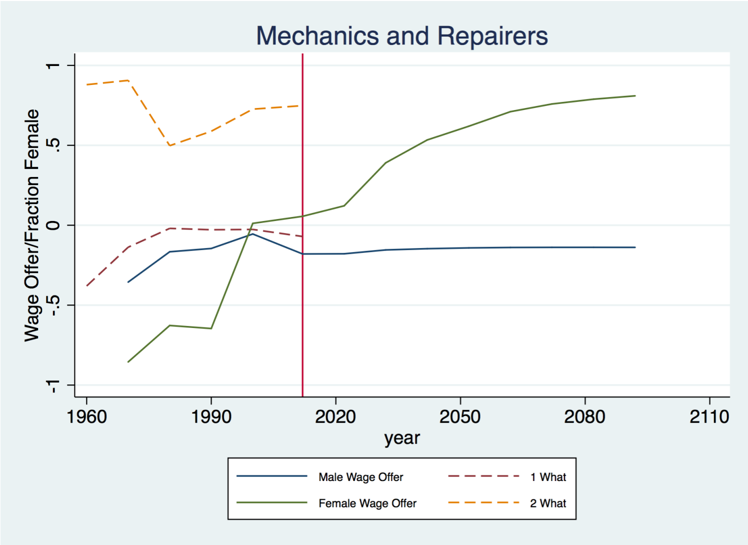


Figure 28: Status Quo: Simulated Occupation Segregation Patterns



10.2 Tables

Table 1: **Average Model Estimates by Occupation**

Occupation	Res. Wage Gap	WTP Gap	Lifetime Income Gap
Teachers, Postsecondary	0.70	0.71	0.53
Teachers, Except Postsecondary	0.50	0.81	0.58
Social Scientists, Lawyers, Judges, Urba	0.56	0.71	0.66
Social, Recreation, and Religious Worker	0.78	0.75	0.54
Writers, Artists, Entertainers, and Athl	0.71	0.57	0.57
Health Technologists and Technicians	0.50	0.70	0.51
Technicians except health	0.76	0.53	0.52
Sales Representatives, Finance and Busin	0.65	0.57	0.55
Sales Workers, Retail and Personal Servi	0.52	0.50	0.51
Administrative Support	0.42	0.83	0.53
Records Processing Occupations, Except F	0.56	0.76	0.59
Financial Records Processing Occupations	0.40	0.91	0.51
Mail and Material Distribution	0.95	0.62	0.56
Adjusters and Investigators	0.56	0.76	0.54
Miscellaneous Administrative Support Occ	0.45	0.77	0.54
Protective Service	0.89	0.48	0.54
Food Preparation and Service Occupations	0.66	0.65	0.55
Health Service Occupations	0.43	0.85	0.55
Cleaning and Building Service Occupation	0.74	0.56	0.50
Private Household and Personal Service	0.43	0.70	0.52
Agriculture, Forestry and Fishing	0.93	0.49	0.55
Mechanics and Repairers	1.63	0.46	0.58
Construction and Extraction	1.50	0.29	0.58
Precision Production Occupations	0.70	0.49	0.54
Metal, Wood, Plastic, Print, Textile	0.72	0.52	0.50
Machine Operators, Fabricators, Assemble	0.86	0.53	0.57
Road, Rail and Water Transportation	1.04	0.42	0.55
Material Moving, Laborers	1.06	0.54	0.54
Executive, Administrative, and Manageria	0.70	0.55	0.55
Management Related Occupations	0.66	0.64	0.55
Engineers, Architects, and Surveyors	1.18	0.55	0.62
Math, Computer, and Natural Science	0.87	0.64	0.59
Health Diagnosing Occupations	0.96	0.70	0.56
Health Assessment and Treating and Thera	0.38	0.93	0.61
Total	0.75	0.63	0.55

Gaps are the ratio of female to male values. Reservation wages and willingness to pay are estimated from the transferable utility matching model using Census and ACS data matched to simulated lifetime income data from the SIPP. Values are the average across data years (1960, 1970, 1980, 1990, 2000, 2012) within occupation.

Table 2: Decomposition of Utility for Male Workers: Panel Evidence

	FE	FEIV1	FEIV2	FEIV3	FEIV4
Fraction Female ($\frac{\gamma^M}{\sigma^M\_\eta}$)	0.666	0.549	1.605	0.00397	-0.0197
	(0.644)	(3.014)	(2.298)	(1.179)	(1.049)
Latent Wage Offer	-0.916	3.248	1.071	1.815	1.703
	(0.564)	(4.501)	(2.091)	(1.875)	(1.705)
Constant	-2.615***				
	(0.294)				
Year dummies	Yes	Yes	Yes	Yes	Yes
Observations	204	204	204	204	204
KP rk F=		0.533	1.930	1.857	8.622

Standard errors in parentheses are clustered at the occupation level.

$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.0001$

Table 3: Decomposition of Utility for Female Workers: Panel Evidence

	FE	FEIV1	FEIV2	FEIV3	FEIV4
Fraction Female ($\frac{\gamma^F}{\sigma^F\_\eta}$)	2.870***	11.11	8.206**	4.363*	4.569**
	(0.577)	(47.53)	(3.001)	(1.808)	(1.540)
Squared distance from parity	-4.383***	16.67	-6.907	-6.980	-7.328
	(0.915)	(283.6)	(6.724)	(4.274)	(3.772)
Cubed distance from parity	-4.506	-86.89	-36.70	7.749	6.664
	(2.605)	(678.1)	(23.72)	(9.780)	(9.718)
Latent Wage Offer	-0.151	7.630	1.037	1.843*	1.932
	(0.299)	(72.56)	(2.165)	(0.886)	(1.039)
Constant	-5.165***				
	(0.224)				
Year dummies	Yes	Yes	Yes	Yes	Yes
Observations	204	204	204	204	204
KP rk F=		0.00234	0.759	5.255	10.03

Standard errors in parentheses are clustered at the occupation level.

$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.0001$

Table 4: Percent increase in women entering an occupation as fraction female of existing workers moves from 20% to 80% (average marginal effect)

Variable	Mean	Std. Dev.	Min.	Max.
Percent increase in women (fixed wage)	1352.072	73.625	1080.847	1442.335
Percent increase in women (equilibrium wage)	124.42	42.05	77.067	277.114
N	34			

Table 5: Fixed effects by occupation

Occupation	Male	Female
Sales Representatives, Finance and Business Services	1.66	1.25
Machine Operators, Fabricators, Assemblers, Testers	2.67	1.1
Material Moving, Laborers	2.79	1.96
Teachers, Postsecondary	0	0
Teachers, Except Postsecondary	1.02	1.48
Social Scientists, Lawyers, Judges, Urban Planners, Librarians	.17	-.15
Social, Recreation, and Religious Workers	-.07	.54
Writers, Artists, Entertainers, and Athletes	.96	.31
Health Technologists and Technicians	-.71	.2
Technicians except health	1.59	1.6
Sales Workers, Retail and Personal Services	2.06	1.85
Administrative Support	.36	2.04
Records Processing Occupations, Except Financial	-1.07	-.32
Financial Records Processing Occupations	-.84	.75
Mail and Material Distribution	1.82	1.39
Adjusters and Investigators	.38	1.17
Miscellaneous Administrative Support Occupations	.34	1.14
Protective Service	1.66	1.73
Food Preparation and Service Occupations	1.66	1.35
Health Service Occupations	-.54	1.28
Cleaning and Building Service Occupations, Except Household	1.22	1.34
Private Household and Personal Service	.33	1.66
Agriculture, Forestry and Fishing	2.22	1.2
Mechanics and Repairers	2.73	1.35
Construction and Extraction	2.93	1.1
Precision Production Occupations	2.06	1.84
Metal, Wood, Plastic, Print, Textile	1.52	.42
Road, Rail and Water Transportation	2.64	1.89
Executive, Administrative, and Managerial Occupations	2.5	2.35
Management Related Occupations	1.33	1.29
Engineers, Architects, and Surveyors	1.38	1.17
Math, Computer, and Natural Science	1.02	.89
Health Diagnosing Occupations	-.22	.33
Health Assessment and Treating and Therapists	-.7	.82

Table 6: Ratio of Female to Male Lifetime Income in Millions

	Parameter		
Cell	GenderPref	NoGenderPref	Difference
Teachers, Postsecondary	0.61	0.69	-0.08
Teachers, Except Postsecondary	0.72	0.91	-0.19
Social Scientists, Lawyers, Judg	0.82	0.96	-0.13
Social, Recreation, and Religiou	0.62	0.77	-0.15
Writers, Artists, Entertainers,	0.74	0.81	-0.06
Health Technologists and Technic	0.61	0.78	-0.17
Technicians except health	0.59	0.60	-0.01
Sales Representatives, Finance a	0.73	0.75	-0.02
Sales Workers, Retail and Person	0.64	0.73	-0.09
Administrative Support	0.61	0.83	-0.22
Records Processing Occupations,	0.71	0.92	-0.21
Financial Records Processing Occ	0.58	0.78	-0.20
Mail and Material Distribution	0.70	0.71	-0.01
Adjusters and Investigators	0.63	0.78	-0.15
Miscellaneous Administrative Sup	0.65	0.84	-0.19
Protective Service	0.65	0.61	0.04
Food Preparation and Service Occ	0.67	0.77	-0.10
Health Service Occupations	0.73	0.99	-0.26
Cleaning and Building Service Oc	0.59	0.60	-0.01
Private Household and Personal S	0.65	0.86	-0.21
Agriculture, Forestry and Fishin	0.71	0.61	0.10
Mechanics and Repairers	0.70	0.56	0.15
Construction and Extraction	0.64	0.50	0.14
Precision Production Occupations	0.74	0.73	0.01
Metal, Wood, Plastic, Print, Tex	0.65	0.67	-0.01
Machine Operators, Fabricators,	0.80	0.75	0.06
Road, Rail and Water Transportat	0.70	0.57	0.13
Material Moving, Laborers	0.68	0.59	0.10
Executive, Administrative, and M	0.68	0.74	-0.05
Management Related Occupations	0.67	0.76	-0.09
Engineers, Architects, and Surve	0.75	0.68	0.07
Math, Computer, and Natural Scie	0.65	0.67	-0.02
Health Diagnosing Occupations	0.61	0.66	-0.05
Health Assessment and Treating a	0.78	1.07	-0.29

Note: GenderPref is ratio of female to male predicted lifetime incomes in the main specification, and NoGenderPref is the same but with the coefficients on fraction female set to zero. The Difference column is GenderPref minus NoGenderPref. Positive differences mean that women earn relatively more due to the preference over the fraction female, and negative differences means that women earn relatively less due to the preference over fraction female.

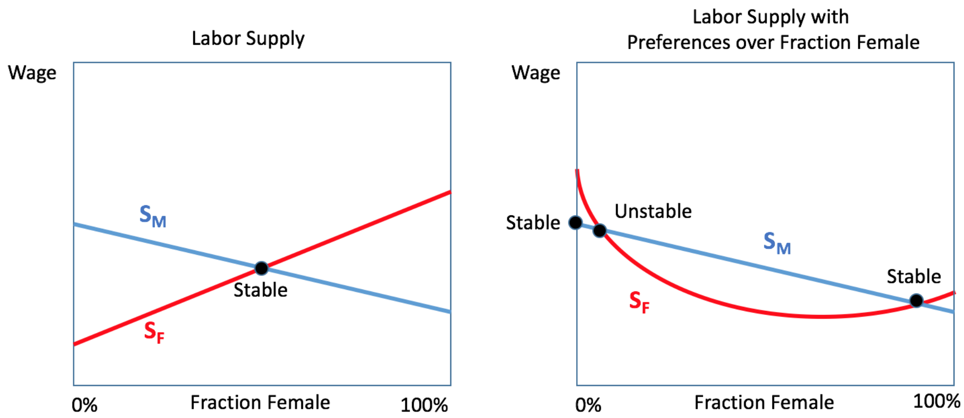
10.3 Pan 2010 Tipping Model

To illustrate the concept of "tipping" I refer to the model of Pan (2015), illustrated below in Figure 29. In this stylized model, an occupation can hire either male or female workers to fill a fixed number of positions. As more women are hired, fewer men are hired, and the fraction female goes up. Assuming upward sloping supply curves for men and women, wages must go up as more workers of either gender are hired. Therefore, assuming equal productivity, firms continue to hire men or women until wages are equalized, which depends on the shape of the supply curves. Tipping points emerge if one or more of the supply curves are not always upward sloping. In the righthand side of Figure 29, we see the case where the female labor supply curve becomes downward sloping at very low fractions female, illustrating a scenario where women are so strongly averse to entering male dominated occupations that hiring more women actually allows the occupation to pay lower wages to women.

In this stylized model, the location of stable and unstable sorting equilibria depends only on the male and female labor supply curves in the occupation, which could depend on how non-wage amenities, wages, and fraction female are valued by men and women.[28] Another key factor affecting the gender mix in the occupation is that firms may not have the same willingness to pay for male and female workers. Firms may values workers of one gender less due to productivity or skill differences, or taste-based or statistical discrimination, or devaluation. Such a gap willingness-to-pay on the firm side will drive a wedge between male and female wages and push the mixed-gender equilibrium in an occupation up or down. Thus both supply and demand factors impact the fraction female in an occupation, and also the tipping point or stability of that fraction female, if there are preferences over occupation gender in the labor supply curves.

[28] In Pan (2015) the locations of the unstable equilibria, and therefore the tipping points, are fixed to be the same across occupations within blue collar or white collar categories.

Figure 29: Stylized Model of Tipping from Pan 2010



10.4 List of occupation Codes

Note that not all component occupations exist in all years of data.

Adjusters and Investigators

Administrative Support

Agriculture, Forestry and Fishing

Cleaning and Building Service

Construction and Extraction

Engineers, Architects, and Surveyors

Executive, Administrative, and Managerial

Financial Records Processing

Food Preparation and Service

Health Assessment and Treating and Therapists

Health Diagnosing

Health Service

Health Technologists and Technicians

Machine operators, Fabricators, Assemblers, Testers

Mail and Material Distribution

Management Related

Material Moving, Laborers

Math, Computer, and Natural Science

Mechanics and Repairers

Metal, Wood, Plastic, Print, Textile

Miscellaneous Administrative Support

Precision Production

Private Household and Personal Services

Protective Service

Records Processing

Road, Rail and Water Transportation

Sales Representatives, Finance and Business Services

Sales Workers, Retail and Personal Services

Social Scientists, Lawyers, Judges

Social, Recreation, and Religious Workers

Teachers, Except Postsecondary

Teachers, Postsecondary

Technicians except health

Writers, Artists, Entertainers, and Athletes

10.5 Joint Likelihood Tobit Type 5

$$
\begin{aligned}
y_{1j}^* &= \overline{WTP}_o^F - \bar{W}_o^F - \xi_j^F - (\overline{WTP}_o^M - \bar{W}_o^M - \xi_j^M) \\
y_{2j}^* &= \bar{W}_o^F + \bar{\xi}_j^F \\
y_{3j}^* &= \bar{W}_o^M + \bar{\xi}_j^M \\
y_{2j} &= y_{2j}^* && \text{if } y_{1j}^* > 0 \\
y_{2j} &= 0 && \text{if } y_{1j}^* \leqslant 0 \\
y_{3j} &= y_{3j}^* && \text{if } y_{1j}^* \leqslant 0 \\
y_{3j} &= 0 && \text{if } y_{1j}^* > 0
\end{aligned}
$$

Let $f_{1,3}$ be the joint density of y_{1j}^* and y_{3j}^*, and likewise $f_{1,2}$.

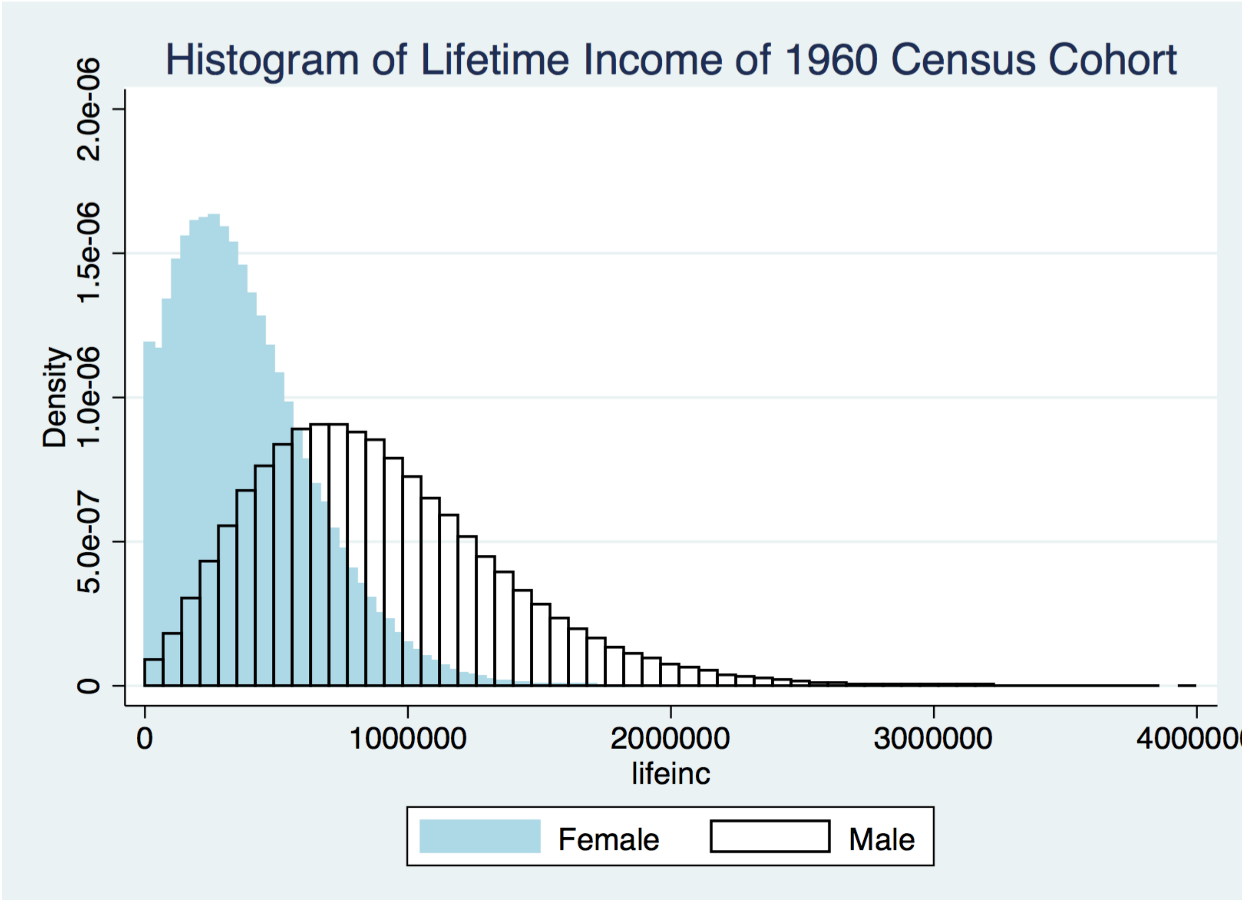
$$
\begin{aligned}
L &= \prod_F \int_{-\infty}^{0} f_{1,3}(y_{1j}^*, y_{3j}) dy_{1j}^* \prod_M \int_0^{\infty} f_{1,2}(y_{1j}^*, y_{2j}) dy_{1j}^* \\
&= \prod_j Pr(y_{1j}^* \leqslant 0, y_{3j})^{\mathcal{I}(y_{3j})} * Pr(y_{1j}^* > 0, y_{2j})^{\mathcal{I}(y_{2j})} \\
&= \prod_j (Pr(y_{1j}^* \leqslant 0 | y_{3j}) * Pr(y_{3j}))^{\mathcal{I}(y_{3j})} * (Pr(y_{1j}^* > 0 | y_{2j}) * Pr(y_{2j}))^{\mathcal{I}(y_{2j})} \\
&= \prod_j (F_1(0|y_{3j}) * f_3(y_{3j}))^{\mathcal{I}(y_{3j})} * (F_{-1}(0|y_{2j}) * f_2(y_{2j}))^{\mathcal{I}(y_{2j})}
\end{aligned}
$$

Where F_1 is the cdf of y_{1j}^*, F_{-1} the cdf of $-y_{1j}^*$, f_3 is the pdf of y_{3j}^*, and f_2 the pdf of y_{2j}^*.

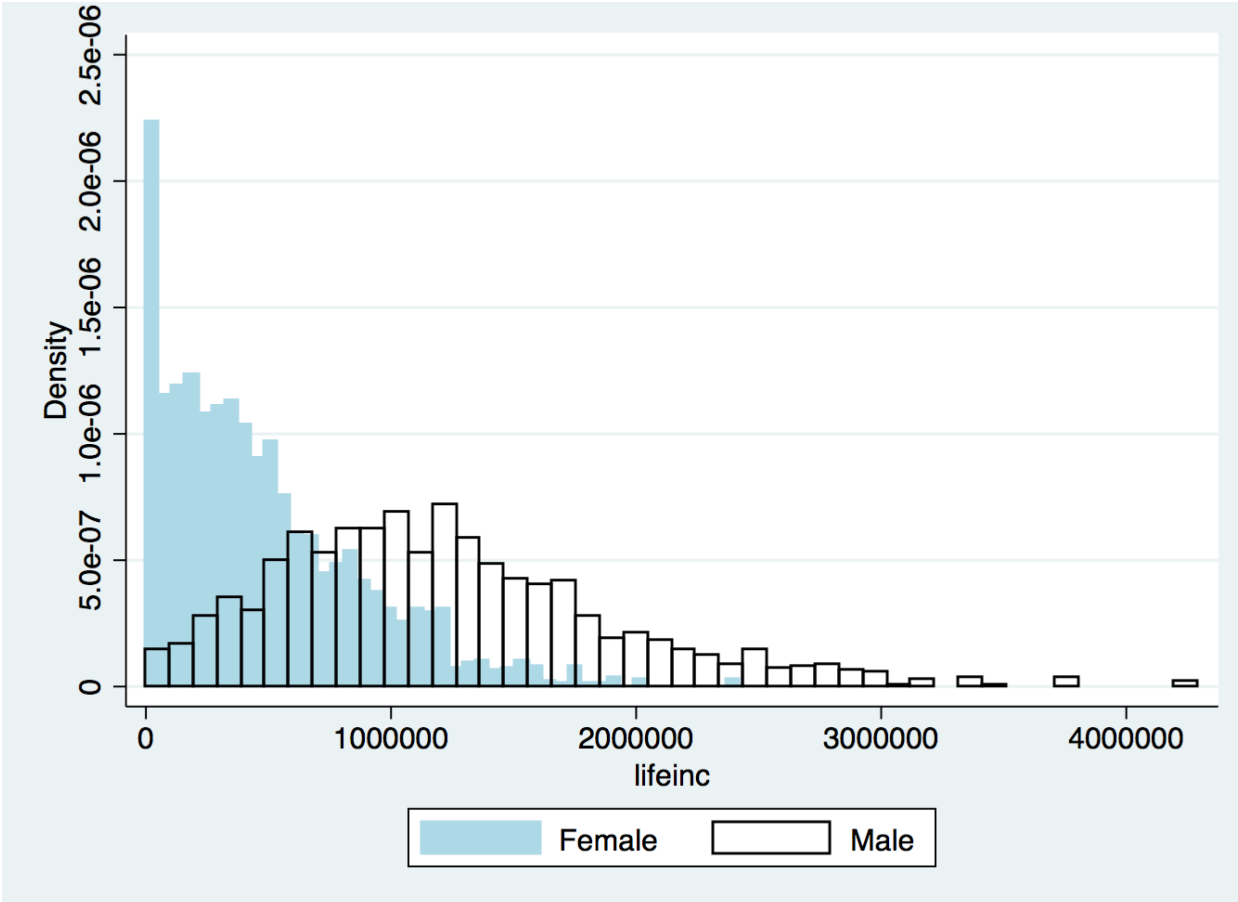
$$
\begin{aligned}
y_{2j} = y_{3j} = 0 \qquad & \text{if} \quad \overline{WTP}_o^M - \bar{W}_o^M - \bar{\xi}_j^M < 0 \\
& \text{and} \quad \overline{WTP}_o^F - \bar{W}_o^F - \bar{\xi}_j^F < 0
\end{aligned}
$$

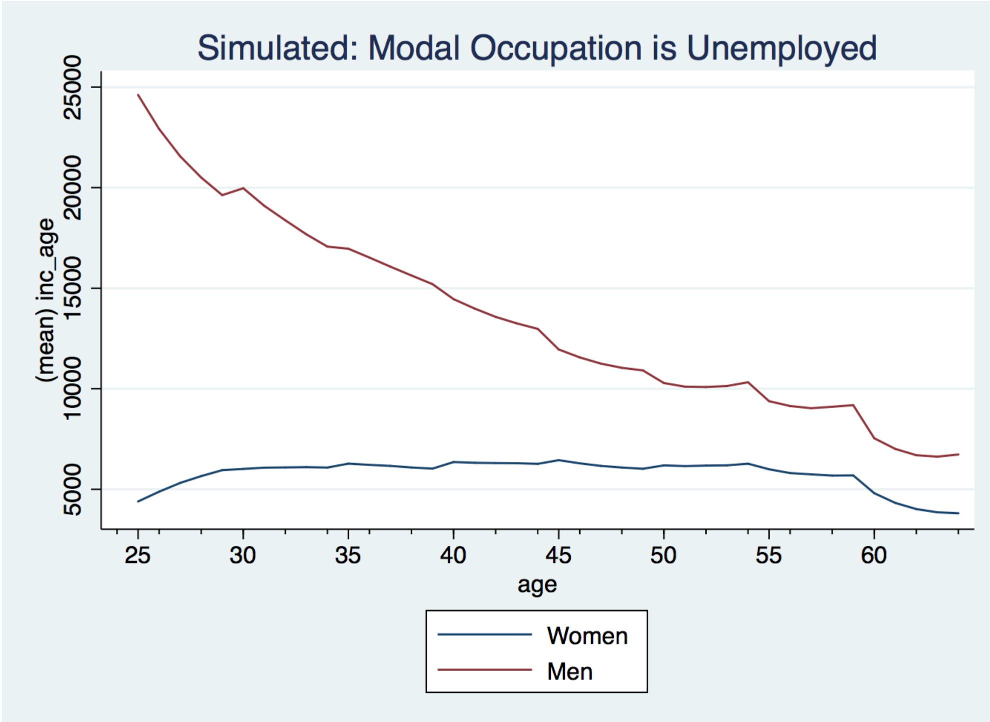
10.6 Lifetime Income Estimation

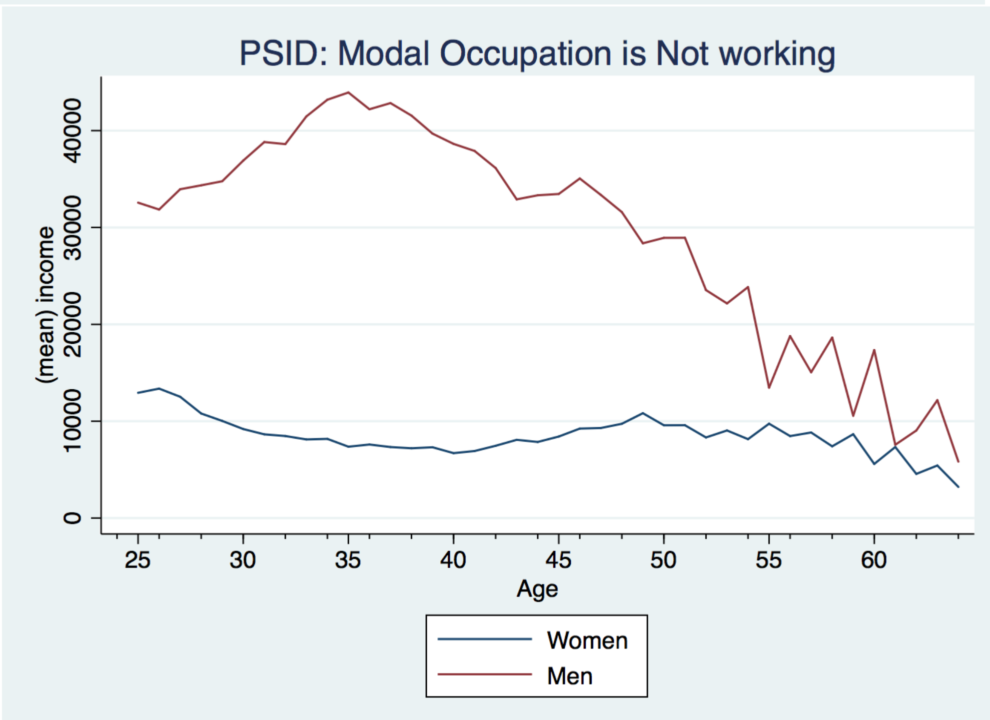
10.6.1 Comparison of Lifetime Income Estimates to PSID

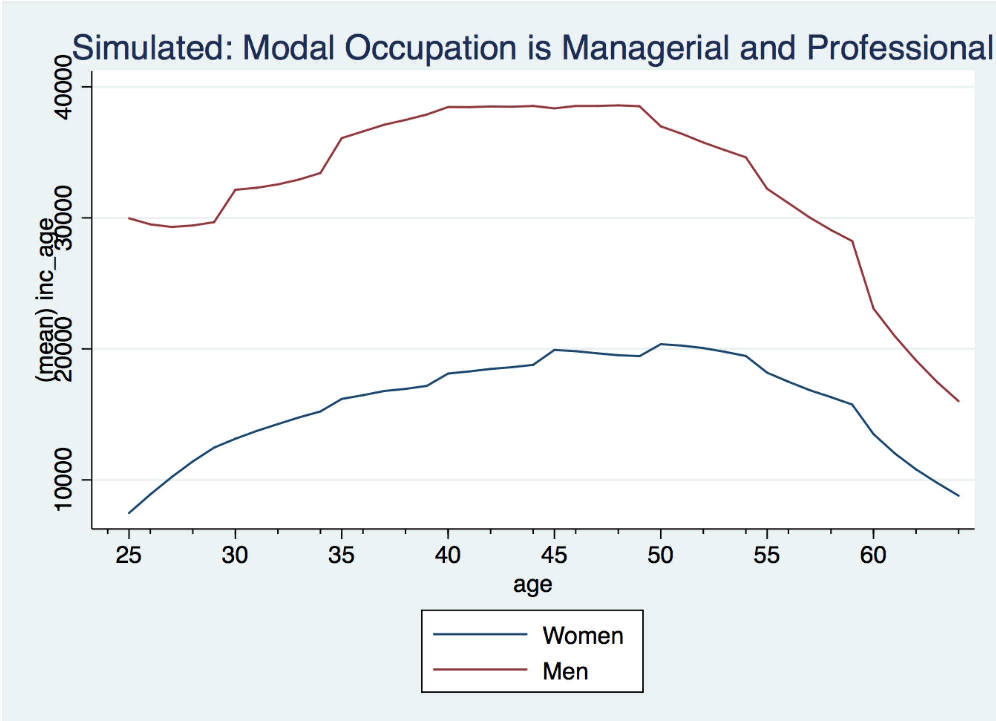


10.6.2 Histogram of lifetime income: PSID ages 25-55

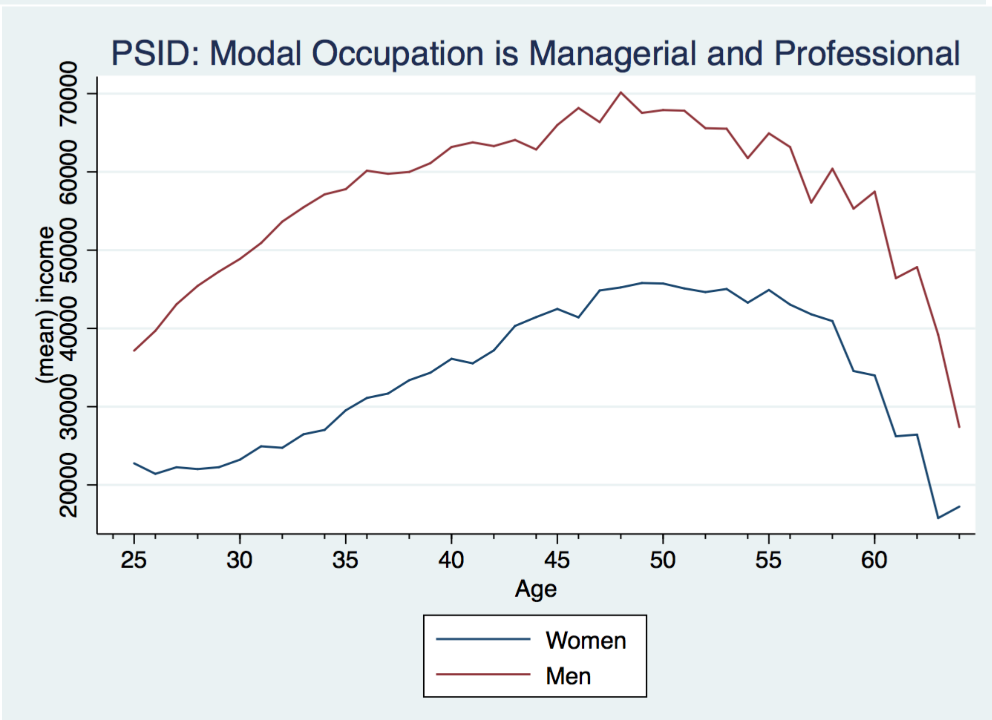



Simulated: Modal Occupation is Unemployed
(mean) inc_age
5000
10000
15000
20000
25000
25
30
35
40
45
50
55
60
age
Women
Men

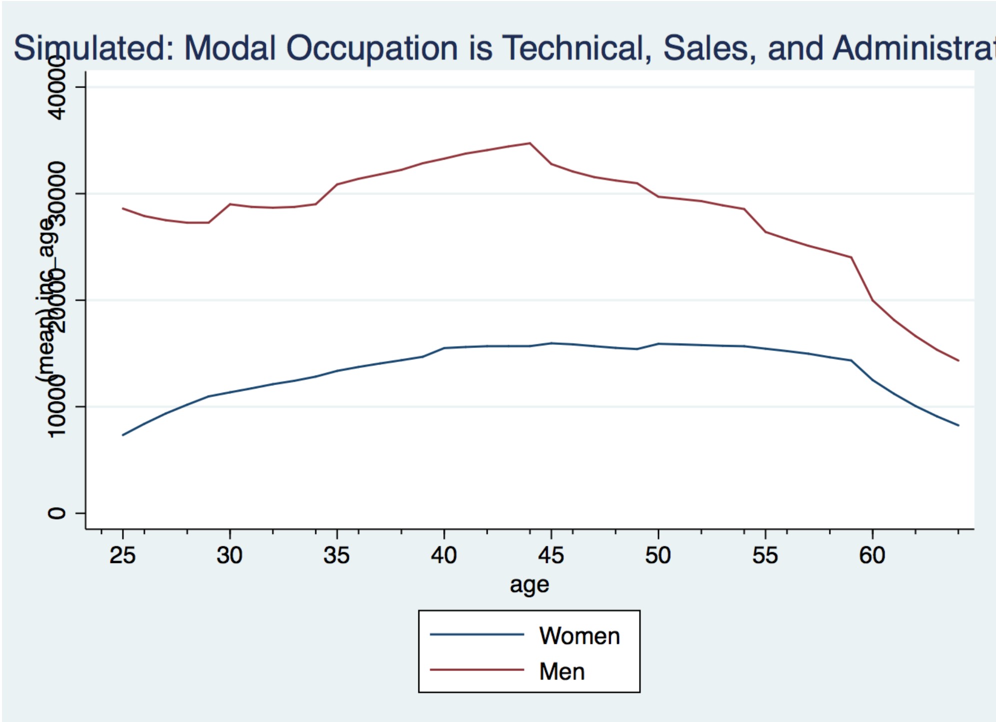

PSID: Modal Occupation is Not working
(mean) income
0
10000
20000
30000
40000
25
30
35
40
45
50
55
60
Age
Women
Men



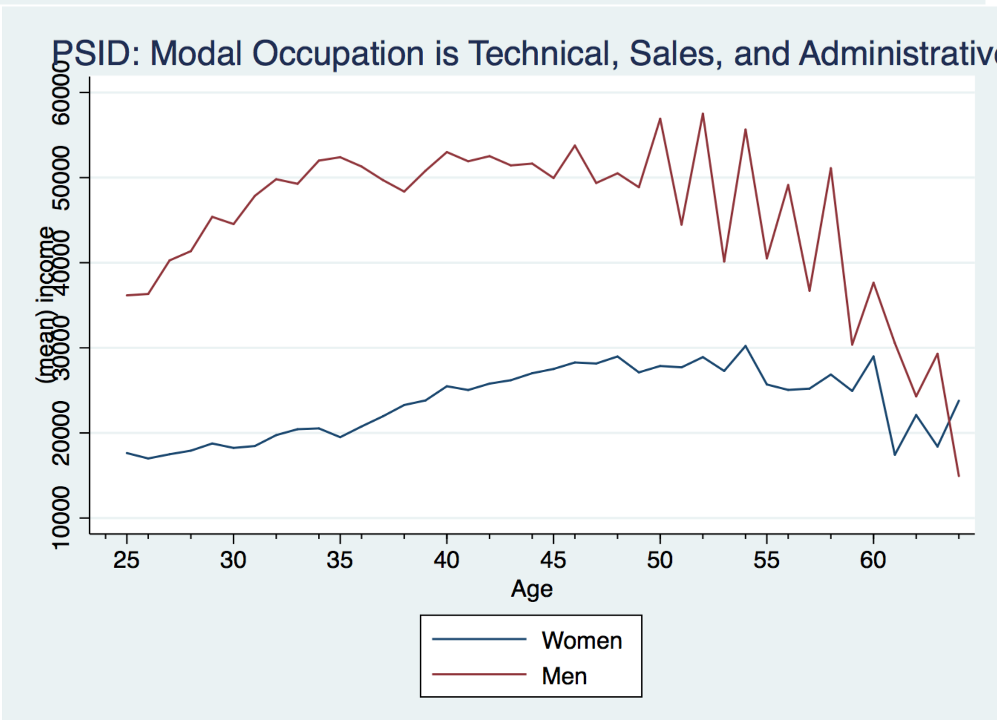
Simulated: Modal Occupation is Managerial and Professional
(mean) inc_age
10000
20000
30000
40000
25
30
35
40
45
50
55
60
age
Women
Men



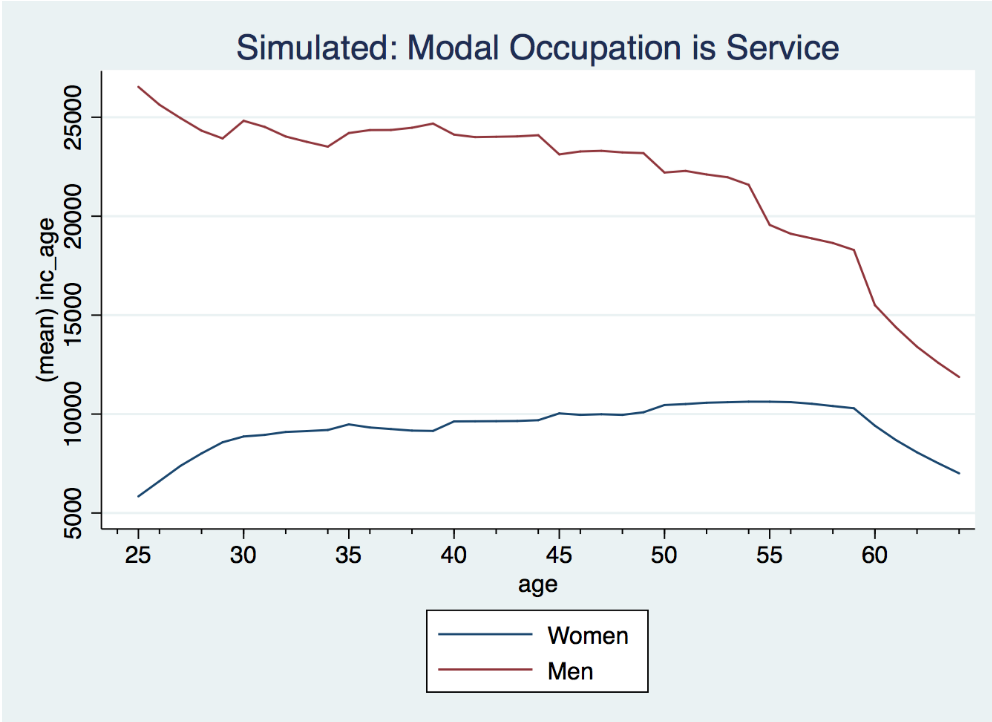
PSID: Modal Occupation is Managerial and Professional
(mean) income
20000
30000
40000
50000
60000
70000
25
30
35
40
45
50
55
60
Age
Women
Men



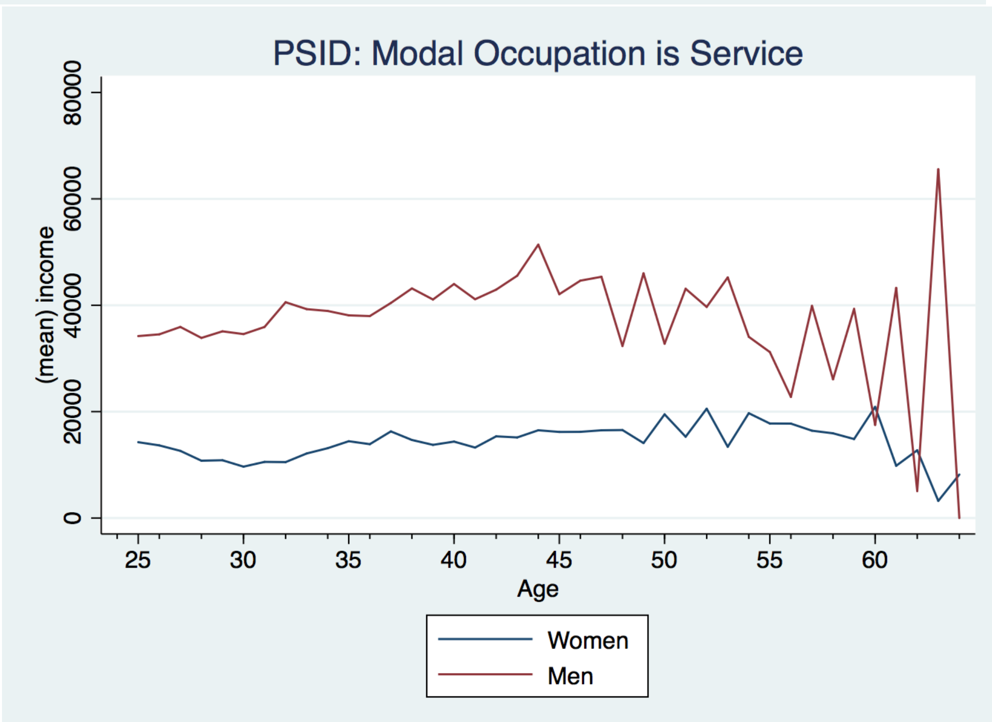
Simulated: Modal Occupation is Technical, Sales, and Administrat
(mean) inc_age
0
10000
20000
30000
40000
25
30
35
40
45
50
55
60
age
Women
Men



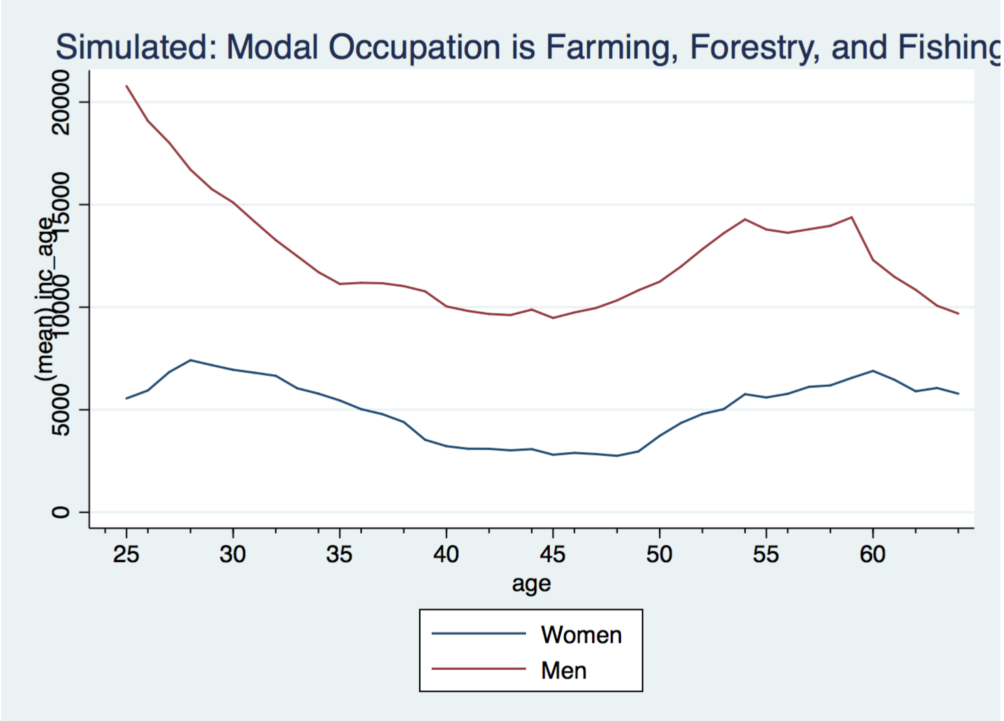
PSID: Modal Occupation is Technical, Sales, and Administrative
(mean) income
10000
20000
30000
40000
50000
60000
25
30
35
40
45
50
55
60
Age
Women
Men

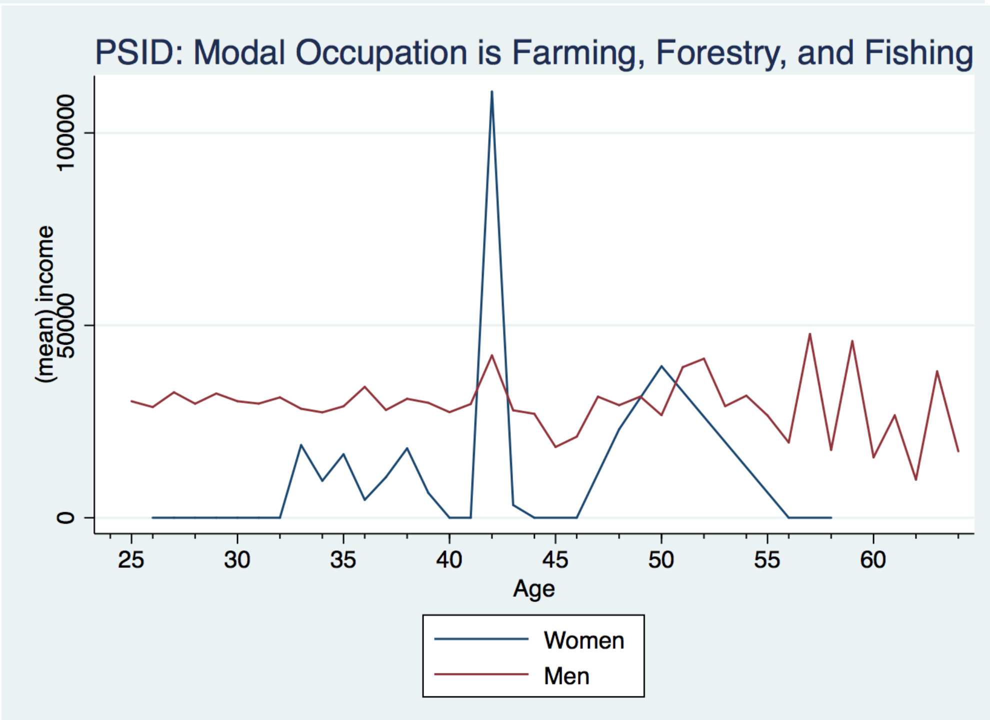


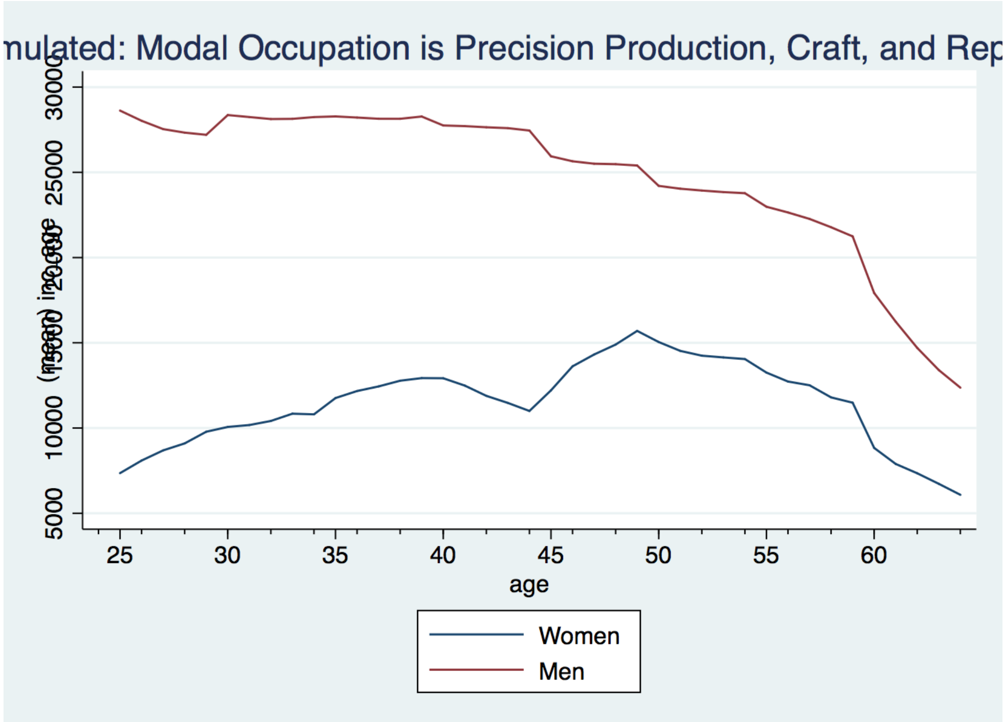
Simulated: Modal Occupation is Service
(mean) inc_age
5000
10000
15000
20000
25000
25
30
35
40
45
50
55
60
age
Women
Men

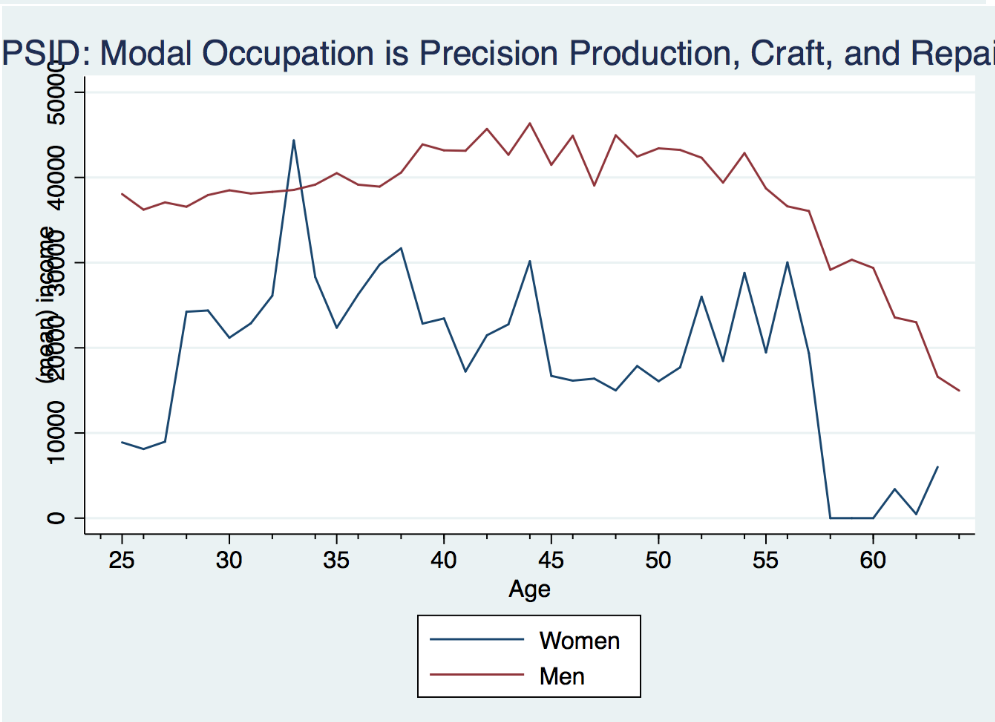


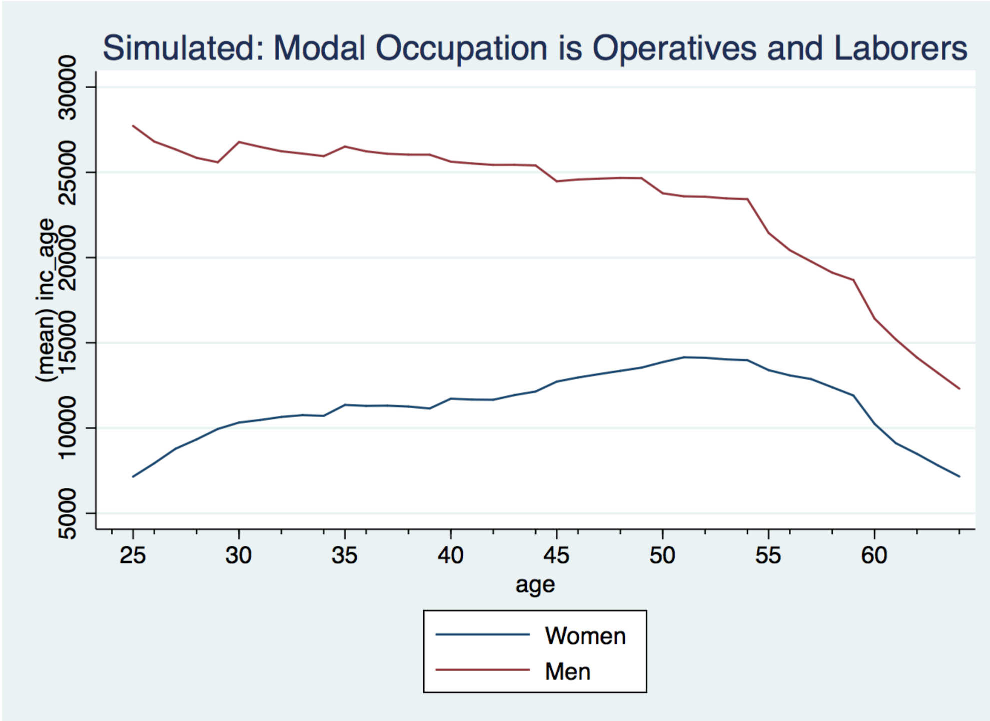
PSID: Modal Occupation is Service
(mean) income
0
20000
40000
60000
80000
25
30
35
40
45
50
55
60
Age
Women
Men

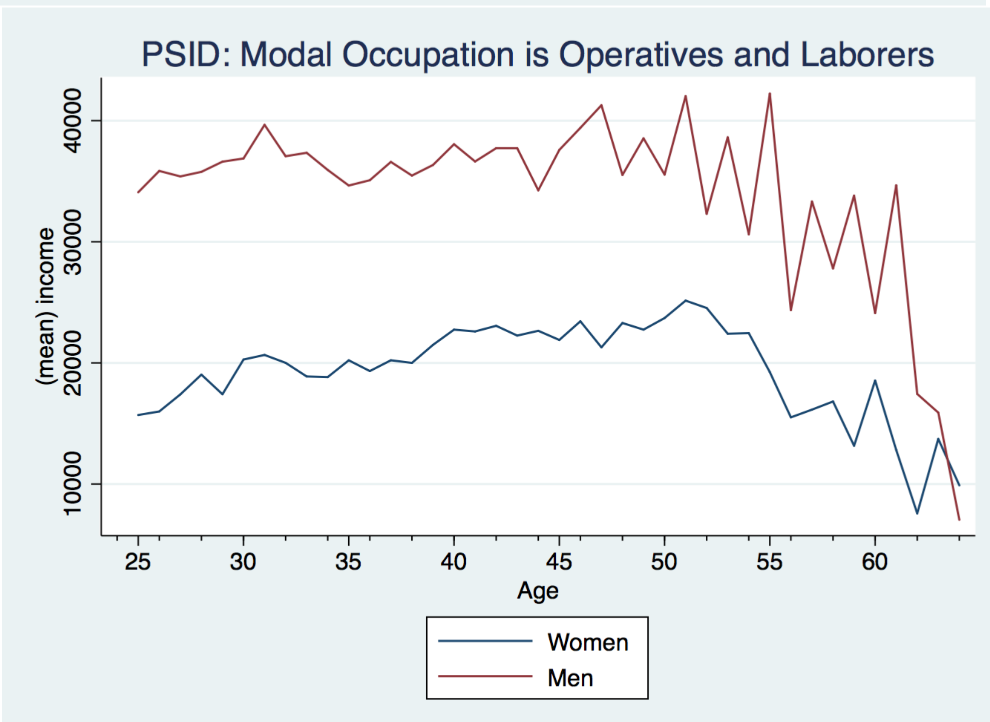

Simulated: Modal Occupation is Farming, Forestry, and Fishing
(mean) inc_age
0
5000
10000
15000
20000
25
30
35
40
45
50
55
60
age
Women
Men


PSID: Modal Occupation is Farming, Forestry, and Fishing
(mean) income
0
50000
100000
25
30
35
40
45
50
55
60
Age
Women
Men


mulated: Modal Occupation is Precision Production, Craft, and Rep
(mean) income
5000
10000
15000
20000
25000
30000
25
30
35
40
45
50
55
60
age
Women
Men


PSID: Modal Occupation is Precision Production, Craft, and Repai
(mean) income
0
10000
20000
30000
40000
50000
25
30
35
40
45
50
55
60
Age
Women
Men


Simulated: Modal Occupation is Operatives and Laborers
(mean) inc_age
5000
10000
15000
20000
25000
30000
25
30
35
40
45
50
55
60
age
Women
Men


PSID: Modal Occupation is Operatives and Laborers
(mean) income
10000
20000
30000
40000
25
30
35
40
45
50
55
60
Age
Women
Men

10.7 More Model Results

10.7.1 Model Job Parameter Estimates by Year

Table 7: Model Parameters in Year 1960, 34 occupations

	Parameter					
Cell	WTP_O^M	WTP_O^F	u_O^M	u_O^F	W_O^M	W_O^F
Teachers, Postsecondary	1.2172	0.7596	-4.4661	-4.7676	0.6044	-0.0037
Teachers, Except Postsecondary	1.0120	0.5831	-3.3220	-2.3175	0.6391	-0.3278
Social Scientists, Lawyers, Judg	1.1729	0.4690	-4.6958	-4.4071	0.3193	-0.2285
Social, Recreation, and Religiou	0.8824	0.4863	-4.3327	-4.8226	0.1994	-0.2268
Writers, Artists, Entertainers,	1.0303	0.5511	-3.8738	-5.0703	0.3755	-0.3118
Health Technologists and Technic	0.9396	0.3388	-5.1452	-4.3801	0.5089	-0.5843
Technicians except health	1.1164	0.4534	-2.9334	-4.4064	0.4432	-0.0877
Sales Representatives, Finance a	1.0251	0.3942	-3.4389	-4.3574	0.2443	-0.4652
Sales Workers, Retail and Person	1.0403	0.3506	-3.0461	-3.6052	0.2672	-0.6304
Administrative Support	0.9958	0.4188	-4.7503	-2.8073	0.6827	-0.4682
Records Processing Occupations,	0.9326	0.3738	-5.8100	-5.3070	0.7269	-0.2652
Financial Records Processing Occ	0.7146	0.3930	-6.2295	-3.6937	0.6474	-0.5516
Mail and Material Distribution	0.9261	0.3033	-3.0598	-4.8226	0.1383	-0.1591
Adjusters and Investigators	1.0260	0.3332	-4.8676	-4.2014	0.4616	-0.4004
Miscellaneous Administrative Sup	0.9045	0.2979	-5.4303	-4.2649	0.5581	-0.4877
Protective Service	0.9635	0.4742	-3.0175	-5.3784	0.1320	-0.2374
Food Preparation and Service Occ	0.7247	0.2850	-3.3989	-3.8144	-0.0790	-0.7870
Health Service Occupations	0.7890	0.2686	-4.9345	-3.2490	0.3408	-0.8057
Cleaning and Building Service Oc	0.8820	0.4148	-3.6455	-4.8723	-0.0454	-0.5735
Private Household and Personal S	0.7351	0.1578	-4.5539	-3.6180	0.0810	-1.1327
Agriculture, Forestry and Fishin	0.7238	0.0253	-3.0179	-5.4079	-0.6743	-1.1074
Mechanics and Repairers	1.0543	0.2546	-2.2921	-7.5489	-0.0515	0.1925
Construction and Extraction	0.9734	-0.1515	-2.5577	-9.7735	-0.1102	0.2706
Precision Production Occupations	0.9952	0.3811	-3.1632	-4.6940	0.2647	-0.4447
Metal, Wood, Plastic, Print, Tex	0.8522	0.1388	-3.8371	-5.2715	0.0894	-0.5540
Machine Operators, Fabricators,	1.0357	0.4154	-2.6545	-4.6327	0.0870	-0.2509
Road, Rail and Water Transportat	0.9883	0.1545	-2.3414	-5.8261	-0.0411	-0.3337
Material Moving, Laborers	0.9102	0.2057	-2.9407	-5.4989	-0.0988	-0.2205
Executive, Administrative, and M	1.2354	0.6225	-1.9349	-3.1201	0.2802	-0.4078
Management Related Occupations	1.1097	0.5735	-3.0822	-3.6301	0.4862	-0.1172
Engineers, Architects, and Surve	1.2312	0.5395	-2.9725	-6.0636	0.4938	0.4299
Math, Computer, and Natural Scie	1.1632	0.6152	-3.0649	-4.4403	0.5061	0.1337
Health Diagnosing Occupations	1.3804	0.6985	-4.3890	-5.6472	0.4494	0.2921
Health Assessment and Treating a	1.0311	0.4883	-5.6106	-2.7835	0.7352	-0.6271
Sample Size	1074937.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: π_O^G, u_O^G, W_O^G are in millions of dollars.

Table 8: Model Parameters in Year 1970, 34 occupations

	Parameter					
Cell	WTP_O^M	WTP_O^F	u_O^M	u_O^F	W_O^M	W_O^F
Teachers, Postsecondary	1.0378	0.5806	-4.3816	-4.7108	0.5620	0.0037
Teachers, Except Postsecondary	1.0738	0.6862	-3.2114	-2.2257	0.7170	-0.1801
Social Scientists, Lawyers, Judg	1.0390	0.5392	-4.6455	-4.3220	0.5998	-0.0712
Social, Recreation, and Religiou	0.8302	0.5640	-4.3208	-4.7094	0.4046	-0.0132
Writers, Artists, Entertainers,	1.0148	0.5079	-3.8433	-5.0638	0.4387	-0.0738
Health Technologists and Technic	1.0415	0.5729	-5.1530	-4.3025	0.6979	-0.2070
Technicians except health	1.0795	0.6269	-2.8440	-4.3821	0.5592	0.1473
Sales Representatives, Finance a	1.0975	0.4910	-3.3800	-4.2919	0.5534	-0.1876
Sales Workers, Retail and Person	1.0291	0.3462	-3.0390	-3.5168	0.5253	-0.5072
Administrative Support	0.9147	0.5382	-4.7109	-2.7612	0.7350	-0.3625
Records Processing Occupations,	0.9238	0.4458	-5.9421	-5.1967	0.7305	-0.1696
Financial Records Processing Occ	0.9110	0.4700	-6.0692	-3.6569	0.8013	-0.4110
Mail and Material Distribution	1.0148	0.4742	-3.0451	-4.8024	0.2984	0.0073
Adjusters and Investigators	0.8593	0.4754	-4.7949	-4.1550	0.6199	-0.1897
Miscellaneous Administrative Sup	0.7916	0.3793	-5.5877	-4.1986	0.7148	-0.3937
Protective Service	0.9883	0.4068	-2.9930	-5.3838	0.4008	-0.1430
Food Preparation and Service Occ	0.8050	0.3274	-3.3697	-3.7957	0.1627	-0.6346
Health Service Occupations	0.7902	0.3312	-5.0258	-3.1907	0.5209	-0.6647
Cleaning and Building Service Oc	0.8716	0.4615	-3.6814	-4.8315	0.2176	-0.4393
Private Household and Personal S	0.8537	0.2601	-4.5727	-3.5939	0.3931	-0.8036
Agriculture, Forestry and Fishin	0.7805	0.1721	-3.1387	-5.4128	-0.1527	-0.6795
Mechanics and Repairers	1.0447	0.1959	-2.3051	-7.3916	0.2366	0.1250
Construction and Extraction	0.9535	-0.1715	-2.5474	-10.5486	0.2043	0.4063
Precision Production Occupations	0.9509	0.3478	-3.1487	-4.6005	0.4288	-0.2615
Metal, Wood, Plastic, Print, Tex	0.9528	0.1959	-3.7276	-5.1083	0.3202	-0.3262
Machine Operators, Fabricators,	0.8873	0.4155	-2.6837	-4.6337	0.2814	-0.2010
Road, Rail and Water Transportat	0.9628	0.1556	-2.3755	-5.8390	0.2066	-0.2036
Material Moving, Laborers	0.9768	0.3139	-2.9740	-5.5051	0.1697	-0.0983
Executive, Administrative, and M	1.1679	0.6119	-1.8793	-3.0921	0.5243	-0.0186
Management Related Occupations	1.0686	0.6239	-3.0727	-3.5556	0.6577	0.0091
Engineers, Architects, and Surve	1.1703	0.7949	-2.9268	-5.9790	0.6045	0.6233
Math, Computer, and Natural Scie	1.1212	0.7256	-3.0163	-4.4516	0.6271	0.2808
Health Diagnosing Occupations	1.2746	0.9399	-4.3047	-5.6772	0.6073	0.6194
Health Assessment and Treating a	1.1237	0.6543	-5.5567	-2.7386	0.9853	-0.2961
Sample Size	240966.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: π_O^G, u_O^G, W_O^G are in millions of dollars.

Table 9: Model Parameters in Year 1980, 34 occupations

	Parameter					
Cell	WTP_O^M	WTP_O^F	u_O^M	u_O^F	W_O^M	W_O^F
Teachers, Postsecondary	1.3748	0.8128	-4.4102	-4.6516	0.7356	0.1508
Teachers, Except Postsecondary	1.4623	0.9805	-3.2654	-2.1377	0.7480	-0.0776
Social Scientists, Lawyers, Judg	1.6884	0.8566	-4.5918	-4.1591	1.0246	0.0235
Social, Recreation, and Religiou	1.4461	0.6149	-4.3058	-4.7094	0.4645	0.0230
Writers, Artists, Entertainers,	1.3309	0.6392	-3.9303	-4.9613	0.6073	-0.0017
Health Technologists and Technic	1.5014	0.8087	-4.9844	-4.1469	0.7346	-0.1354
Technicians except health	1.4403	0.8664	-2.8511	-4.2699	0.6475	0.1214
Sales Representatives, Finance a	1.5355	0.6205	-3.3657	-4.1246	0.6485	-0.0213
Sales Workers, Retail and Person	1.4253	0.6057	-3.0255	-3.4817	0.6055	-0.2862
Administrative Support	1.1677	0.7664	-4.6899	-2.6380	0.8627	-0.2706
Records Processing Occupations,	0.9697	0.5726	-5.8679	-5.1914	0.6367	-0.1459
Financial Records Processing Occ	1.2462	0.6748	-6.2362	-3.5644	0.8515	-0.2816
Mail and Material Distribution	1.2029	0.7953	-3.0563	-4.6795	0.3832	0.0877
Adjusters and Investigators	1.1869	0.6818	-4.8040	-4.0445	0.6702	-0.1096
Miscellaneous Administrative Sup	1.2339	0.8201	-5.5452	-4.0982	0.7322	-0.2705
Protective Service	1.5128	0.4958	-2.9611	-5.3039	0.3773	-0.0614
Food Preparation and Service Occ	1.3383	0.8023	-3.3975	-3.7147	0.1914	-0.4413
Health Service Occupations	0.9987	0.6614	-4.9138	-3.1112	0.5256	-0.4620
Cleaning and Building Service Oc	1.2666	0.7861	-3.6561	-4.7970	0.2495	-0.2821
Private Household and Personal S	1.1041	0.5837	-4.5779	-3.5497	0.4141	-0.6208
Agriculture, Forestry and Fishin	1.3360	0.4549	-3.2015	-5.3199	0.1116	-0.2800
Mechanics and Repairers	1.4022	0.6539	-2.2954	-7.4806	0.3295	0.4568
Construction and Extraction	1.4108	0.0701	-2.5438	-10.1161	0.3074	0.3248
Precision Production Occupations	1.3469	0.7042	-3.1777	-4.6069	0.4817	-0.0694
Metal, Wood, Plastic, Print, Tex	1.3167	0.7088	-3.8322	-5.0682	0.3390	-0.2291
Machine Operators, Fabricators,	1.3133	0.5785	-2.7852	-4.5573	0.3511	-0.0316
Road, Rail and Water Transportat	1.3700	0.4922	-2.3480	-5.6906	0.2977	-0.0137
Material Moving, Laborers	1.4206	0.6603	-2.9591	-5.2711	0.2470	0.0316
Executive, Administrative, and M	1.5642	0.9717	-1.8949	-2.9339	0.7733	0.1509
Management Related Occupations	1.4323	0.9185	-3.0146	-3.4329	0.7502	0.1193
Engineers, Architects, and Surve	1.5239	0.7330	-2.9549	-5.9674	0.6929	0.5293
Math, Computer, and Natural Scie	1.4245	1.0840	-3.0084	-4.2767	0.7577	0.4043
Health Diagnosing Occupations	1.9595	1.3815	-4.3489	-5.4980	1.2891	1.0410
Health Assessment and Treating a	1.2099	0.8703	-5.5545	-2.6367	1.0373	-0.1329
Sample Size	1657234.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: π_O^G, u_O^G, W_O^G are in millions of dollars.

Table 10: Model Parameters in Year 1990, 34 occupations

	Parameter					
Cell	WTP_O^M	WTP_O^F	u_O^M	u_O^F	W_O^M	W_O^F
Teachers, Postsecondary	1.4215	1.0229	-4.3910	-4.6260	0.7381	0.3134
Teachers, Except Postsecondary	1.5292	1.0224	-3.3468	-2.1575	0.8015	0.0645
Social Scientists, Lawyers, Judg	1.6181	1.0624	-4.6556	-4.0730	1.0464	0.2708
Social, Recreation, and Religiou	1.5008	0.8352	-4.3250	-4.6979	0.4950	0.1664
Writers, Artists, Entertainers,	1.4082	0.8594	-3.9331	-4.8456	0.6154	0.1963
Health Technologists and Technic	1.6239	0.9248	-4.9472	-4.1142	0.7941	0.0679
Technicians except health	1.8339	0.8854	-2.8416	-4.2376	0.6855	0.3427
Sales Representatives, Finance a	1.5266	0.9000	-3.3313	-4.0149	0.6728	0.1972
Sales Workers, Retail and Person	1.3488	0.8793	-3.0389	-3.4382	0.6006	-0.0939
Administrative Support	1.6771	0.8259	-4.6861	-2.6205	0.8507	-0.1150
Records Processing Occupations,	0.9872	0.8573	-5.8446	-5.1435	0.6042	0.0226
Financial Records Processing Occ	1.1747	0.8550	-6.2849	-3.5399	0.8385	-0.1375
Mail and Material Distribution	1.3651	0.8265	-3.0673	-4.6278	0.3738	0.2458
Adjusters and Investigators	1.2171	0.7767	-4.8024	-4.0299	0.6897	0.0381
Miscellaneous Administrative Sup	1.1751	0.8711	-5.6052	-4.0450	0.7583	-0.1342
Protective Service	1.5901	0.7314	-2.9706	-5.1774	0.4157	0.2626
Food Preparation and Service Occ	1.3811	0.8797	-3.3672	-3.6656	0.1822	-0.2664
Health Service Occupations	1.1954	0.9196	-4.9125	-3.0666	0.5577	-0.3206
Cleaning and Building Service Oc	1.3528	0.8199	-3.6332	-4.7489	0.2303	-0.1403
Private Household and Personal S	1.1982	0.7356	-4.5968	-3.5431	0.4331	-0.4284
Agriculture, Forestry and Fishin	1.3050	0.6502	-3.1816	-5.3037	0.0868	-0.0887
Mechanics and Repairers	1.5007	0.8642	-2.3077	-7.2150	0.3290	0.6378
Construction and Extraction	1.4193	0.5269	-2.5147	-9.9853	0.2809	0.5783
Precision Production Occupations	1.4282	0.9097	-3.2165	-4.5722	0.4912	0.1027
Metal, Wood, Plastic, Print, Tex	1.5301	0.6235	-3.8668	-5.0806	0.3225	-0.0859
Machine Operators, Fabricators,	1.4648	0.8149	-2.8230	-4.5154	0.3420	0.1201
Road, Rail and Water Transportat	1.3434	0.7588	-2.3645	-5.6735	0.2556	0.1975
Material Moving, Laborers	1.3442	0.5969	-2.9418	-5.2648	0.2203	0.1547
Executive, Administrative, and M	1.6418	1.1066	-1.9196	-2.8495	0.8096	0.3597
Management Related Occupations	1.5265	1.1827	-3.0332	-3.3598	0.7992	0.2995
Engineers, Architects, and Surve	1.6579	1.1230	-2.9598	-5.7571	0.7579	0.7689
Math, Computer, and Natural Scie	1.6781	1.1797	-3.0117	-4.1879	0.8027	0.5831
Health Diagnosing Occupations	2.1196	1.5187	-4.4414	-5.3547	1.3918	1.1024
Health Assessment and Treating a	1.4239	1.2457	-5.5958	-2.5758	1.1073	0.1225
Sample Size	1858609.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: π_O^G, u_O^G, W_O^G are in millions of dollars.

Table 11: Model Parameters in Year 2000, 34 occupations

	Parameter					
Cell	WTP_O^M	WTP_O^F	u_O^M	u_O^F	W_O^M	W_O^F
Teachers, Postsecondary	1.4420	1.1501	-4.4311	-4.6181	0.7311	0.4153
Teachers, Except Postsecondary	1.4003	1.1580	-3.3246	-2.1140	0.7816	0.0902
Social Scientists, Lawyers, Judg	1.6724	1.2999	-4.6655	-4.0902	1.0763	0.4614
Social, Recreation, and Religiou	1.2656	0.9248	-4.3809	-4.6966	0.5482	0.2370
Writers, Artists, Entertainers,	1.4115	1.0542	-3.9135	-4.8239	0.6319	0.3656
Health Technologists and Technic	1.5519	0.9116	-4.9338	-4.1392	0.8163	0.1849
Technicians except health	1.5804	1.0562	-2.8604	-4.2346	0.7813	0.5162
Sales Representatives, Finance a	1.4957	1.1377	-3.3765	-3.9837	0.6743	0.3182
Sales Workers, Retail and Person	1.4777	0.9332	-3.0550	-3.4484	0.6210	0.0617
Administrative Support	1.3810	0.9830	-4.7224	-2.6534	0.8438	-0.0003
Records Processing Occupations,	1.1595	0.8594	-5.8807	-5.1703	0.5482	0.0747
Financial Records Processing Occ	1.3184	1.0960	-6.1801	-3.5664	0.8478	-0.0183
Mail and Material Distribution	1.4668	0.8400	-3.0875	-4.6731	0.3391	0.3294
Adjusters and Investigators	1.2270	0.9134	-4.7949	-3.9796	0.6665	0.1246
Miscellaneous Administrative Sup	1.1992	1.0463	-5.5961	-4.0873	0.7394	-0.0086
Protective Service	1.5107	0.9617	-2.9458	-5.1247	0.4579	0.4173
Food Preparation and Service Occ	1.4230	1.1193	-3.3490	-3.6797	0.2019	-0.1310
Health Service Occupations	1.1816	0.9489	-4.9640	-3.0635	0.5697	-0.2201
Cleaning and Building Service Oc	1.2343	0.9122	-3.6698	-4.7459	0.2423	-0.0192
Private Household and Personal S	1.1597	0.8602	-4.5978	-3.5068	0.5029	-0.2861
Agriculture, Forestry and Fishin	1.2721	0.7399	-3.1905	-5.3018	0.1066	0.0818
Mechanics and Repairers	1.3908	0.8065	-2.3321	-7.4017	0.3296	0.7421
Construction and Extraction	1.4289	0.7245	-2.5569	-9.5942	0.2854	0.5417
Precision Production Occupations	1.3174	1.0547	-3.2717	-4.5768	0.4766	0.1887
Metal, Wood, Plastic, Print, Tex	1.4164	0.8117	-3.8794	-5.1480	0.2991	0.0404
Machine Operators, Fabricators,	1.4028	0.9471	-2.8748	-4.5470	0.3341	0.1974
Road, Rail and Water Transportat	1.3166	0.8678	-2.3967	-5.6964	0.2539	0.3538
Material Moving, Laborers	1.2597	0.8491	-2.9928	-5.2885	0.2142	0.2500
Executive, Administrative, and M	1.6264	1.2588	-1.9368	-2.8519	0.8453	0.5498
Management Related Occupations	1.5012	1.1989	-3.0498	-3.3391	0.8400	0.4372
Engineers, Architects, and Surve	1.5286	1.0979	-3.0099	-5.7484	0.7842	0.8869
Math, Computer, and Natural Scie	1.5252	1.3913	-2.9501	-4.1623	0.8053	0.6729
Health Diagnosing Occupations	1.9179	1.4929	-4.4379	-5.3372	1.2720	1.2231
Health Assessment and Treating a	1.4127	1.1189	-5.5600	-2.5833	1.1763	0.2522
Sample Size	1738053.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: π_O^G, u_O^G, W_O^G are in millions of dollars.

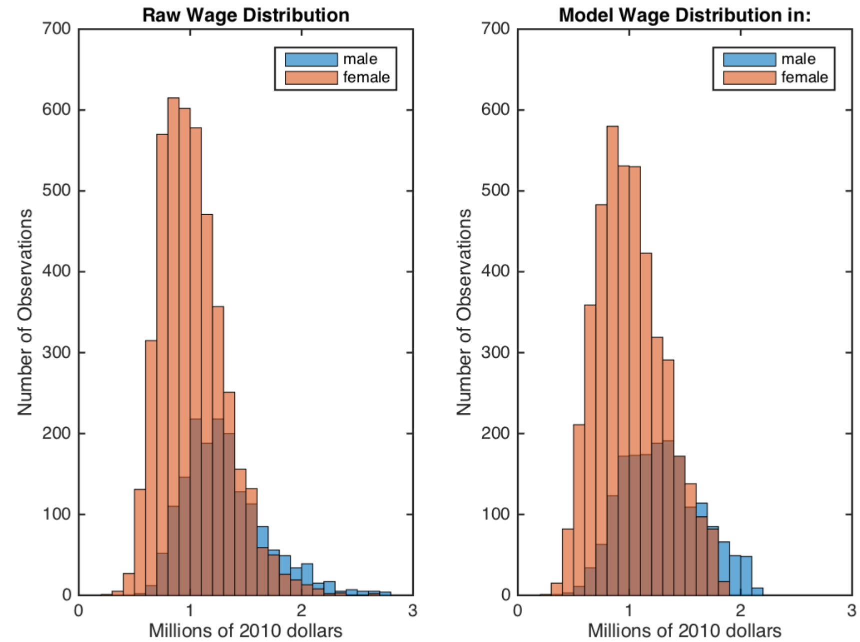
Table 12: Model Parameters in Year 2012, 34 occupations

	Parameter					
Cell	WTP_O^M	WTP_O^F	u_O^M	u_O^F	W_O^M	W_O^F
Teachers, Postsecondary	1.5608	1.1675	-4.4684	-4.5895	0.7045	0.4667
Teachers, Except Postsecondary	1.7397	1.1186	-3.3111	-2.0748	0.7801	0.1475
Social Scientists, Lawyers, Judg	1.6434	1.4886	-4.7228	-4.0513	1.1312	0.5822
Social, Recreation, and Religiou	1.4694	1.0101	-4.3148	-4.6205	0.5406	0.2912
Writers, Artists, Entertainers,	1.6915	1.0729	-3.9245	-4.8027	0.6125	0.4532
Health Technologists and Technic	1.7555	1.1267	-4.8795	-4.1567	0.8308	0.2806
Technicians except health	1.6944	1.2244	-2.8843	-4.2467	0.8139	0.5883
Sales Representatives, Finance a	1.4449	1.1210	-3.3881	-4.0110	0.6312	0.3698
Sales Workers, Retail and Person	1.5155	0.9346	-3.0584	-3.4472	0.5875	0.1234
Administrative Support	1.4332	1.1228	-4.7174	-2.6769	0.8317	0.0442
Records Processing Occupations,	1.0122	0.9501	-5.8731	-5.1711	0.5607	0.1380
Financial Records Processing Occ	1.1267	1.0160	-6.3201	-3.6148	0.8668	0.0365
Mail and Material Distribution	1.5170	0.9194	-3.0925	-4.6176	0.2730	0.3629
Adjusters and Investigators	1.3445	1.0603	-4.7962	-3.9760	0.5893	0.1432
Miscellaneous Administrative Sup	1.2197	0.9477	-5.6115	-4.0840	0.7133	0.0494
Protective Service	1.5032	0.9663	-2.9530	-5.0619	0.4670	0.5076
Food Preparation and Service Occ	1.2101	1.0318	-3.2994	-3.6188	0.1334	-0.1024
Health Service Occupations	1.3082	0.9183	-4.8893	-3.0074	0.5064	-0.1860
Cleaning and Building Service Oc	1.3930	0.9825	-3.6342	-4.7555	0.1823	0.0318
Private Household and Personal S	1.2612	0.8717	-4.5884	-3.4867	0.4407	-0.2224
Agriculture, Forestry and Fishin	1.4483	0.8146	-3.1800	-5.2743	0.0603	0.1703
Mechanics and Repairers	1.5542	0.9125	-2.3806	-7.4307	0.2727	0.7797
Construction and Extraction	1.5589	0.4001	-2.5966	-9.4066	0.2287	0.7637
Precision Production Occupations	1.4116	0.9894	-3.3167	-4.5754	0.4019	0.2169
Metal, Wood, Plastic, Print, Tex	1.5229	0.7905	-3.9841	-5.1799	0.2033	0.0789
Machine Operators, Fabricators,	1.4746	0.9317	-2.9499	-4.5624	0.2653	0.2286
Road, Rail and Water Transportat	1.4784	0.8664	-2.4077	-5.6869	0.1787	0.4277
Material Moving, Laborers	1.3408	0.8706	-2.9768	-5.3507	0.1341	0.2672
Executive, Administrative, and M	1.6637	1.3326	-1.9524	-2.8419	0.8280	0.6294
Management Related Occupations	1.5802	1.2634	-3.0466	-3.3560	0.8608	0.5493
Engineers, Architects, and Surve	1.6220	1.3404	-3.0308	-5.7120	0.8077	1.0230
Math, Computer, and Natural Scie	1.5359	1.4125	-2.9627	-4.1466	0.8056	0.7492
Health Diagnosing Occupations	2.2452	1.5624	-4.4854	-5.2399	1.4683	1.2922
Health Assessment and Treating a	1.5627	1.3546	-5.5500	-2.5317	1.2430	0.3723
Sample Size	974328.0000	0.0000	0.0000	0.0000	0.0000	0.0000

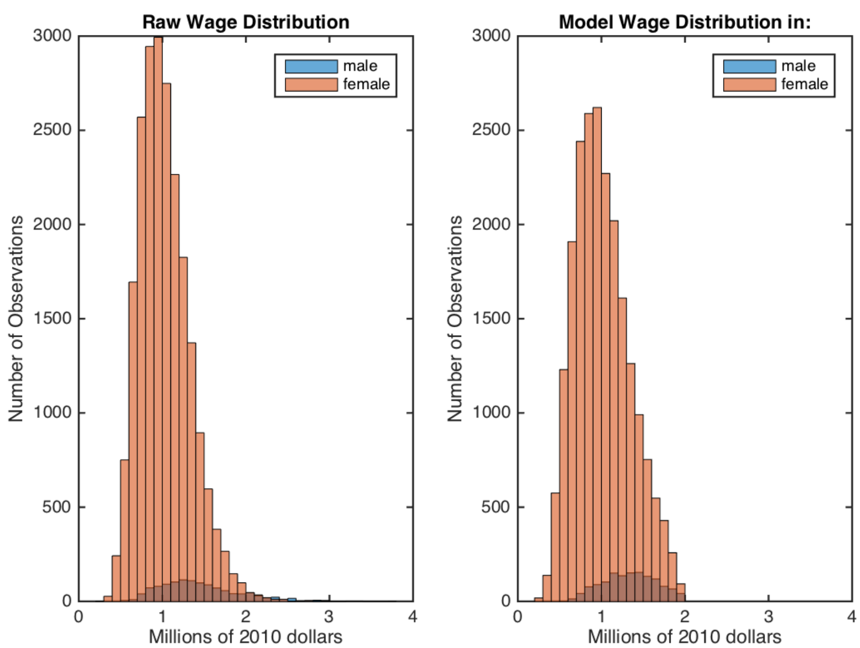
Note: π_O^G, u_O^G, W_O^G are in millions of dollars.

10.7.2 Model Fit for Years 2000/2012

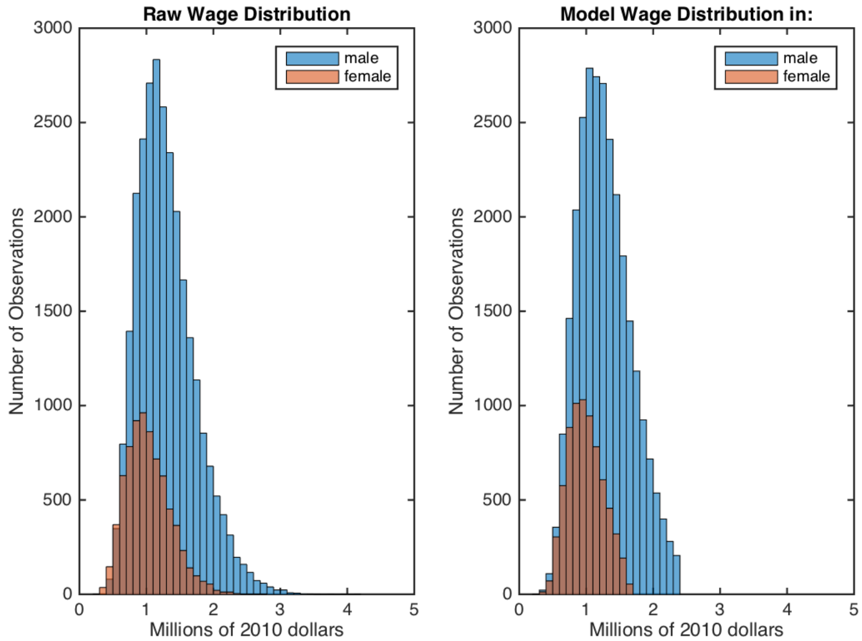
RecordsProcessingOccupations, data=0.72 model=0.708



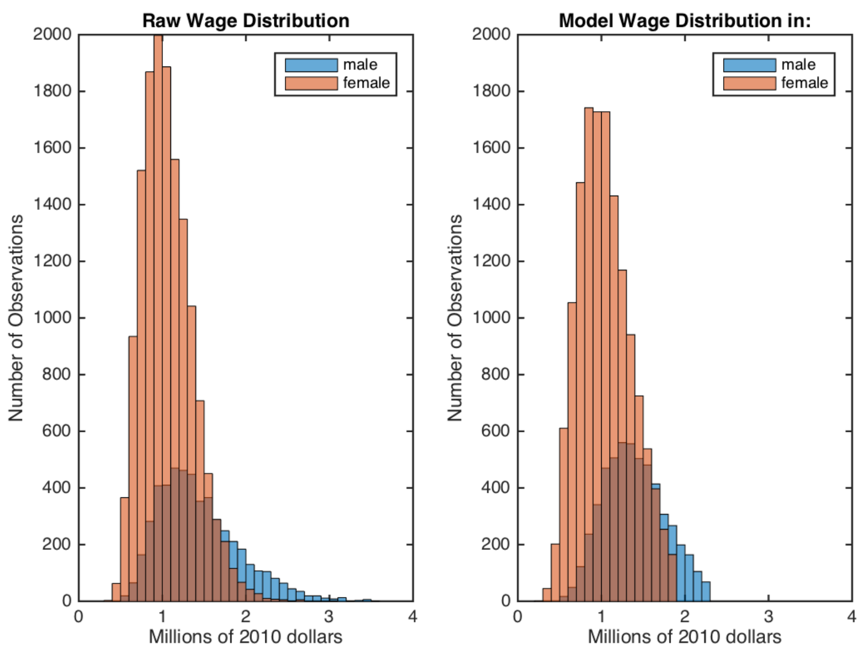
FinancialRecordsProcessingOcc data=0.949 model=0.941



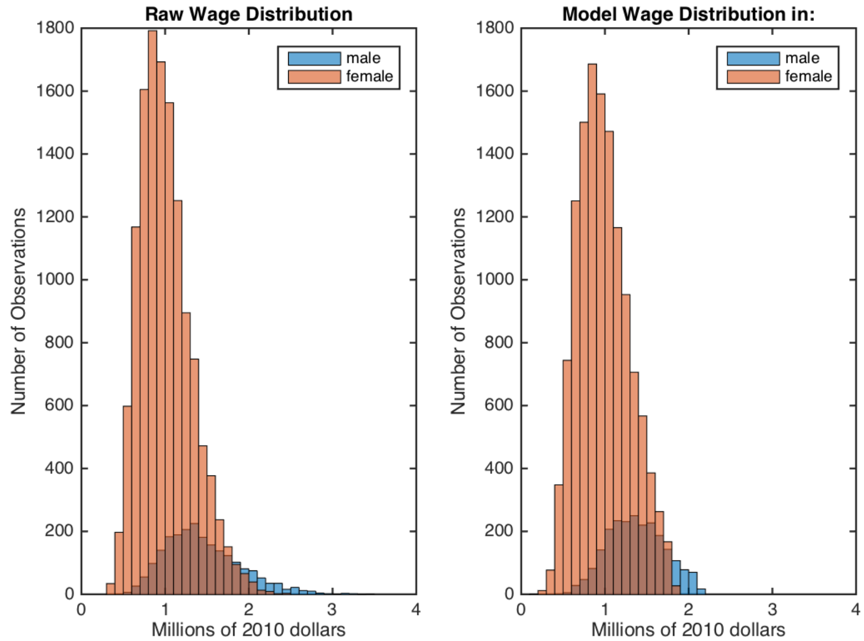
MailandMaterialDistribution data=0.216 model=0.208



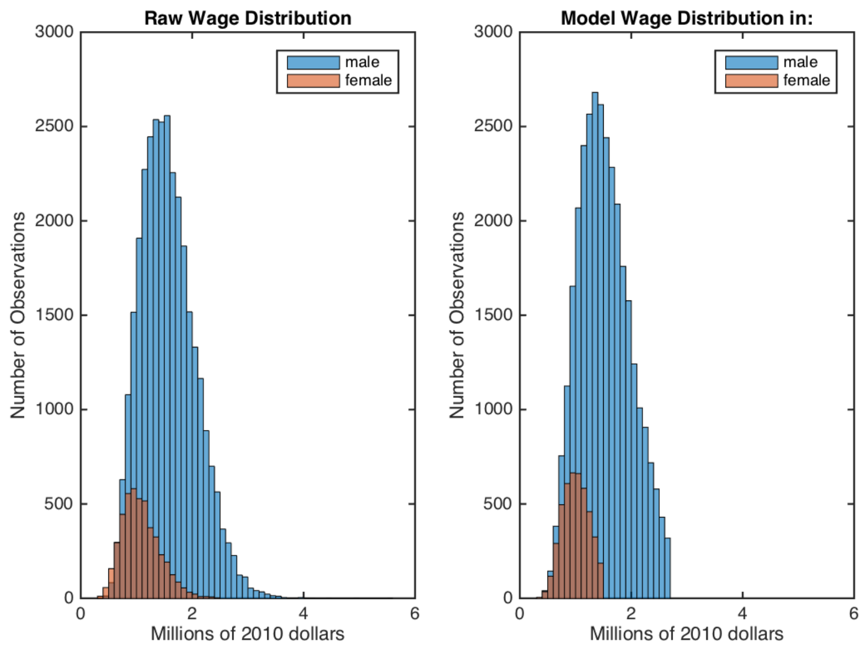
AdjustersandInvestigators data=0.743 model=0.726



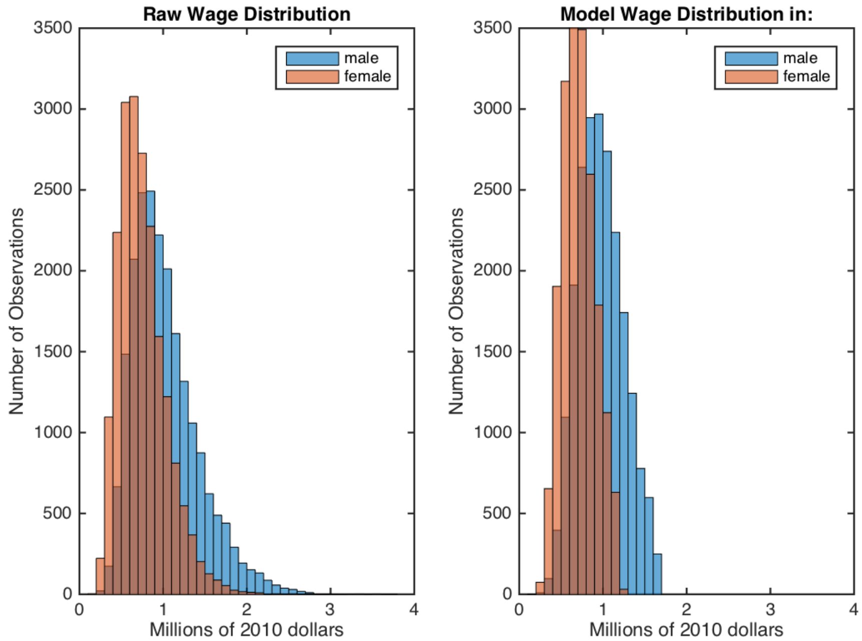
MiscellaneousAdministrativeSup data=0.856 model=0.85



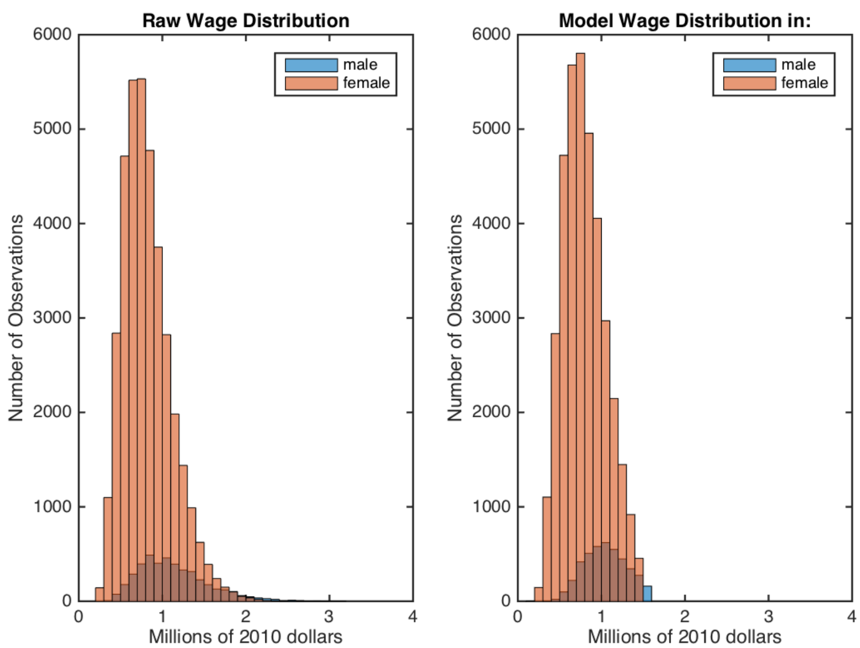
ProtectiveService data=0.127 model=0.122



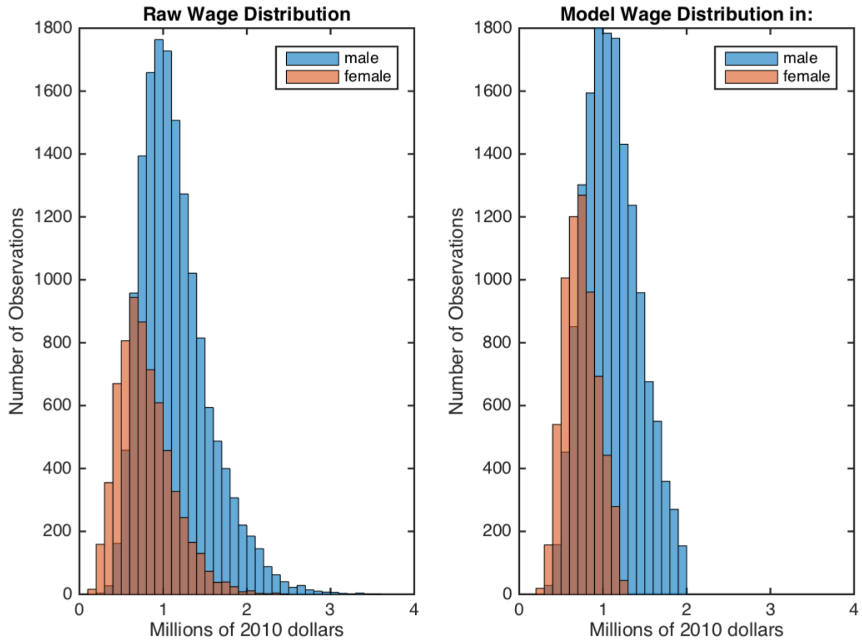
FoodPreparationandServiceOcc data=0.484 model=0.47



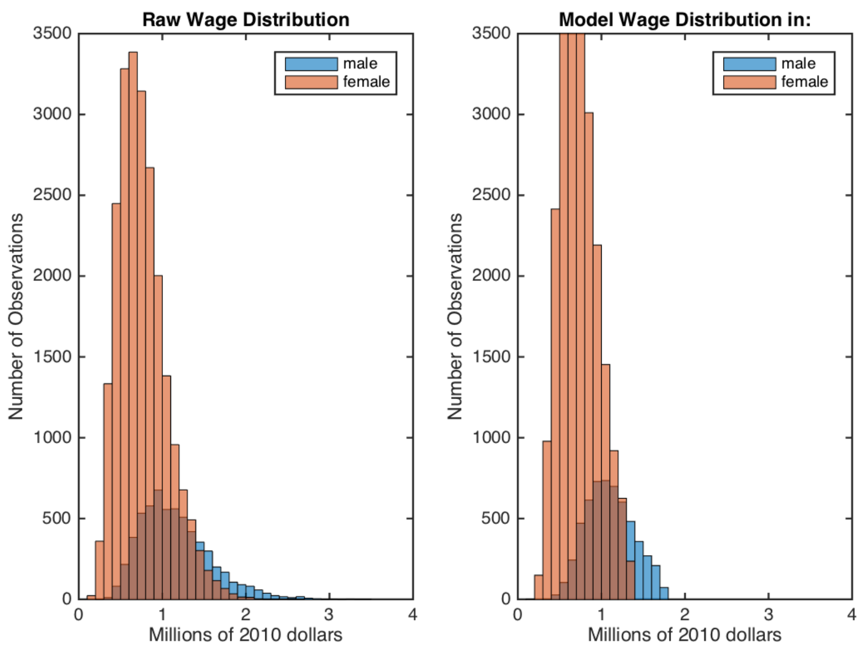
HealthServiceOccupations data=0.896 model=0.898



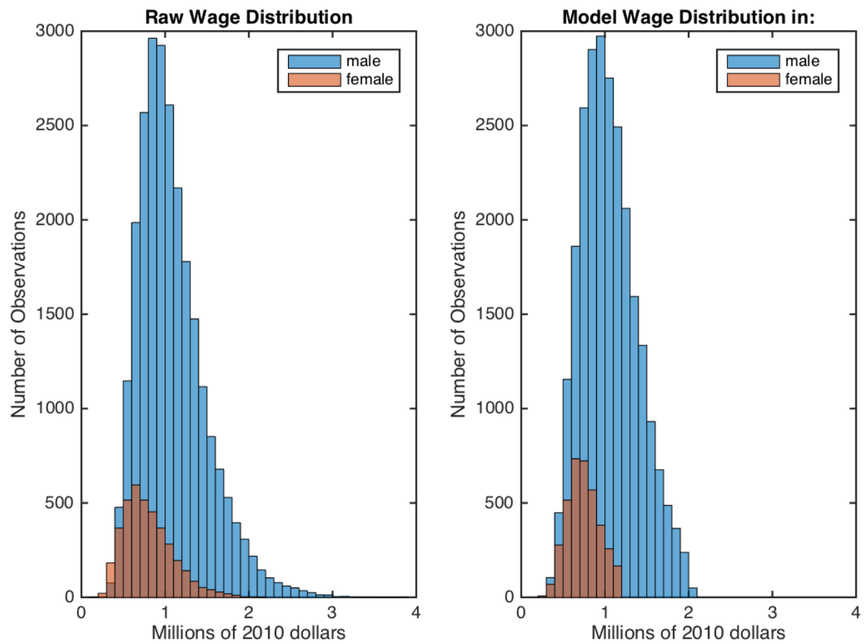
CleaningandBuildingServiceOc data=0.302 model=0.299



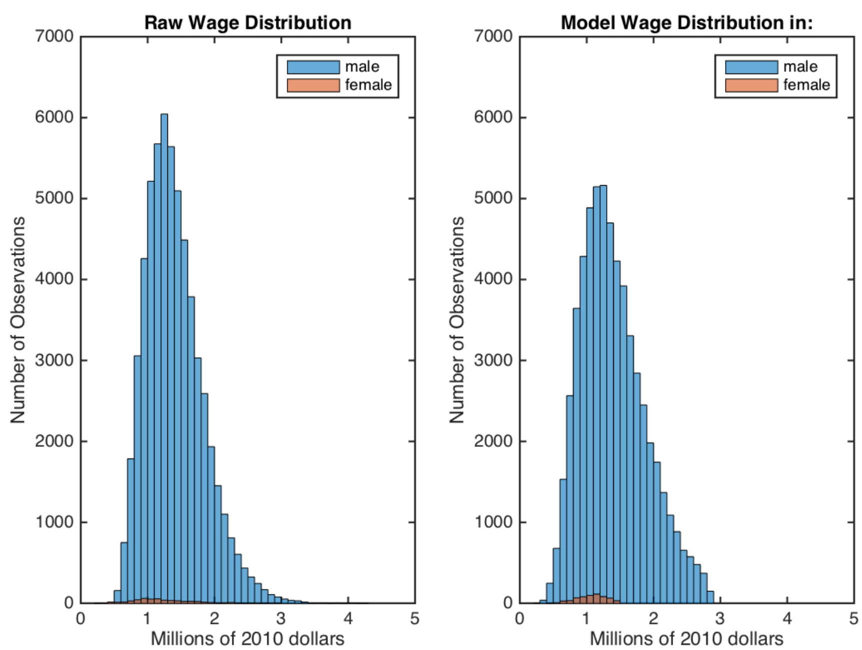
PrivateHouseholdandPersonalS data=0.792 model=0.806



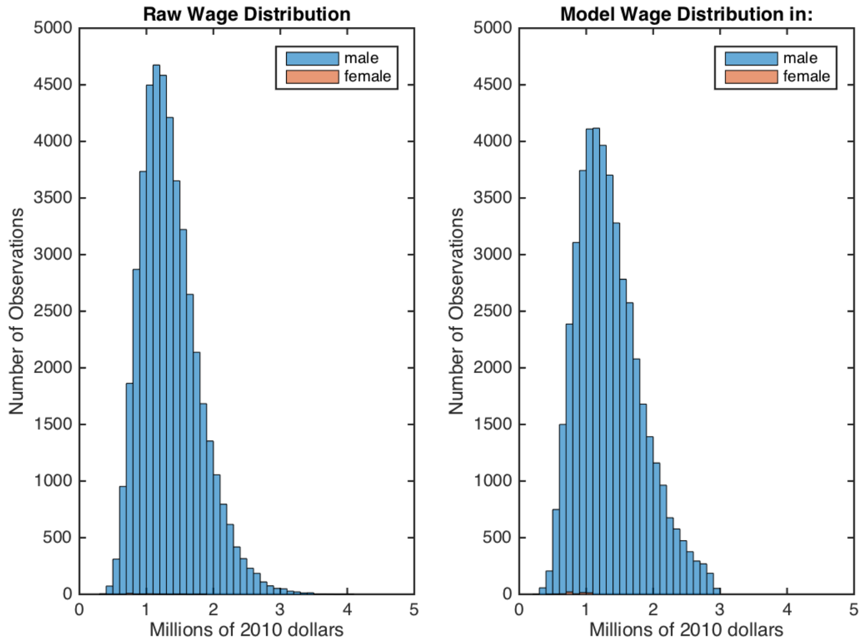
Agriculture,ForestryandFishin data=0.135 model=0.129



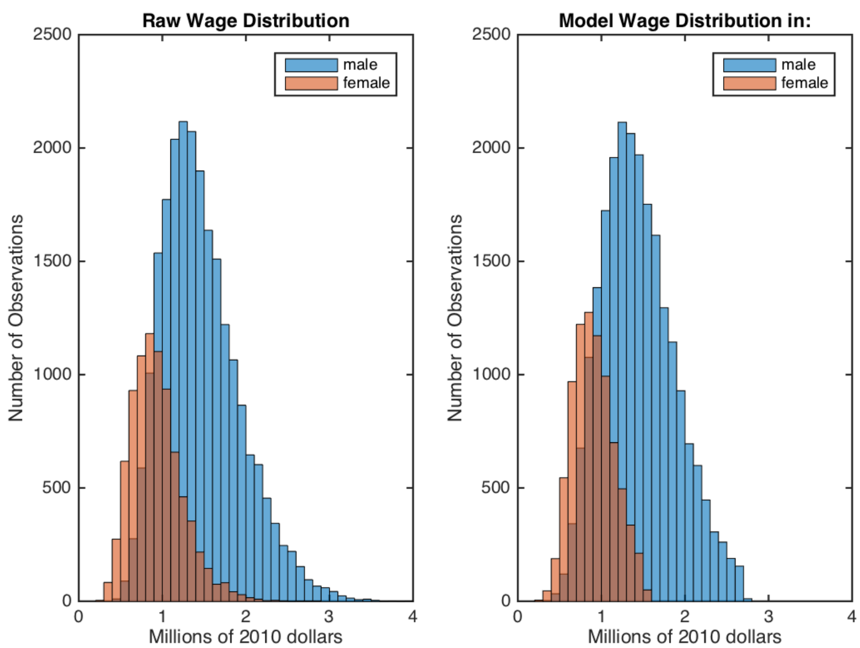
MechanicsandRepairers data=0.008 model=0.01



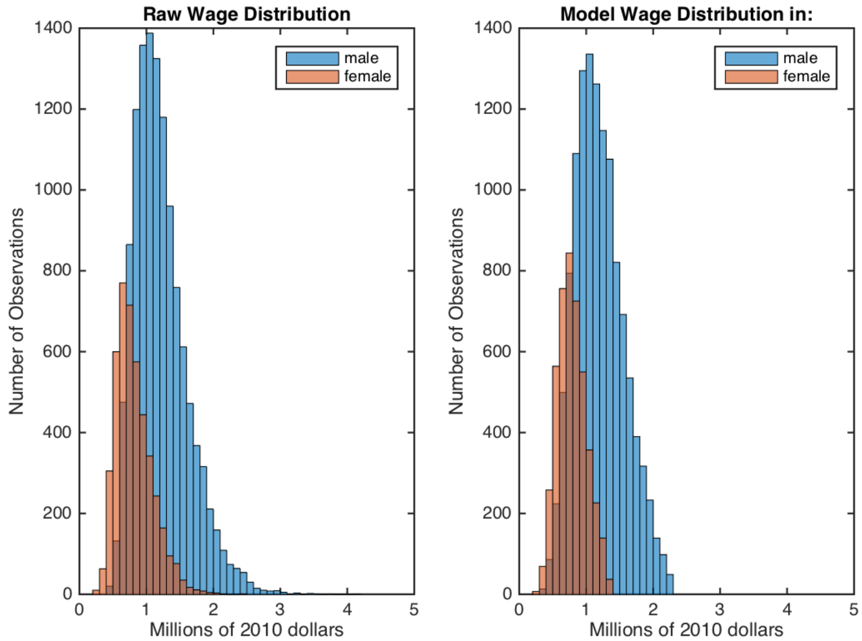
ConstructionandExtraction data=0.001 model=0.001



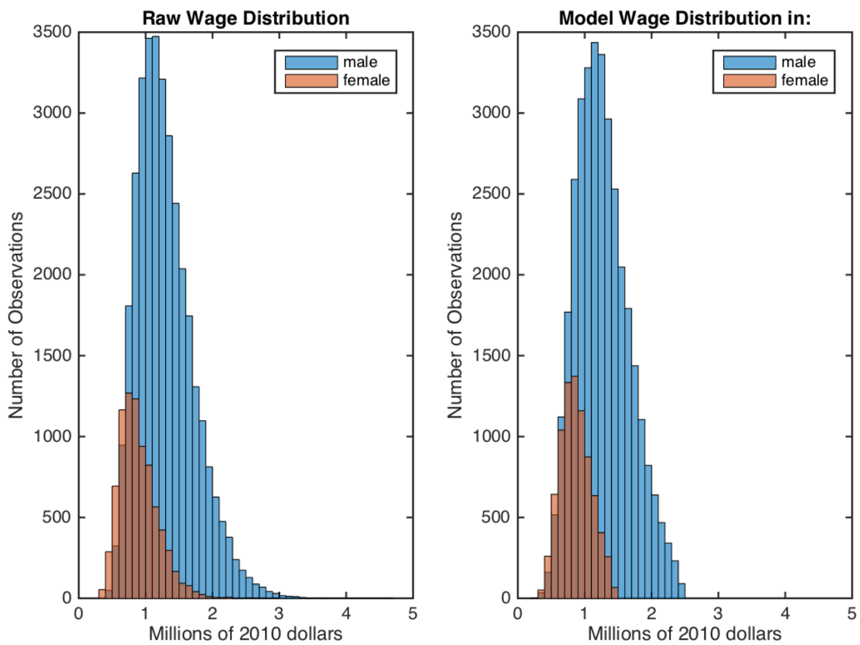
PrecisionProductionOccupations data=0.268 model=0.264



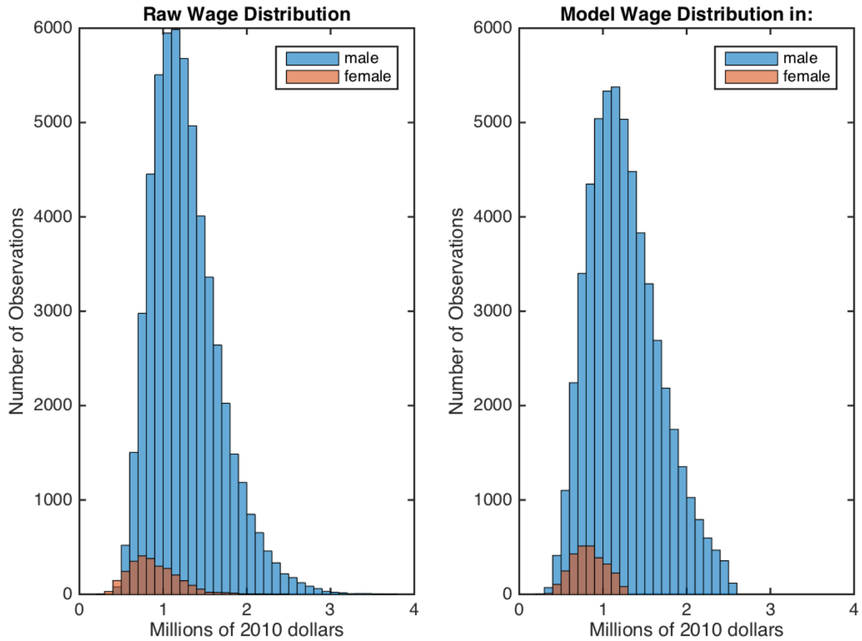
Metal,Wood,Plastic,Print,Tex data=0.269 model=0.273



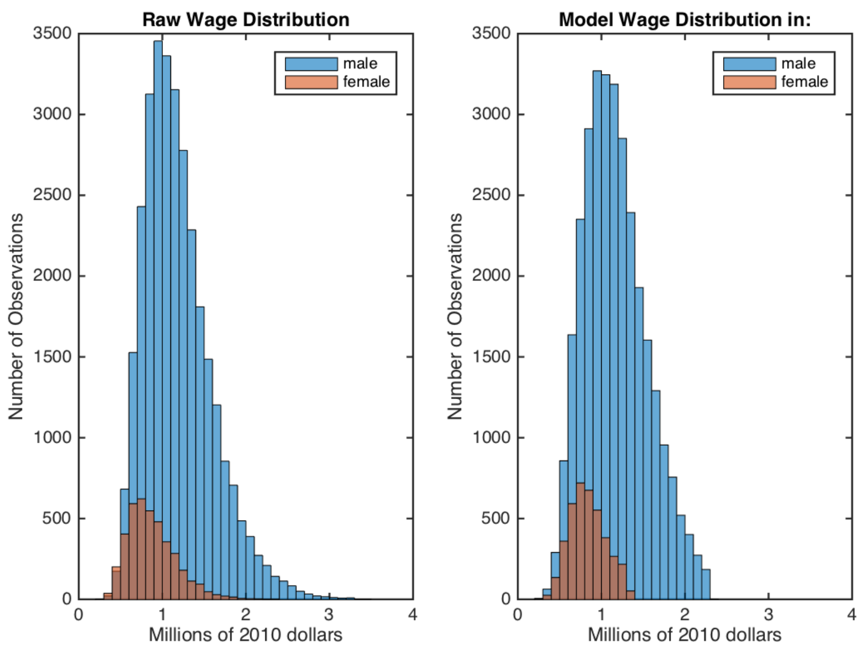
MachineOperators,Fabricators, data=0.195 model=0.193

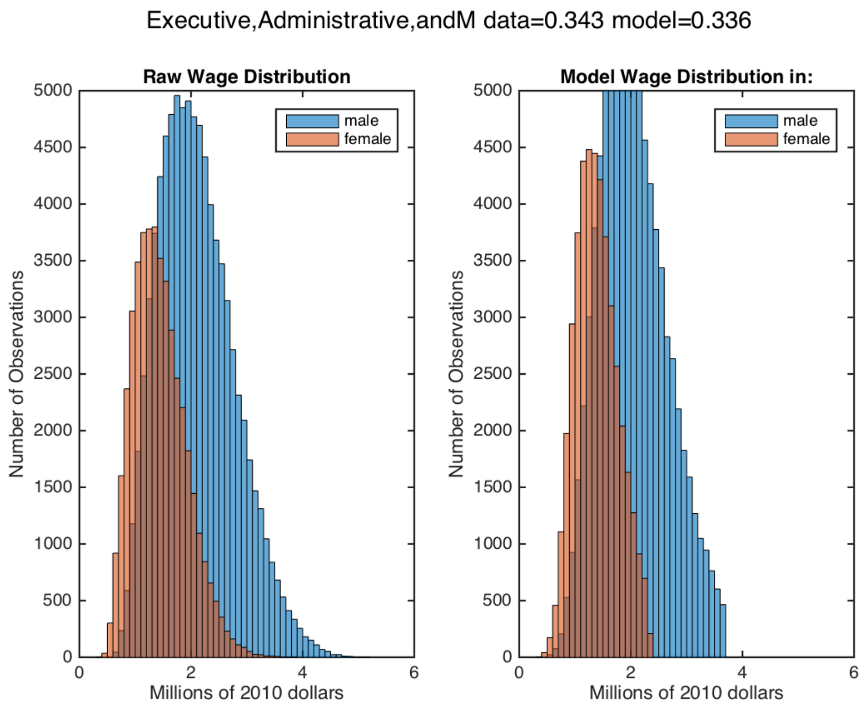


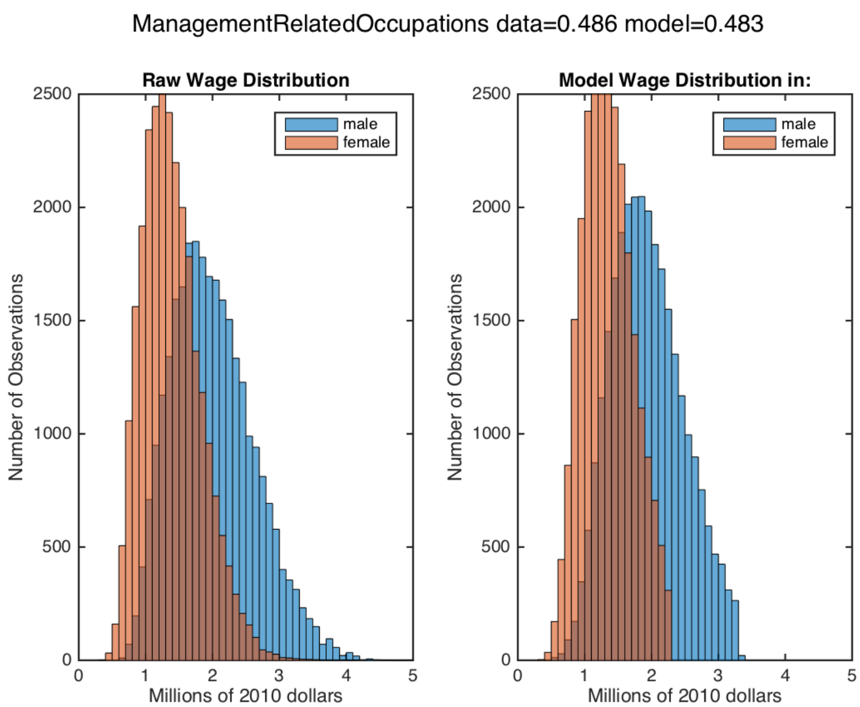
Road,RailandWaterTransportat data=0.047 model=0.049

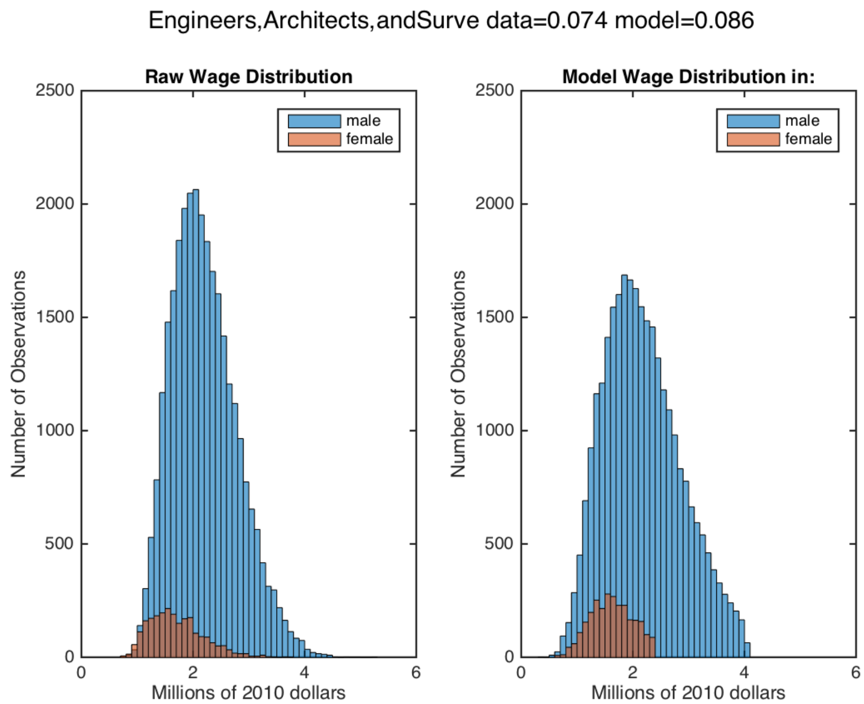


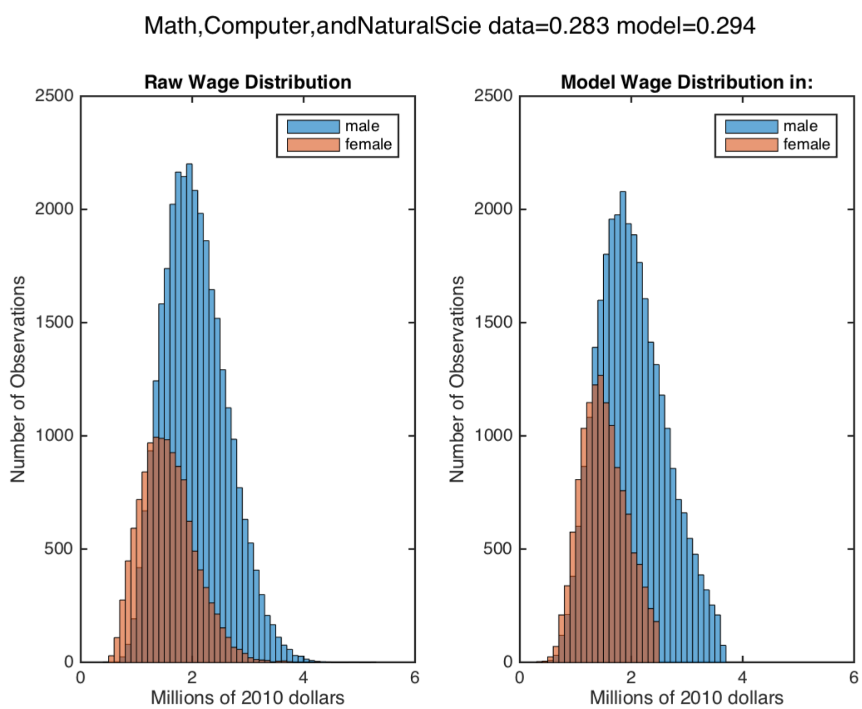
MaterialMoving,Laborers data=0.116 model=0.114



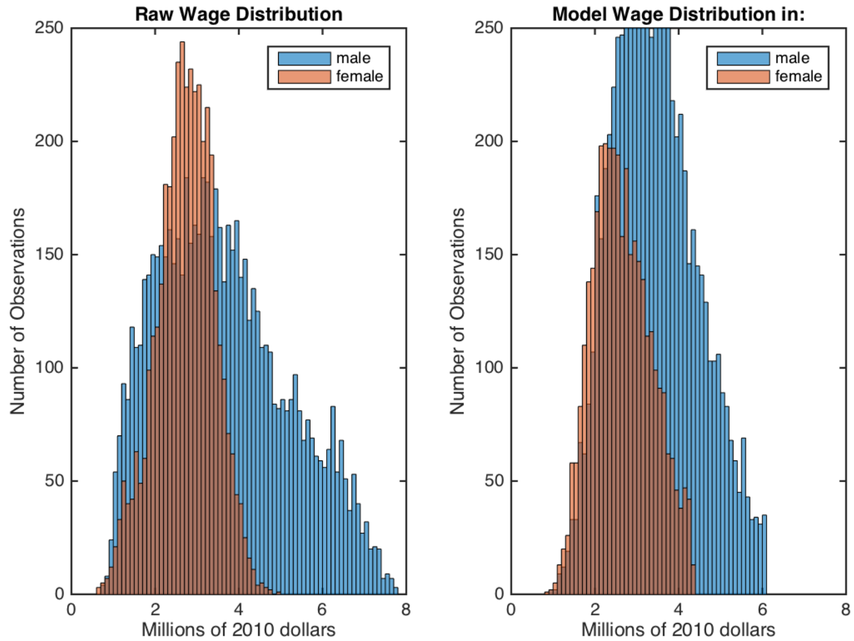

Executive,Administrative,andM data=0.343 model=0.336
Raw Wage Distribution
Model Wage Distribution in:
male
female
Number of Observations
Millions of 2010 dollars


ManagementRelatedOccupations data=0.486 model=0.483
Raw Wage Distribution
Model Wage Distribution in:
male
female
Number of Observations
Millions of 2010 dollars

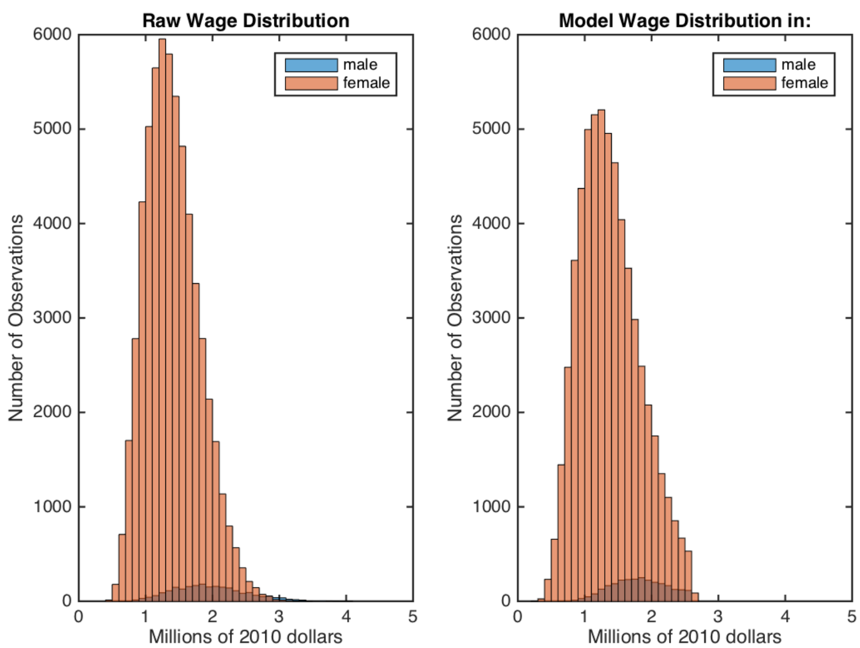

Engineers,Architects,andSurve data=0.074 model=0.086
Raw Wage Distribution
Model Wage Distribution in:
male
female
Number of Observations
Millions of 2010 dollars
0
500
1000
1500
2000
2500
0
2
4
6


Math,Computer,andNaturalScie data=0.283 model=0.294
Raw Wage Distribution
Model Wage Distribution in:
male
female
Number of Observations
Millions of 2010 dollars
0
500
1000
1500
2000
2500
0
2
4
6

HealthDiagnosingOccupations data=0.369 model=0.326



HealthAssessmentandTreatinga data=0.962 model=0.956



References

Addison, J. T., Wang, S., & Ozturk, O. D. (2017). The Occupational Feminization of Wages. *ILR Review*(Online). doi: 10.1177/0019793917708314

Albanesi, S., & Sahin, A. (2017). The Gender Unemployment Gap. *NBER Working Paper 23743*.

Altonji, J. G., & Card, D. (1991). *The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives* (No. January). doi: 10.1017/CBO9781107415324.004

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China Syndrome: Local labor market impacts of import Competition in the United States. *American Economic Review*, *103*(6), 2121–2168. doi: 10.1257/aer.103.6.2121

Baker, M., & Cornelson, K. (2018). Gender-Based Occupational Segregation and Sex Differences in Sensory, Motor, and Spatial Aptitudes. *Demography*, *55*(5), 1749–1775.

Berry, S. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, *25*(2), 242–262.

Bertrand, M. (2011). 17 - New Perspectives on Gender. *Handbook of Labor Economics, Volume 4B*, *4*(11), 1543–1590. Retrieved from `http://dx.doi.org/10.1016/S0169-7218(11)02415-4` doi: 10.1016/S0169-7218(11)02415-4

Blau, F. D., Brummund, P., & Liu, A. Y. H. (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. doi: 10.1007/s13524-012-0151-7

Caetano, G., & Maheshri, V. (2017). School segregation and the identification of tipping behavior. *Journal of Public Economics*, *148*, 115–135. Retrieved from http://dx.doi.org/10.1016/j.jpubeco.2017.02.009 doi: 10.1016/j.jpubeco.2017.02.009

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics*, *123*(1), 177–218.

Chiappori, P.-A., Salanié, B., & Weiss, Y. (2015). Partner Choice and the Marital College Premium: Analyzing Marital Patterns Over Several Decades. *Working Paper*, 1–58.

Choo, E., & Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, *114*(1), 175–201. doi: 10.1086/498585

Cortes, P., & Pan, J. (2017). Discussion paper series. *IZA Discussion Paper No. 10672*(6770).

Crawford, V. P., & Knoer, E. M. (1981). Crawford and Knoer 1981 ECTA.pdf. *Econometrica*, *49*(2), 437–450.

DeLeire, T., & Levy, H. (2004). Worker Sorting and the Risk of Death on the Job. *Journal of Labor Economics*, *22*(4), 925–953. doi: 10.1086/423159

Dupuy, A., & Galichon, A. (2017). A Note on the Estimation of Job Amenities and Labor Productivity. *IZA Discussion Paper 10900*.

Flood, S., King, M., Ruggles, S., & Robert, W. J. (2015). *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0.* Minneapolis: University of Minnesota. Retrieved from http://doi.org/10.18128/D030.V4.0

Fox, J. T. (2010). *Estimating Matching Games with Transfers.* Retrieved from http://www.nber.org/papers/w14382 doi: 10.3386/w14382

Galichon, A., & Salanié, B. (2013a). Cupids Invisible Hand: Social Surplus and Identification in Matching Models. *Open Access publications from Sciences Po*.

Galichon, A., & Salanié, B. (2013b). Cupids Invisible Hand: Social Surplus and

Identification in Matching Models. *Open Access publications from Sciences Po*.

Galichon, A., & Salanié, B. (2015). Cupids Invisible Hand: Social Surplus and Identification in Matching Models. *Open Access publications from Sciences Po*.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2017). Bartik Instruments : What, When, Why and How. *Working Paper*. Retrieved from `https://f247968a-a-62cb3a1a-s-sites.googlegroups.com/site/isaacsorkin/Bartik{_}Instruments.pdf?attachauth=ANoY7crDS4jlf2AwQ9sV{_}yXsmyiVzBOTZKZt1TXOvfVQHh6nRfbDedmYfQJr3t2WqZA97LDle5ts8-jB87LCRbVm7EcEo8PB27o{_}osQnOP8vPUVF51ImkFlQivimxqFd9`

Harris, J. M. (2018). *Do Wages Fall when Women Enter an Occupation ? New Evidence using an Instrumental-Variables Approach.* Retrieved from `https://sites.google.com/site/jorgenharris`

Hausman, J. A. (1996). *Valuation of new goods under perfect and imperfect competition* (Vol. I) (No. January). Retrieved from `http://www.nber.org/chapters/c6068.pdf` doi: 10.3386/w4970

Hsieh, C.-t., Hurst, E., Jones, C. I., & Klenow, P. J. (2016). The Allocation of Talent. *Growth (Lakeland)*.

Kambourov, G., & Manovskii, I. (2008). Rising Occupational and Industrial Mobility in the United States: 1968-1997. *International Economic Review*, *49*(1), 41–79.

Levanon, A., England, P., & Allison, P. (2009). Occupational Feminization and Pay: Assessing Causal Dynamics Using 1950-2000 U .S. Census Data. *Social Forces*, *88*(2), 865–891. doi: 10.1353/sof.0.0264

Lordan, G., & Pischke, J.-s. (2015). Does Rosie Like Riveting ? Male and Female Occupational Choices. *NBER Working Paper 22495*.

Macpherson, D. a., & Hirsch, B. T. (1995). Wages and Gender Composition: Why do Women's Jobs Pay Less? *Journal of Labor Economics*, *13*(3), 426. doi: 10.1086/298381

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, *69*(2), 307–342. doi: 10.1111/1468-0262.00194

Olivieri, E. (2014). Occupational Choice and the College Gender Gap.

Pan, J. Y. (2015). Gender Segregation in Occupations : The Role of Tipping and Social Interactions. *Journal of Labor Economics*, *33*(2), 365–408. doi: 10.1086/678518

Reed, W. R., & Dahlquist, J. (1994). Do women prefer women's work? *Applied Economics*, *26*(12), 1133–1144. doi: 10.1080/00036849400000111

Roth, A., & Sotomayor, M. (1990). *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., & Sobek, M. (2015). *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]*. Minneapolis: University of Minnesota. Retrieved from `http://doi.org/10.18128/D010.V6.0`

Salanié, B. (2014a). Identification in Separable Matching with Observed Transfers. *Working Paper*, 1–7. Retrieved from `http://www.columbia.edu/{~}bs2237/IdentObservedTransfers.pdf`

Salanié, B. (2014b). Identification in Separable Matching with Observed Transfers. , 1–7. Retrieved from `http://www.columbia.edu/{~}bs2237/IdentObservedTransfers.pdf`

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*. doi: 10.1080/0022250X.1971.9989794

Shapley, L. S., & Shubik, M. (1972). *The assignment game 1: The core.*

U.S. Bureau of Labor Statistics. (2017). *Job Openings and Labor Turnover*

Survey. Downloaded May 2017 from FRED Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/categories/32241/downloaddata.

U.S. Census Bureau. (2017). *Survey of Income and Program Participation.* Downloaded December 2016 from NBER: http://www.nber.org/data/survey-of-income-and-program-participation-sipp-data.html.

Usui, E. (2008). Job satisfaction and the gender composition of jobs. *Economics Letters*, *99*(1), 23–26. doi: 10.1016/j.econlet.2007.05.019

Wiswall, M., & Zafar, B. (2018). Preference for the Workplace, Human Capital, and Gender. *The Quarterly Journal of Economics*, *133*(1), 457–507. doi: 10.1017/CBO9781107415324.004