A Theory of Dual Job Search and Sex-Based Occupational Clustering

Alan Benson*

Updated October 2013

JEL Codes: J24, J16, J31, J12, J61

Keywords: Human Capital, Occupational Choice, Marriage,

Job Search, Economic Geography, Earnings Differentials

Abstract. This paper theorizes and provides evidence for the segregation of men into clustered occupations and women into dispersed occupations in advance of marriage and in anticipation of future colocation problems. Using the Decennial Census, and controlling for occupational characteristics, I find evidence of this general pattern of segregation, and also find that the minority of the men and women who depart from this equilibrium experience delayed marriage, higher divorce, and lower earnings. Results are consistent with the theory that marriage and mobility expectations foment a self-fulfilling pattern of occupational segregation with individual departures deterred by earnings and marriage penalties.

---

* Assistant Professor, Carlson School of Management, University of Minnesota. This paper is based upon Chapter 2 of my doctoral dissertation at the MIT Sloan School of Management. I am indebted to David Autor for comments and feedback throughout the project. I also thank Dan Fehder, Stephanie Hurder, Thomas Kochan, Ferran Mane, Matt Marx, Kathy McGinn, Paul Osterman, Al Roth, participants in the MIT Economics Labor Lunch, MIT Economic Sociology Working Group, the 2011 LBS Transatlantic Doctoral Conference, the 2011 IWAEE conference, the 2011 MOOD conference, the 2011 People and Organizations conference at Wharton, and colleagues at the Institute for Work & Employment Research for comments. Replication materials are available upon request. The usual disclaimer applies. Correspondence: bensona@umn.edu; Carlson School of Management, 321 19th Ave S. 3-300RR, Minneapolis, MN 55455. m: 760-898-2905.

Introduction

In 2010, 67% of families with at least one labor force participant featured dual-earner couples, up from 41% in 1980.[1] Given the frequency that young, career-oriented workers relocate for work or schooling and the impact these decisions may have on the career trajectories of the partner, it is not surprising that researchers' attention to the colocation problem has also grown. Mincer (1978) and Sandell (1977) introduce a neoclassical approach to household relocation decisions, noting that decisions optimizing the family's net career outcomes may impair the careers of "tied stayers" or "tied movers." Frank (1978) hypothesizes that the co-location problem will concentrate women in large cities. Costa and Kahn (2000) attribute the increasing concentration of college-educated "power-couples" in large metropolitan areas in part to their ability to sustain two simultaneous careers. Empirical studies have challenged the symmetry of these models due to evidence that household relocation decisions favor husbands' careers. Compton and Pollak (2007) find only the education of the husband predicts migration to large metropolitan areas. Others find household migration usually advances husbands' careers to the detriment of their wives (Boyle et al. 2001, Clark and Huang 2006, Jacobsen and Levin 2000, McKinnish 2008, Long 1974).

However, by emphasizing why and for whom household relocation decisions are made, research on couples' job search neglects the broader challenges and opportunities presented by variation in the geographic flexibility of jobs and potential endogeneity in occupational choice. Specifically, newly-trained physicists, computer scientists, naval architects, and nuclear engineers are typically geographically-constrained and require calculated moves early in their careers. In contrast, administrative assistants, dentists, school teachers, nurses, and general

[1] Figures based on author's calculation. Includes full-time, part-time, and unemployed labor force participants where both heads are age 18-65.

managers are ubiquitous and typically enjoy relative geographic-flexibility. Given that formative training and career investments are typically incurred prior to marriage, and given the difficulty of reconciling two careers each demanding calculated moves, do young men and women self-segregate into geographically-constrained and geographically-flexible occupations in anticipation of the colocation problem?

Section I introduces a model whereby early specific career investments, couples' desire to colocate, and variation in the geographic flexibility of jobs result in a coordination problem that prompts men and women to sort into geographically-clustered and dispersed occupations in advance of marriage, a phenomenon I refer to as "sex-based occupational clustering."[2] In the model, occupations are clustered or dispersed for reasons exogenous to their sexual composition, and highly-clustered occupations compensate for the disamenity of geographic constraints but penalize couples with two such careers due to the colocation problem. The model is consistent with a number of well-known features of inequality, including the tendency for women to be "tied-movers" when a family relocates for work and the segregation of women into lower-paying jobs. It also yields specific testable predictions regarding the marriage and earnings penalties that deter desegregation, and implies these predictions should be greatest for those a high degree of formal education.

Section II constructs an occupational clustering index. Using the 2000 Decennial Census, for each occupation, I use the index to calculate the share of workers in that occupation that would need to relocate to equalize its workers per capita in every MSA. I find workers in

---

[2] The model and timing is similar to others in which the household division of labor and childrearing expectations affect men's and women's *ex ante* training investments (see, for example, Becker 1991, Echevarria and Merlo 1999, Engineer and Welling 1999, and Hadfield 1999). However, while these models' *ex ante* investments follow from specialization in household production (and the inability to efficiently contract domestic work), *ex ante* investments in the model proposed here follow from variation in the geographic constraint and flexibility across occupations.

clustered occupations tend to have higher rates of relocation-for-work among never-married and single-earning men and women.

Section III confirms sex-based occupational clustering in which occupations dominated by women tend to be geographically-dispersed while those dominated by men tend to be geographically-clustered, after controlling for occupations' hours worked, mean age, physical strength requirements, the extent it involves assisting and caring for others, and the required math level. To provide an intuition of the magnitude, prime age dual-earner couples featuring a husband in an occupation that is more clustered than his wife's outnumber the reverse about two-to-one.

Section III also suggests men and women are deterred from departing from this pattern by earnings and marriage penalties. Women who enter geographically-clustered occupations experience lower earnings, later marriage, and higher divorce rates compared with those entering dispersed occupations. Men, in contrast, enjoy earnings premia and lower divorce rates in clustered occupations. Both sex-based occupational clustering and related effects on earnings, mobility, and marriage are most-pronounced among the college-educated, suggesting the highly-educated employ more occupationally-specific skills.

I. A Model of Occupational Clustering and Sex Segregation

This section provides sufficient conditions for equilibrium sex-based occupational clustering. The model considers risk-neutral workers who initially differ only in sex and who are interested in low marriage search costs and high family earnings. It is symmetric in that earnings and the costs of marriage search are not conditioned directly on sex.[3] The timing is as follows:

[3] Relaxing symmetry by conditioning earnings on sex initially shifts the mixed strategy Nash equilibrium and makes one pure strategy Nash equilibrium superior to the other, and at the extreme, results in a unique equilibrium. Conditioning earnings directly on sex may be justified by pregnancy (Polachek 1981; Engineer and Welling 1999), comparative advantage in childrearing (Becker 1965, 1985; Mincer and Polachek 1974),

1. Men and women choose to receive training[4] in either a clustered or dispersed occupation. Training is costless and binary, but may only be made in one occupation. Workers differ only in sex and are otherwise homogenous (i.e. in earnings by occupation and in marriage search costs). Workers choose training to maximize expected utility, which is additively-separable in their expected Period 2 marriage search costs and their Period 3 expected cumulative family earnings. If a sex is indifferent, its workers choose training in dispersed occupations.
2. Men and women engage in a costly marital search. Marriage is assortive, but only imperfectly so due to probabilistic "unconditional love" between pairs with two clustered or two dispersed occupations.
3. Families choose between two locations to maximize joint earnings. The key assumption is that compensating differentials make expected personal wages in clustered occupations superior to those in dispersed occupations only if that worker can choose a location to maximize personal wages. Therefore, expected family wages are greatest when one spouse enters a clustered occupation and the other enters a dispersed occupation.

The environment results in a coordination problem in *ex ante* human capital investments analogous to the classic "battle of the sexes" game, with two pure-strategy Nash equilibria conforming to sex-based occupational clustering and an unstable mixed-strategy equilibrium.

---

discrimination (Becker 1971), or natural developmental factors advantaging women in educational attainment (Buchmann, DiPrete, and McDaniel 2006; Goldin, Katz, and Kuziemko 2006). Conditioning the disutility of marriage search by sex may be justified, for example, by differences and men's and women's fertility by age.

[4] "Training" in an occupation should be interpreted broadly, and may include formal schooling, on-the-job training, progression through career ladders, and accrued tacit knowledge of how to succeed in one's field, formation of professional networks, and so on.

Conceptually, occupations are assumed to be clustered or dispersed for reasons exogenous to their sexual composition. This may be due to production technology (engineers or scientists with highly-specialized skills may benefit from local knowledge spillovers), input supply (mining is clustered around natural resources), or product markets (physicians and hairdressers are dispersed because they provide services that are universally demanded and must be done in-person). Ellison and Glaeser (1997) examine industrial agglomeration in greater detail. Household production might also be understood as a "dispersed" occupation, in that the productivity of a non-labor force participant is likely to be robust to exogenous relocations.[5]

The exposition proceeds by backward induction.

A. Period 3: Job Search Environment and Expected Household Earnings

The earnings subgame adapts a household relocation framework in the spirit of Mincer (1978). Unlike Mincer's model, employment opportunities in some occupations are location-invariant (dispersed occupations, such as secretaries) while employment opportunities in others differ by location (clustered occupations, such as nuclear engineers).

Let W_{ij} denote expected family earnings, which is the expected sum of the heads' wage rates, where i and j represent the occupations of the family heads. Occupations i and j may be clustered ("c") or dispersed ("d"). Because wages are not directly conditional on sex, W_{ij} can take three values: W_{cc}, W_{cd}, and W_{dd}.

To characterize W_{ij}, consider the following job search environment. Assume there are two locations, no relocation costs, and spouses must work in the same location. Families choose

[5] Engineer and Welling (1999) propose a similar model in which men and women marry out of love (analogous to $p_u = 1$ in period 2 of this model), resulting in a coordination problem that prompts men and women to specialize in domestic or external labor in advance of marriage. While their model's coordination problem arises from the complementarities of external and household labor (*i.e.* domestic production cannot be efficiently outsourced), this model coordination problem arises from the complementarities of clustered and dispersed occupations (*i.e.* individuals are bound by the desire to colocate).

location to maximize expected joint earnings and locate randomly if indifferent between the two locations. Because both spouses gain utility from W_{ij}, consider the decision rule of the family as the decision rule consistent with the individual interests spouses. Wages in dispersed occupations are location-invariant and fixed at w_D, while wages in clustered occupations are w_H in one location and w_L in the other, with $w_H > w_L$. Wages w_L may represent the best outside-option for geographic mismatch, which may involve foregoing opportunities for career-advancing work relocations and job transfers, switching occupations, or exiting the labor force entirely.[6] The next section confirms occupational clustering is strongly and positively correlated with relocation for work for never-married and for married men and women.

The first key assumption is that wages in clustered occupations compensate workers for the disamenity of geographic constraint and the *ex ante* risk of the colocation problem (*i.e.* displacement resulting in w_L), such that $w_H > w_D > w_L$. The second key assumption is that, if the worker is employed in one of the two locations randomly (*i.e.* due to locating for the spouse), expected wages in dispersed occupations are higher than those in clustered occupations, ie. $w_D \geq 0.5w_H + 0.5w_L$.[7] Intuitively, if a household chooses location to maximize one worker's wages (*i.e.* because the other spouse has a dispersed occupation and earns w_D regardless of location), expected personal wages will be superior in the clustered occupation, and if the household does not choose location for the worker (*i.e.* because the spouse has a clustered occupation), expected

6 Relocation is treated here as a single event that occurs after the location of w_H is revealed in Period 3. However, since relocation is costless, there is no discounting, and workers are risk-neutral, there is no material difference between this model and one in which a worker with a clustered occupation must dynamically relocate to receive w_H.

7 These assumptions could be treated as properties of a model where compensating differentials lead workers of one type (e.g. men) to segregate into occupations featuring a disamenity (e.g. geographic constraint), just as dangerous occupations pay a compensating differential and attract workers who are less risk-averse with regard to injury. Here, the disamenity of constraint for one sex is reduced (increased) by the segregation of the other sex into dispersed (clustered) occupations.

personal wages are superior in the dispersed occupation. I present evidence for compensating differentials for men and colocation penalties for women in the results section.

Note that, in the case of pure segregation by sex, workers of the sex segregating into clustered occupations always earn w_H, while workers of the other sex always earn w_D and are deterred from clustered occupations by the probability of w_L. While the first assumption prompts workers of one sex to enter clustered occupations, the second assumption captures the anti-complementarity of clustered occupations that exacerbates the colocation problem and deters workers of the other sex from also entering clustered occupations.[8]

For families with two clustered occupations in which high-wage opportunities are geographically-uncorrelated, there is a probability of 0.5 that wages w_H do not overlap and that one spouse earns w_L.[9] Expected wages are therefore $W_{cc} = 0.5(w_H + w_H) + 0.5(w_H + w_L)$. For families with exactly one clustered occupation, the family selects the location where the spouse with the clustered occupation earns w_H, and the family earns $W_{cd} = w_H + w_D$.[10] For families with two dispersed occupations, the location choice is immaterial and the family earns $W_{dd} = 2w_D$. The first key assumption implies $W_{cd} > W_{dd}$ and the second key assumption implies $W_{cd} \geq W_{cc}$.

B. Period 2: Marriage Search Environment and Expected Marriage Costs

---

8 Because clustering is treated as an exogenous feature of occupations and the empirical section treats clustering as relative, "pooling" into either clustered or dispersed occupations is neither meaningful or testable. Rather, we might instead interpret the latter assumption: the exogenously-determined distribution of occupations sufficiently wide such that some occupations are indeed sufficiently constrained, while others are indeed sufficiently flexible, to meaningfully impact the careers of spouses. Alternatively, rejecting the model's predictions may suggest the exogenous geographic distribution of occupations is insufficient to substantially impact spouses' careers.

9 In practice, opportunities are likely to be correlated, as about 4.5% of all dual-earner couples in the 2000 Census 5% PUMS work in the same occupation, more than would be expected by chance. The career compatibility of marriage partners who meet in the same city or in the same industry or occupation are also likely correlated. However, to the degree restricting the marital search to those within your same occupation, industry, or expected future location results in costly delayed marriage (ie. in period 2), the model's predictions are upheld.

10 In equilibrium, this is also consistent with research finding women are more-likely than men to be tied-movers.

In Period 2, workers engage in a marriage search. Suppose unmarried workers incur cost $c > 0$ to randomly-encounter an unmarried member of the opposite sex with whom, given that they have the opposite occupation type (they have "compatible careers"), they would marry. Furthermore, a share of matches $p_u \in [0, 1]$ would marry out of "unconditional love" (*i.e.* even if they have the same occupation type).[11] This quality is assumed to be match-specific and independent of past match quality. Therefore, with probability $(1 - p_u)$, a couple with the same occupational type will not marry and will delay marriage by re-entering the "pool" of unmarried workers, re-incurring the search cost. Let ψ_m and ψ_f denote the share of unmarried men and women in the pool of unmarried workers who work in clustered occupations.

Noting that pairs only delay marriage if they are in the same occupation type and they are not members of the subset p_u, the probability a worker of sex *i* in a clustered occupation marries in a given period of search is $[1 - \psi_j(1 - p_u)]$, and the probability a worker of sex *i* in a dispersed occupation marries in a given period is $[1 - (1 - \psi_j)(1 - p_u)]$. Because workers who re-enter the pool are again randomly-matched to a spousal candidate of the opposite sex, the probability a worker of sex *i* marries a worker of sex *j* in the opposite occupation type is equal to the probability the worker marries a worker in the opposite occupation type given that the worker marries that period. For example, the probability a worker of sex *i* in a clustered occupation marries a worker of sex *j* in a dispersed occupation is $(1 - \psi_j)[(1 - \psi_j) + p_u\psi_j]^{-1}$. These probabilities are used to calculate expected future family earnings in Period 1.

[11] Alternatively, in a continuous-time approach, potential marital matches may be thought of as a Markovian arrival process with a random non-mO*NETary match-specific utility component, where workers adopt a higher reservation match quality when the spousal candidate has the same occupation (and thereby lower expected future family earnings in Period 3). Then p_u arises endogenously as the probability the match quality exceeds the reservation value, increasing with the variance of the non-mO*NETary match-specific utility, and decreasing with the wage gaps. Parameters c and $p_u c$ would then respectively denote the average search costs associated with encountering the set and the subset of workers whom would marry.

Let p_i denote the share of sex i entering clustered occupations, where $i \in \{\text{m, f}\}$. Noting that the expected search duration is the inverse of the probability of marriage in any search period, we may weight p_i by the search durations to compute share of the unmarried pool of sex i in a clustered occupation:

$$\psi_i = \frac{p_i[1 - \psi j(1 - p_u)]^{-1}}{(1-p_i)[1 - (1 - \psi j)(1 - p_u)]^{-1} + p_i[1 - \psi j(1 - p_u)]^{-1}} \tag{1}$$

Lastly, note that the expected marriage search costs are given by the expected search duration multiplied by the marriage search cost c.

C. Period 1: Training Choice

In Period 1, risk-neutral men and women consider expected future wages and marriage search costs, and select their training. Because workers are risk-neutral and the utility of wages and disutility of marriage search costs are additively separable, a male trains in a clustered occupation if

$$\left(\frac{p_u\psi_f}{(1-\psi_f) + p_u\psi_f}\right)W_{cc} + \left(\frac{1-\psi_f}{(1-\psi_f) + p_u\psi_f}\right)W_{cd} - c\left(1 - \psi_f(1 - p_u)\right)^{-1} >$$

$$\left(\frac{\psi_f}{\psi_f + p_u(1-\psi_f)}\right)W_{cd} + \left(\frac{p_u(1-\psi_f)}{\psi_f + p_u(1-\psi_f)}\right)W_{dd} - c\left(p_u + \psi_f(1 - p_u)\right)^{-1} \tag{2}$$

And a female trains in a clustered occupation if

$$\left(\frac{p_u\psi_m}{(1-\psi_m) + p_u\psi_m}\right)W_{cc} + \left(\frac{1-\psi_m}{(1-\psi_m) + p_u\psi_m}\right)W_{cd} - c\left(1 - \psi_m(1 + p_u)\right)^{-1} >$$

$$\left(\frac{\psi_m}{\psi_m + p_u(1-\psi_m)}\right)W_{cd} + \left(\frac{p_u(1-\psi_m)}{\psi_m + p_u(1-\psi_m)}\right)W_{dd} - c\left(p_u + \psi_m(1 - p_u)\right)^{-1} \tag{3}$$

There are three Nash equilibria. First, consider the training decision of the atomistic male or female in the case that all men pursue training in clustered occupations and all women pursue training in dispersed occupations, such that $p_m = 1$ and $p_f = 0$. By equation (1), $p_m = 1$ and $p_f = 0$ implies $\psi_m = 1$ and $\psi_f = 0$. In this scenario, (2) implies it is incentive-compatible for the atomistic male to enter clustered occupations if $W_{cd} - c > W_{dd} - cp_u^{-1}$, and (3) implies it is incentive-compatible for the atomistic female to enter dispersed occupations if $W_{cd} - c \geq W_{cc} - cp_u^{-1}$. These inequalities follow from the key assumptions regarding the wage environment, $p_u^{-1} \geq 1$, and $c > 0$. The second pure strategy Nash equilibrium is the symmetric case where all females pursue training in clustered occupations and all males pursue training in dispersed occupations. The third equilibrium is the unstable mixed strategy Nash equilibrium, whereby men and women sort into clustered occupations in an equal proportion that increases with W_{cc}, declines with W_{dd}, approaches 0.5 with W_{cd} for $p_u \neq 0$, and approaches 0.5 with c for $p_u \neq 1$. The pure strategy equilibria correspond to Hypothesis 1.

Comparative statics show that as sex *j* departs from the pure strategy Nash equilibrium, sex *i*'s rate of substitution of expected costs in marriage search with costs in expected wages depends on p_u. When $p_u = 1$ (*i.e.* workers only marry out of love, regardless of occupational type), then all costs of sex *i* are borne by expected wages at a rate of $\psi_j W_{cc} + (1 - \psi_j)W_{cd}$ if sex *i* segregates into clustered occupations and at a rate $(1 - \psi_j)W_{cd} + \psi_j W_{dd}$ if sex *i* segregates into dispersed occupations. When $p_u = 0$ (*i.e.* workers would not marry out of love someone of the same occupational type), then all costs of sex *i* are borne by marriage search costs at a rate $-c\psi^{-1}$ for workers entering clustered occupations and $-c(1 - \psi)^{-1}$ for workers entering dispersed occupations. For $p_u \in (0, 1)$, workers substitute between the expected wage penalties of the colocation problem and the expected costs of delayed marriage, yielding both wage penalties and marriage search penalties for men and women who depart from the equilibrium. These correspond to Hypotheses 2 and 3.

The results section tests for and identifies the pure-strategy Nash equilibrium. It then tests the predictions corresponding to the Nash equilibrium—that men and women departing from this equilibrium incur wages and marriage search costs. Specifically, these predictions are[12]:

(i) Sex-based occupational clustering: The degree to which a worker's occupation is geographically-clustered varies by sex

(ii) Wage penalties for deviation: The sex segregating into clustered occupations has higher wages in clustered occupations than in dispersed occupations, and the sex segregating into dispersed occupations has higher wages in dispersed occupations than in clustered occupations

(iii) Marriage penalties for deviation: The sex in which the majority sort into clustered occupations marries earlier in clustered occupations than in dispersed occupations, and the sex in which the majority sort into dispersed occupations marries earlier in dispersed occupations than in clustered occupations

In addition, results are expected to be strongest in occupations that require the greatest formal skills training, specifically those where a majority of workers have bachelor's degrees. In the United States, education through high school is highly-general, with specialization largely beginning with college coursework, post-secondary vocational training, and internships. It is this specialization that determines the geographic scope of the job search, the importance of calculated relocations for work, the robustness of the job to relocations for a spouse's career, and

[12] Although this paper does not examine predictions regarding educational investments, existing empirical research has found that career pathways are formed early in one's life course and reflect the segregation in skill investments as predicted by the model. Daymont and Andrisani (1984) find that women are more likely than men to major in education, the humanities, and health or biology, and less likely to major in engineering, science or math, or professional studies, the latter of which will be shown to typify geographically-clustered occupations. Studies by McDonald and Thornton (2007) and Black et al. (2008) find that the majority of the sex pay gap among college-educated workers may be explained by choice of college majors, which is also consistent with segregation by women into majors leading to lower-paying, geographically-dispersed occupations.

the magnitude of the sunk training costs involved in switching occupations in response to a colocation problem. In contrast, workers with less formal skills training may find job-switching more tenable. For example, although gaming cage workers (who account for and exchange chips at casino cages) are highly-clustered and dominated by less-educated women, relocation to a city without a gaming industry may not be very costly since foregone accumulated skills are minimal and potentially transferable to other less-skilled service-sector occupations (such as a cashiers).

D. Potential Extensions

The model makes a number of simplifying assumptions to focus on the coordination problem posed by occupationally-specific educational investments made in advance of knowing the spouse or from where job opportunities will come.

In the model, workers care only for joint family earnings and marriage. In reality, workers also choose their occupations according to their personal tastes. If some proportion of workers would select their occupation out of love of the occupation, the remaining will segregate into clustered and dispersed occupations, although the wage and marriage penalties for departing will fall as the proportion of workers selecting their occupation out of love for the occupation rises.

In the model, all workers receive training, all workers get married to partners of the opposite sex, and all people work. The results section considers exceptions conceptually. I interpret workers with less formal education as workers who should be able to change jobs cheaply, reducing the predicted magnitude of results. I present results for non-marriage. I also present separate results for married non-labor force participants, whose domestic work might be thought of as a “dispersed” occupation. I do not examine same-sex couples in this paper because predictions are theoretically ambiguous, and due to methodological issues.

In the model, workers are sequentially trained, are married, and then observe wage offers in different locations. In reality, foundational career investments, marriage search, and work relocation decisions are often long and overlapping processes rather than discrete sequential decisions.[13] Men and women may delay marriage search until after they have settled down in a long-term job, or they may intentionally search for a spouse within the same occupation or industry so that the geographic distribution of career advancement opportunities, while restricted, will be correlated. As a "Battle of the Sexes" game, the model requires some cost for avoiding the colocation problem after investments into a clustered or dispersed occupation have been made. As such, the model's predictions are robust to foregoing marriage until long-term location is known, restricting the marriage search to those whose location or occupation is known to be compatible, divorce due to differences in geographic preferences, or switching occupations after the spouse is known, as long as these alternatives are costly. Predictions are not robust to costlessly delaying the marriage search, costlessly restricting the marriage search, or costlessly switching among occupations or spouses.

Exceptions to these simplifying assumptions present a wide range of opportunities for testing other specific predictions of the model. The objective of this paper is to introduce a general life course framework for analyzing the colocation problem, and test its basic predictions for segregation, earnings, and marriage by sex and education.

II. Measuring Occupational Clustering

The model's predictions are driven by anti-complementarities when both spouses in a dual-earner couple work in occupations that reward self-serving relocation decisions for career advancement, and the benefits of having one spouse in an occupation that is geographically-

[13] However, the general timing of the model is consistent with Gautier, Svarer, and Teulings (2010), who interpret the tendency of young, single Danish workers to migrate to large cities, marry, and then move outside the city as evidence initial early-career location decisions are partly governed by the desire to reduce marriage search costs.

flexible and robust to work relocations for the spouse. Operationalizing the geographic constraint and flexibility of occupations is not straightforward, and directly examining work relocations by occupation is problematic for at least three reasons. First, job search behaviors and opportunities are likely to be endogenous by age, sex, and marital status. Second, the penalties for tied-spouses are difficult to estimate using data on relocations, particularly among "tied-stayers." Third, data sources with work relocation information such as the CPS have insufficient statistical power to estimate the likelihood of relocation for work for individual occupations. The key advantages of focusing instead on occupational clustering is that it avoids these dilemmas, as the geographic distribution of occupations is likely to be (relatively) exogenous to their sexual composition and demographics, occupational clustering captures the geographic constraint while ignoring tied moves and stays, and occupational clustering may be estimated with the (much more powerful) Decennial Census files. NO*NETheless, an occupation's geographic distribution is an imperfect proxy for what is desired: the degree occupations truly reward calculated relocation decisions or the degree to which they are truly robust to spousal relocations. Geographic dispersion may overstate the geographic flexibility of jobs that require state licensure or an accumulated local clientele. Such jobs will introduce measurement error.

To measure occupational clustering, I construct an occupation-level data set from labor force participants residing in metropolitan areas in the 5% public-use microdata sample of the 2000 U.S. Decennial Census. The Census features the large number of observations (7.6 million metropolitan workers) across the 283 metropolitan statistical areas and the 474 six-digit SOC occupations necessary to analyze occupational clustering. The Bureau of Labor Statistics distinguishes six-digit SOC classifications as those with similar job duties and skills, which is arguably consistent with occupational skill-specificity that makes it costly to switch jobs after marriage. I control for two occupational characteristics quantified by the O*NET Database[14]: the degree to which aptitude for "explosive physical strength," "assisting and caring for others," and the use of "mathematics to solve problems" affects job performance. Measures for each vary from 0 to 100. These represent the typical human capital and taste-based explanations for occupational segregation (see Anker 1997 for a review). I then use a Duncan-type index for polytomous dissimilar variables (see Duncan and Duncan 1955; Mele 2007). Duncan's D is commonly used to study occupational stratification by summing the absolute differences between the proportions of two groups (such as men and women). Because the variable of interest is polytomous rather than dichotomous dissimilarity (metropolitan areas rather than sex), the generalized version allows analysis of clustering. This "clustering index" of occupation *j* takes the form

$$\mathrm{C_j} = \frac{1}{2}\sum_{i=1}^{I}\left|\frac{n_{ij}}{n_j} - \frac{n_i - n_{ij}}{n - n_j}\right|$$

where *I* represents the set of metropolitan areas, *n* represents the counts of workers aged 18-65 in the labor force, and subscripts denote counts within metropolitan areas *i* and occupations *j*. For illustration, if occupation *j* represents exactly 1% of the workforce in every metropolitan area, the absolute sum of the differences (and the clustering index for that occupation) is zero, regardless of the sex composition across metropolitan areas or other occupations. As workers are concentrated into fewer metropolitan areas, the clustering index converges to one. Like the original Duncan's D, the generalized "clustering index" also has an intuitive interpretation: it is the proportion of workers within an occupation that would need to relocate for the share of the

[14] O*NET (short for the Occupational Information Network) is sponsored by the Department of Labor and the Employment and Training Administration, and has served as the standard reference for job definitions since the Dictionary of Occupational Titles was phased out in 1991. It provides continuously-updated and standardized measures of occupational characteristics and is constructed from surveys of job incumbents and occupation analysts.

occupation to be uniform across metropolitan areas. For example, in 2000 and within metropolitan areas, there were seventeen registered nurses per thousand labor force participants in the United States, and the (relatively low) clustering score of 0.084 for registered nurses implies 8.4% of registered nurses would need to relocate for there to be seventeen registered nurses per thousand workers in every metropolitan area in the United States. Following this interpretation, it is also intuitive to see that the index places greater weight on clustering in more populous cities. Functionally, occupational clustering in more-populous cities affects the national mean (the right hand term) more-so than clustering in small cities.

[TABLE 1]

Table 1 lists the three non-military occupations with the highest and lowest clustering indices by education and sex. Educational categories are assigned from the most-likely educational category for a randomly-sampled worker in that occupation (*i.e.* the modal educational attainment). Highly-clustered occupations that employ highly-educated men tend to be in specialized sciences and engineering, while those for highly-educated women tend to be diverse and smaller occupations, such as museum curators and archivists. Highly-educated and dispersed occupations that are dominated by men include physicians and managers, while those dominated by women are larger occupations, many of which are in teaching and education.

Among occupations employing less-educated workers, men dominate technical occupations, including both dispersed occupations (such as computer and auto repair) and clustered occupations (assemblers). Highly-dispersed, some-college, female-dominated occupations are some of the largest in terms of total employment, including registered nurses, secretaries and administrative aides, auditing clerks, customer service representatives, and nursing home aides.

Interestingly, clustered occupations dominated by men with less formal skills training (for example, machinists, technicians, and extraction workers) appear to involve a great degree of highly-specific vocational skills. Workers in these occupations are likely to involve foregone wages for exogenous relocations, but may not be recognized by measures of formal education. In contrast, less-educated but highly-clustered occupations dominated by females appear to involve highly-general and transferable skills, and are likely to be less-penalized by relocation. I find that less-educated men earn a premium in clustered occupations, but not women (see the results section).

This index is then normalized by log-transformation. Among labor force participants aged 18-65, the log-transformed clustering index has a mean of -2.17, a standard deviation of 0.55, a minimum of -3.26, and a maximum of -0.23 across all occupations.

[TABLE 2]

Table 2 summarizes the clustering index by sex across educational attainment and marital status. To provide an intuition, the mean clustering index of the occupations held by men is 0.247 log-points (28%) greater than that for women. The differences are larger across columns than rows, implying sex is a more powerful predictor of the degree to which a worker's occupation is clustered than other demographic characteristics. Men and women in clustered occupations are generally: more educated, are more likely to have a separated or absent spouse, and work full-time. Clustered occupations differ little in whether workers have children, although the interpretation is complicated by selection bias caused by the lower labor force participation for women with children.

Before checking the model's predictions for occupational clustering and related wage and marriage penalties, I examine the model's implication that occupational clustering prompts workers to relocate for work, since wages are location-invariant only in dispersed occupations. I do so by estimating the probability workers and couples "relocate for new job or job transfer" using the March CPS Supplements (2003-2010) that use the 2000 Census SOC codes. Households attribute about 10% of all relocations primarily for work-related reasons, with other major reasons including family, change in marital status, and upgrading housing. I restrict the sample to prime age workers (25 to 40) because most workers are done with schooling, are highly at-risk of both relocation and marriage, and likely to be selecting job offers to accumulate skills. The clustering variable is appended to the March CPS data from calculations performed using the 2000 Census. Analysis is conducted for never-married workers, one-earner couples, and two-earner couples. For two-earner couples, propensity to relocate is examined using the maximum of the heads' log-clustering indices. The effect of clustering is calculated separately when this corresponds to the husband's or wife's occupation. Specifically, logistic regression's estimated log-odds a family relocates for work is given by

$$\ln\left(\frac{\widehat{p_i}}{1-\widehat{p_i}}\right) = \hat{\beta}_0 + \hat{\beta}_1 \ln C_{i,H}\, X_i + \hat{\beta}_2 \ln C_{i,W}\,(1 - X_i)$$

where $\widehat{p_i}$ is the logit function's estimated probability that household *i* relocates for work, $C_{i,H}$ is the husband's clustering index, and X_i is an indicator for $C_{i,H} \geq C_{i,W}$. Results are shown in Table 3.

[TABLE 3]

Consistent with the model, never-married prime-age men and women are both more likely to relocate for work when they work in clustered occupations. This lends support for the model's implication that geographically-clustered occupations reward and induce relocation for work among those who are most-able to relocate.

Among households where the husband works, families in which the husband's occupation is more-clustered are more-likely to relocate for work. The CPS includes very few married couples where only the wife works that move for work-related reasons, making similar estimates for women unreliable. Regressions 7 and 8 show that the propensity to relocate for work among two-earner couples are about equally-sensitive to whether the husband or wife has the more-clustered occupation. Children have a negative effect on propensity to relocate in all regressions. Restricting the sample to bachelor's degree-holding workers and households (not shown) yields estimated effects of occupational clustering on relocation for work that are, for all models, higher than those from the full sample.

III. Results for Segregation, Wages, and Marriage

First, I test for sex-based occupational clustering (Hypothesis 1). Upon establishing the clustering of men into clustered occupations, I test the corresponding predictions regarding the wage penalties (Hypothesis 2) and marital penalties (Hypothesis 3) predicted to affect the minority who depart from this equilibrium.

A. Sex-Based Occupational Clustering in the Family and Labor Market

The model predicts men or women will segregate differentially into clustered and dispersed occupations, rather than occupational clustering being unconditioned on sex. These results will be used to identify the corresponding equilibrium predictions around marriage and wages.

Table 4 compares labor force participation and occupational clustering within families and breaks down tabulations by the mean age of the heads. Consistent with the model, men in both age groups are significantly and substantially more-likely to work in the more-clustered occupation.

[TABLE 4]

The table also provides intuitive measures of the magnitude of sex-based occupational clustering. In 2000, among both younger and older non-power couples and power-couples (those with two bachelor's degree holders), couples in which the husband's occupation had a clustering score greater than the wife's occupation outnumbered the reverse nearly two-to-one. Male-breadwinner families outnumber female-breadwinner families more than six-to-one. Household labor may be thought to be a "very-dispersed" occupation, and highly-amenable to relocation.[15]

Table 5 examines sex-based occupational clustering in the economy as a whole by presenting logistic regression results estimating the likelihood a worker is female given the log-clustering index of his or her occupation using the Decennial Census. First, I run bivariate regressions for all 474 occupations. Next, I perform additional analyses to isolate the effect of geographic clustering from alternative neoclassical explanations for occupational segregation. To address potential alternative explanations for occupational segregation, I control for the occupation's usual hours, its mean age, its requirements of physical strength, its requirements of assisting and caring for others, and its math requirements. Next, I restrict the sample to the subset of "highly-educated" occupations that further excludes 366 of those where the majority of workers have not obtained a bachelor's degree (or greater).[16] This subset focuses on occupations for which early-career training investments are likely to be greatest and most occupationally-

[15] However, household labor may also be complementary to external labor market labor for other reasons other than geographic constraint and flexibility. Either complementarities in domestic and external production, or Ricardian specialization in domestic and external production along with an inability to efficiently outsource domestic production, are typically essential to household models (see Becker 1991 for a discussion).

[16] The only occupation omitted for being "physical" that would otherwise appear among the "highly-educated" occupations is "55-1010: Military Officer Special and Tactical Operations Leaders/Managers." This occupation is male-dominated, highly-clustered, and omitted anyway as a military occupation. Incidentally, military officers have attracted special research attention because they are perceived to be very challenging for dual earner couples. For example, Gill and Haurin (2002) examine how the decision to pursue military officer training is affected by their wives' earnings and labor force attachment.

specific, raising the cost of mid-career occupation switching and generating the theory's specific prediction that segregation will be most punctuated among high-skill occupations. In contrast, sex-based occupational clustering should break down if switching were costless and did not require leaving productive skills fallow, which may be the case among low-skill occupations. For example, switching from a gaming cage worker (clustered) to a cashier (dispersed) due to a spouse's relocation may involve less foregone income than switching from one highly-technical specialty to one employing broader skills at the destination city.

[TABLE 5]

Consistent with the theory, Table 5 finds a pattern of sex-based occupational clustering whereby men tend to enter more geographically-clustered occupations than do women. Controls for the occupation's mean age and usual hours worked are shown to be weak predictors of occupational clustering. Not surprisingly, among all occupations, women segregate into jobs requiring less physical strength and greater care work. Among highly-educated occupations, the relationships between these occupational characteristics and the sex composition are weaker. Including further controls reduces the magnitude but retains the significance of occupational clustering in the full regressions. As predicted, occupational clustering is much more-pronounced in occupations in which the majority of workers have bachelor's degrees. Interestingly, O*NET controls are weakly correlated with sex; caring and physicality are significant but weak predictors among all occupations, and mathematical skill is a significant but weak predictor among highly-educated occupations. Controlling for occupational characteristics calculated from the Census and O*NET has a small and statistically insignificant effect on the occupational clustering coefficient.

To provide an intuition for the magnitude, Regression 2 estimates that a worker in an occupation with a log-clustering score one standard deviation above the mean (more-clustered) has a 36% probability of being female, and one with a log-clustering score one standard deviation (more-dispersed) below the mean has a 55% probability of being female. For workers in high-skill occupations, Regression 6 estimates these to be 34% and 78%, respectively.[17]

[FIGURE 1]

Figure 1 provides a scatterplot of the clustering index against female composition for both the full set of non-military occupations and then for the subset of highly-educated occupations where the majority of workers have bachelor's degrees or greater. In both samples, there is a downward and statistically significant slope indicating that highly clustered occupations tend to have fewer women, particularly among highly-educated occupations. It is also clear that larger occupations tend to be less clustered and more female-dominated, while smaller occupations tend to be more clustered and more male-dominated. Generally, the largest occupations—secretaries, retail salespeople, school teachers, cashiers, customer service representatives, and nurses—all perform tasks that must be done in-person and serve a population's basic needs. On the other hand, the smaller occupations often perform specialized tasks specific to industries that benefit from clustering, and tend not to produce non-transportable in-person services, with very few exceptions ("septic tank servicers and sewer pipe cleaners" and "motion picture projectionists" are among the smallest twenty occupations and are relatively dispersed). It also may be the case that larger occupations are not only dispersed, but also offer well-functioning and competitive ("thick") labor markets in any given city, making them attractive to potential tied-movers.

[17] Estimates are calculated at the mean values for average age and usual hours.

Graphical analysis and Cooks D show that influential observations yielding the significant downward OLS regression line include teaching, health, and administrative support occupations (highly female, highly dispersed) and engineering occupations (highly male, highly clustered). Influential observations working against the downward trend include sewing machine operators and gaming cage workers (highly female, highly clustered) and truck drivers (highly male, highly dispersed). More broadly, exceptions tend to be among occupations requiring less formal education, and thereby potentially less of a sunk cost for switching less-skilled occupations. Indeed, the second panel of Figure 1 illustrates Table 5's finding that occupational clustering is more-pronounced in the more highly-educated occupations, with few exceptions. The most influential residuals among occupations where the majority of workers have bachelor's degrees are "Chief Executives" and "Clergy," both of which are large, male-dominated, highly-dispersed occupations, although these are problematic for other reasons.

Both family and economy-wide results are consistent with sex-based occupational clustering in which men segregate into clustered occupations. Following the model, this then yields predictions regarding how wage premia of clustered occupations vary by sex.

B. Wage-Age Profiles by Sex, Education, and Clustering

Having established the segregation of men into geographically-clustered occupations, this section tests the model's corresponding predictions regarding the wage premia in clustered occupations by sex. Formally, in perfect segregation men always enjoy the compensating differential $w_H > w_D$ in clustered occupations, while women are deterred from clustered occupations by the risk post-marital geographic displacement (w_L). In contrast, the model assumes that dispersed occupations are robust to spousal relocations, and fixed at w_D. The model thereby predicts higher wages in clustered occupations for men and higher wages in dispersed occupations for women. The model predicts these properties keep equilibrium segregation stable.

I also examine related properties of the model. First, the sex-pay gap should be driven by clustered occupations, not by dispersed occupations. This is because women are more-likely to be penalized by geographic displacement in clustered occupations, while wages in dispersed occupations are fixed for both sexes (and all geographies) w_D.[18]

To analyze wage premia in clustered occupations at different ages by sex and education, I separate workers in the 2000 Decennial Census into terciles by occupational clustering score; workers in occupations with occupational clustering scores in the sixty-seventh percentile and above (greater than 0.164) are denoted as workers in "clustered occupations," and workers with scores in the thirty-third percentile and below (less than 0.092) are denoted as workers in "dispersed occupations." Table 6 reports estimated median wages in clustered and dispersed occupations by education, sex, and age. Estimates are made using quantile regression to predict real median wages as a function of age (linear and quadratic terms), estimated separately for each sex and at four education levels.

[TABLE 6]

Consistent with the model, median wages among men in clustered occupations are greater than median wages among men in dispersed occupations throughout the working life and at every educational category. Consistent with the model's prediction regarding the relative wage gap, the premium for working in clustered occupations among women is relatively small in the

[18] Selection bias is likely to complicate straightforward testing since labor force exit and occupational switching are not directly observed and would be expected to mitigate the earnings penalties of the colocation problem. For estimation purposes, this is particularly troublesome since workers may be prompted to exit the labor force or switch occupations due to a spouse's higher earnings (a labor supply-reducing income effect) or demanding career, both of which would be expected to be characteristics of clustered occupations. This bias may attenuate results.

early-career and negative in the late-career. For both men and women, I expect the premia for clustered occupations to overestimate the premium for the family because it neglects the effect on the spouse's earnings. The overestimate is believed to be particularly great for women, since husbands tend to be in more-clustered occupations than their wives.

Table 6 also confirms that the sex wage gap is more pronounced in clustered occupations, suggesting wage penalties may have a deterrent effect on women. Within highly-educated occupations, the male wage premium is between -9% and 12% in dispersed occupations and between 5% and 40% in clustered occupations, depending on age and education. While the wage gap is similar throughout the life course in dispersed occupations, the gap in clustered occupations tends to grow throughout the life course. Similarly, among less-educated occupations, the wage gap grows more rapidly in clustered occupations.

[FIGURE 2]

Figure 2 illustrates the striking magnitude of the wage gap for highly-educated women in clustered occupations at later ages, which is absent for women in dispersed occupations. At early ages, women in clustered occupations enjoy greater median wages than those in dispersed occupations. However, median wages for highly-educated women in clustered occupations rapidly decelerate at later ages. For women with graduate degrees, median wages peak around age forty, and fall below highly-educated women in dispersed occupations in the mid-forties, while wages for highly-educated women in dispersed occupations continue to rise and (in contrast) closely track wages for highly-educated men in dispersed occupations. As a result, women in clustered occupations earn much lower wages than observably comparable men, while women in dispersed occupations have wages that are similar for comparable men.

One interpretation for the rapid deceleration is that highly-educated women in clustered occupations are more likely than men to be (or to have been) tied-stayers or tied-movers, and these career impediments manifest lower median wages, particularly when compared to men of equal age and education in clustered occupations. In contrast, wages for highly-educated women in dispersed occupations are very similar to those of men of equal age; the wage disparity is stark among highly-educated women in clustered occupations, and this disparity grows along the life course. This interpretation is consistent with Bertrand, Goldin, and Katz's (2009) finding that much of the sex pay gap among MBAs may be explained by the likelihood of job discontinuity (particularly from childrearing) within ten years after completing education. The distinction between highly-clustered and highly-dispersed occupations suggests that the fall from career ladders may be more common or pronounced among women who enter clustered occupations such as banking or consulting rather than dispersed occupations such as medicine or education; like childrearing, the tied-mover/tied-stayer problem affects women more than men.

Among women with some college and associate's degrees, those with clustered occupations do tend to earn more than those in dispersed occupations, but differences are very small. One interpretation is that retraining (ie. job-switching) costs for these occupations are low, and therefore it is feasible to retrain cheaply if a spouse relocates. Indeed, many of the female-dominated less-educated occupations that are highly-clustered (eg. textile occupations and gaming workers) appear to involve less-accumulation of on-the-job and tacit skills as those dominated by men (largely in the technical trades, manufacturing, and extraction).

Analysis of wage profiles by clustering, age, sex, and education supports the model's prediction that clustered occupations are associated with wage premiums for educated members of the sex generally entering clustered occupations (men), and high-but-decelerating premiums for educated members of the sex generally entering dispersed occupations (women).

C. Occupational Clustering and Marital Search

Because men are more-likely to enter geographically-clustered occupations, the model predicts that women who depart from occupational selection patterns by entering geographically-clustered occupations will, on average, experience later first-marriage. Following the conceptual framework, this is due to the relatively-slow arrival rate of marriage candidates with compatible (dispersed) occupations. Likewise, men who depart from occupational selection patterns by entering geographically-dispersed occupations will also experience a longer average marital search, as the arrival rate of marriage candidates with clustered occupations is more slow. Formally, in the case of pure segregation, men and women always marry in the first period at cost c, and are deterred from the other occupation by the expected cost cp_u^{-1}, where $p_u \in [0, 1]$ is the probability of marrying a random spousal candidate out of love and $p_u^{-1} \geq 1$.

[FIGURE 3]

Figure 3 examines this hypothesis by reporting the probability of being ever-married by age, education, and whether the individual's occupation is above or below median clustering. Consistent with the hypothesis, women who enter clustered occupations experience later first-marriage. The quartile ages for first-marriage among highly-educated women are 24, 28, and 34 in clustered occupations and 23, 26, and 30 in dispersed occupations. By age thirty, 63% of highly-educated women in clustered occupations are ever-married, versus 74% of those in dispersed occupations. The effect among less-educated women is smaller, with an age-at-first-marriage quartiles at 21, 24, and 30 in clustered occupations and 21, 24, and 29 in dispersed occupations. Evidence is weaker for men. Quartiles among highly-educated men are the same in clustered and dispersed occupations: 24, 28, and 34. Among less-educated men, working in a clustered occupation has an opposite effect as women; quartiles in clustered occupations are 23, 25, and 32, while those in dispersed occupations are 23, 26, and 34.

One challenge of the Census data is that it does not distinguish cohabitating partners from married partners. This may bias results if women who enter clustered occupations are also more-likely to prefer cohabitation to marriage.

[FIGURE 4]

As noted by Mincer (1978), the colocation problem may cause marital tension. If the colocation problem is particularly-likely in families where women enter geographically-clustered occupations as is argued here, then Mincer's logic may be extended to use occupational clustering to predict divorce. Figure 4 shows that ever-married women in clustered occupations are more-likely to be currently divorced and not remarried.[19]

Consistent with the theory, women who enter clustered occupations are more likely to delay or forego marriage than women who enter dispersed occupations, particularly among the college-educated. While theory predicts the reverse for men (who, when entering dispersed occupations, delay marriage to find a female with a clustered occupation), evidence for this is weak among non-college-educated men, and there is no evidence of delayed marriage among college-educated men entering dispersed occupations. One interpretation is that college-educated men in dispersed occupations have distaste for delaying marriage to seek out a wife with a clustered occupation who would be the higher-earning "leading spouse."

[19] Differences were also tested in a regression framework as follows. Eight regressions were run—one for each combination of education, sex, and clustering. The sample frame includes non-never-married individuals (*i.e.* those at-risk of being divorced). The dependent variable is a "divorced" marital status, and the independent variables are clustering, age, age-squared. All coefficients in all four regressions are statistically significant with $p<0.01$. Evaluated at age forty, women in clustered occupations had a higher probability of being divorced and men had a lower probability of being divorced, each with a $p<0.01$. The estimated difference for less-educated men, however, is very small (2%, standard error of 0.4%).

IV. Discussion and Future Work

This paper has three principal goals: to broaden the theoretical scope of the colocation problem to include its effects on men's and women's occupations, marriage timing, and wages; to calculate and to examine the economic significance of an index that captures the geographic ubiquity of an occupation; and to test a basic set of predictions regarding equilibrium segregation and the penalties predicted to keep it stable. A broader implication of the theory is that men's and women's expectations of future mobility may help perpetuate occupational segregation, the sex-pay gap, disproportionate marriage market penalties for women entering geographically-clustered jobs. A further implication is that these features of inequality are perpetuated by their aggregately self-fulfilling expectations of future mobility.

Results confirm the segregation of one sex (men) into geographically-clustered occupations.[20] Results suggest career concerns penalize men who enter dispersed occupations, and both wage and marriage concerns penalize women who enter clustered occupations. However, results suggest less-educated men suffer later marriage when entering dispersed occupations and more-educated women suffer lower mid- and late-career median wages in clustered occupations.

As an explanation of occupational segregation, the principal concern is that geographically-clustered occupations possess other qualities affecting segregation. Indeed, it is clear that dispersed and clustered occupations differ in several ways. Clustered occupations tend to be specific to clustered industries, which may be subject to the natural and technological

[20] It remains unclear why all societies appear to exist on the equilibrium where women segregate into dispersed occupations. One possibility is that men's rational expectations to be able to relocate for work, and women's expectation to be exogenously relocated, is a legacy of women's historic emphasis on domestic work. Alternatively, childrearing may disproportionately interrupt women's careers, and workers sort to avoid interrupting the dominant career.

agglomerative forces described by Ellison and Glaeser (1997), while dispersed occupations either tend to be specific to dispersed industries (e.g. education or healthcare) or be employed in supportive capacities across potentially-clustered industries (e.g. general managers or administrative assistants).

Controlling for occupational characteristics is inherently difficult, and this study tries to overcome these obstacles in two ways. First, this study uses O*NET and the Census to control for job characteristics offered by the literatures in neoclassical economics and sociology—these include average working hours, wage, physicality, or requirement of assisting and caring for others. I do not find evidence that these measures reduce the explanatory power of geographic clustering on sex. Second, certain results presented here are uniquely consistent with sex-based occupational clustering. Occupational clustering by sex becomes a more pronounced among occupations with formal education (*i.e.* those for which the cost of switching occupations after spouse and location is revealed is high). Deviating from clustered and dispersed occupations has a differential effect on wages and marriage outcomes by sex.

Notwithstanding these results, this study does not attempt to estimate the disamenity cost of geographic constraint by sex, nor offer an explanation why modern economies are universally consistent with a “men segregate into clustered occupations”-equilibrium, even though language and national boundaries tend to segregate marriage markets. One potential explanation is that this segregation is a legacy of a time when men enjoyed the ability to work anywhere, and as women entered the labor market, they continued to enter jobs still amenable to household labor and geographic flexibility. Unfortunately, such predictions will be difficult to test. One possibility is that sex-based occupational clustering may be examined in countries with limited international mobility, and where educated workers tend to live in one city.

In addition to explaining occupational segregation, wages, and marriage outcomes, the model may also help contextualize other labor market phenomena. First, because workers in dispersed occupations implicitly pay for the amenity of geographic flexibility, the model presents an explanation for why women segregate into lower-paying occupations and college majors with lower earnings (Blau and Kahn 2000, McDonald and Thornton 2007, Black et al. 2008). Second, the model provides an explanation for occupation switching and equilibrium skills mismatch, and may explain why women are more-likely to possess fallow skills (Johansson and Katz 2006, Ofek and Merrill 1997; see also McGoldrick and Robst 1996 for a critical view). Third, the model may explain why the expansion of women with bachelor's and graduate degrees are being absorbed by the most highly-dispersed occupations, including physicians, lawyers, dentists, and general managers. Fourth, the model helps explain "statistical discrimination" in training by proposing why women have higher turnover rates due to exogenous spousal relocation (for studies explaining discrimination in training by differences in turnover rates by sex, see Duncan and Hoffman 1979, Gronau 1988, Lillard and Tan 1986, and Royalty 1996). Fifth, the model explains why women are persistently more-likely than men to be tied-movers (Mincer 1978, Nivalainen 2004). In short, a wide variety of labor market phenomena might be better-understood by adopting a life-cycle approach and treating occupational choice as endogenous to geographic constraints.

Works Cited

Anker, Richard. 1997. “Theories of Occupational Segregation by Sex: An Overview.” *International Labour Review* 136(3): 315-339.

Becker, Gary. 1965. “A Theory of the Allocation of Time.” *Economic Journal,* 75(299): 493-517.

_____. 1971. *The Economics of Discrimination*. Chicago: University of Chicago Press.

_____. 1985. “Human Capital, Effort, and the Sexual Division of Labor.” *Journal of Labor Economics*, 3(1): S33-S58.

_____. 1991. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bertrand, Marianne, C. Goldin, and L. F Katz. 2009. Dynamics of the gender gap for young professionals in the corporate and financial sectors. Working Paper No. 14681, National Bureau of Economic Research, Cambridge, MA.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2008. “Gender Wage Disparities Among the Highly Educated.” *Journal of Human Resources*, 43(3): 630-59.

Boyle, Paul et al. 2001. “A Cross-National Comparison of the Impact of Family Migration on Women's Employment Status.” *Demography*, 38(2): 201-213.

Buchmann, Claudia, and Thomas A. DiPrete. 2006. “The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement.” *American Sociological Review*, 71(4): 515-41.

Clark, William A. V., and Youqin Huang. 2006. “Balancing move and work: women's labour market exits and entries after family migration.” *Population, Space and Place*, 12(1): 31-44.

Compton, Janice, and Robert A. Pollak. 2007. "Why Are Power Couples Increasingly Concentrated in Large Metropolitan Areas?" *Journal of Labor Economics* 25(3): 475-512.

Costa, Dora L., and Matthew E. Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990." *Quarterly Journal of Economics*, 115(4): 1287-1315.

Daymont, Thomas N., and Paul J. Andrisani. 1984. "Job Preferences, College Major, and the Gender Gap in Earnings." *The Journal of Human Resources,* 19(3): 408-428.

Duncan, Greg J., and Saul Hoffman. 1979. "On-The-Job Training and Earnings Differences by Race and Sex." *The Review of Economics and Statistics*, 61(4): 594-603.

Duncan, Otis Dudley, and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review,* 20(2): 210-217.

Echevarria, Cristina, and Antonio Merlo. 1999. "Gender Differences in Education in a Dynamic Household Bargaining Model." *International Economic Review* 40(2): 265-86.

Ellison, Glenn, and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy,* 105(5): 889-927.

Engineer, Merwan, and Linda Welling. 1998. "Human Capital, True Love, and Gender Roles: Is Sex Destiny?" *Journal of Economic Behavior and Organization*, 40(2): 155-78.

Frank, Robert. 1978. "Family Location Constraints and the Geographic Distribution of Female Professionals." *Journal of Political Economy*, 86(1): 117-30.

Gautier, Pieter, Michael Svarer, and Coen Tuelings. 2010. "Marriage and the City: Search Frictions and Sorting of Singles." *Journal of Urban Economics*, 67(2): 206-18.

Gill, Leroy, and Donald Haurin. 2002. "Wherever He May Go: How Wives Affect their Husband's Career Decisions." *Social Science Research*, 27(3): 264-79.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives,* 20(4): 133-4A.

Gronau, Reuben. 1988. "Sex-Related Wage Differentials and Women's Interrupted Labor Careers-the Chicken or the Egg." *Journal of Labor Economics*, 6(3): 277-301.

Hadfield, Gillian. 1999. "A Coordination Model of the Sexual Division of Labor." *Journal of Economic Behavior and Organization* 40(2): 125-53.

Jacobsen, Joyce P., and Laurence M. Levin. 2000. "The effects of internal migration on the relative economic status of women and men." *Journal of Socio-Economics*, 29(3): 291-304.

Johansson, M., and K. Katz. 2007. "Wage differences between women and men in Sweden: the impact of skill mismatch." An Institute for Labour Market Policy Evaluation working paper, No. 2007: 13.

King, Miriam, et al. 2010. Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. Machine-Readable Database. Minneapolis: University of Minnesota.

Lillard, L. A, and H. W Tan. 1986. "Private Sector Training. Who Gets It and What Are Its Effects?" Rand Corporation.

Long, Larry H. 1974. "Women's Labor Force Participation and the Residential Mobility of Families." *Social Forces* 52: 342-48.

McGoldrick, KimMarie, and John Robst. 1996. "Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification." *The American Economic Review*, 86(2): 280-284.

McDonald, Judith A., and Robert J. Thornton. 2008. "Do New Male and Female College Graduates Receive Equal Pay?" *Journal of Human Resources*, 42(1): 32-48.

McKinnish, Terra. 2008. “Spousal Mobility and Earnings.” *Demography*. 45(4): 829-849.

Mele, Angelo. 2007. “How to Measure Segregation.” Working paper available online at http://www.eea-esem.com/files/papers/EEA-ESEM/2007/299/ppsjul07v14.pdf.

Mincer, Jacob. 1978. “Family Migration Decisions.” *The Journal of Political Economy*, 86(5): 749-773.

Nivalainen, Satu. 2004. “Determinants of Family Migration: Short Moves vs. Long Moves.” *Journal of Population Economics*, 17(1): 157-75.

Ofek, Haim, and Yesook Merrill. 1997. “Labor Immobility and the Formation of Gender Wage Gaps in Local Labor Markets.” *Economic Inquiry*, 35(1): 28-47.

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

Royalty, Anne Beeson. 1996. “Effects of Job Turnover on the Training of Men and Women.” *The Industrial and Labor Relation Review*, 49(3): 506.

Ruggles, Teven, J. et al. 2010. Integrated Public Use Microdata Series: Version 5.0. Machine-readable database. Minneapolis: University of Minnesota.

TABLE 1–Most and Least-Clustered Occupations by Education and Sex

Modal Education	Type	Occupations with Male Majority	Occupations with Female Majority
Grad. Degree (n=30)	Most-Clustered	Economists (0.46)	Oth. Health Pract. (0.41)
		Astron. & Phys. (0.45)	Audiologists (0.27)
		Legislators (0.38)	Archivists & Curators (0.27)
	Most-Dispersed	Physicians (0.12)	Educ. admins (0.08)
		Dentists (0.13)	Counselors (0.10)
		Clergy (0.14)	Spec. Ed. Teachers (0.15)
	No. Occs	*19*	*11*
College Degree (n=85)	Most-Clustered	Mine/Petrol Engineers (0.69)	Bio. Technicians (0.36)
		Marine Engineers (0.63)	Rec. Therapists (0.34)
		Nucl. Engineers (0.57)	Budget Analysts (0.3)
	Most-Dispersed	Wholesale Sales Reps (0.09)	Elementary Teachers (0.06)
		Other Managers (0.1)	Secondary Teachers (0.08)
		Pharmacists (0.12)	HR Managers (0.09)
	No. Occs	*52*	*33*
Some College/ Asso. (n=146)	Most-Clustered	Avionics Techs (0.54)	Occ. Therapist Aides (0.42)
		Geo/Petrol Techs (0.49)	Brokerage Clerks (0.42)
		Tool & Die Makers (0.47)	New Acct. Clerks (0.34)
	Most-Dispersed	Retail Sales Supers (0.05)	Bookkeeping Clerks (0.05)
		Gen. Ops Managers (0.07)	Admin Support Supers (0.05)
		Non-Ret. Sales Supers (0.08)	Secretaries (0.06)
	No. Occs	*69*	*77*
HS or Less (n=213)	Most-Clustered	Tire Builders (0.7)	Textile Wind. Setters (0.79)
		Shoe Machinists (0.71)	Gaming Cage Work. (0.67)
		Oth. Extraction Work. (0.65)	Textile Knit. Setters (0.64)
	Most-Dispersed	Food Serv. Managers (0.06)	Retail Salespersons (0.04)
		Janitors (0.07)	Receptionists (0.05)
		Auto Service Techs (0.08)	Hairdressers (0.06)
	No. Occs	*171*	*42*

Notes: Data Source: 2000 Decennial Census 5% PUMS. Clustering indices in parentheses. Four military occupations are excluded, each of which are highly-clustered and majority-male.

TABLE 2–Differences in Log-Clustering Indices to Unconditional Mean Among Females, by Sex, Education, Marital Status, Full-Time Status, and Children

	Males		Females	
	Mean	SD	Mean	SD
Total	0.247	(0.563)	≡ 0	(0.523)
By Education				
High School Equiv. or Less	0.250	(0.568)	0.026	(0.556)
Associate's/ Some Coll.	0.194	(0.557)	-0.052	(0.489)
College Degree	0.220	(0.537)	0.002	(0.484)
Graduate Degree	0.301	(0.488)	0.088	(0.488)
By Marital Status				
Married, Spouse Present	0.246	(0.548)	-0.012	(0.514)
Married, Spouse Absent	0.275	(0.583)	0.039	(0.544)
Separated	0.242	(0.547)	0.056	(0.522)
Divorced	0.235	(0.548)	0.029	(0.513)
Widowed	0.237	(0.553)	0.029	(0.531)
Never married	0.201	(0.563)	-0.002	(0.523)
By Full-Time Status				
Works ≥ 35 hours per week	0.246	(0.550)	0.023	(0.512)
Works < 35 hours per week	0.211	(0.561)	-0.020	(0.522)
By Children				
Has Own-Children in Household	0.248	(0.547)	0.005	(0.513)
Does Not	0.223	(0.558)	-0.005	(0.522)
Observations	3,954,006		3,691,964	

Notes: Data Source: 2000 Decennial Census 5% PUMS. Log-clustering indices are normalized to the unconditional mean among females (-2.286) by subtraction. Standard errors by sex are both less than 0.0003. Standard errors by subcategory are all less than 0.002. Includes labor force participants aged 18-65.

TABLE 3–Logistic Regression Estimating Probability of Relocating for Work, All Education Levels

	Never-Married Workers		Married Couples with Spouses Present					
			One Spouse ILF				Both ILF	
	Men	Women	Husb-Only ILF		Wife-Only ILF			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log-Clustering	0.207**	0.295**	0.309**	0.298**	0.113	0.127		
	(0.054)	(0.065)	(0.069)	(0.069)	(0.281)	(0.282)		
Husband's Log-C							0.193**	0.175*
x (Husb More Clust)							(0.069)	(0.070)
Wife's Log-C							0.163*	0.148*
x (Wife More Clust)							(0.069)	(0.070)
Family Has Child				-0.598**		-0.769*		-0.946**
				(0.112)		(0.310)		(0.070)
Constant	-3.025**	-2.953**	-2.979**	-2.464**	-3.801**	-3.199**	-3.887**	-3.222**
	(0.113)	(0.145)	(0.139)	(0.167)	(0.637)	(0.676)	(0.129)	(0.136)
LR Chi-Squared	14.7**	20.2**	19.8**	44.9**	0.2	5.8*	7.8*	172.4**
Observations	47354	40616	27201	27201	2810	2810	63233	63233

Notes: From CPS March Supplement 2003-2010. Standard errors in parentheses. Excludes workers reporting the primary reason for relocation was for reasons other than work or job transfer. Occupations are those reported within one year of migration.

** Significant at the 1 percent level.

* Significant at the 5 percent level.

TABLE 4–Husbands and Wives Labor Force Participation and Relative Occupational Clustering by Mean Age and Whether Both Have Bachelor's Degrees

	All Couples		"Power" Couples	
Family Type	Age 25-45	Age 46-65	Age 25-45	Age 46-65
Dual-Earner Couples				
Husband's Occ More Clustered	61.6%	62.6%	57.2%	61.0%
Wife's Occ More Clustered	34.2%	32.8%	33.9%	30.5%
Same Occupation	4.2%	4.6%	8.9%	8.5%
Single-Earner Couples				
Only Husband ILF	86.5%	71.6%	92.6%	78.8%
Only Wife ILF	13.5%	28.4%	7.4%	21.2%
Observations	1,334,326	938,715	224,702	160,367

Notes: Data Source: 2000 Decennial Census 5% PUMS. Includes married couples with spouse present. "Power" couples are defined as families where both heads have bachelor's degrees. dual-earner couples, 31% of couples are single-earner couples, and 7% of couples have no labor Among all couples, 62% of couples are force participants. Among power couples, 71% are dual-earner couples, 26% are single-earner couples, and 3% have no labor force participants. "Ages" are mean age of husband and wife.

TABLE 5–Logistic Regression Predicting Workers' Sex (Female = 1, Male = 0) by Occupational Clustering

	All Occupations		Majority-Bachelor's Holding Occupations	
	(1)	(2)	(3)	(4)
Log-Clustering	-0.818**	-0.467**	-1.536**	-1.250**
	(0.183)	(0.177)	(0.252)	(0.196)
Occ. Mean Age		-0.004		-0.140**
		(0.022)		(0.035)
Occ. Mean Hours		-0.023**		-0.024**
		(0.001)		(0.001)
Occ. Explosive Strength		-0.020**		0.009
		(0.073)		(0.010)
Occ. Assisting and Caring		0.040**		0.001
		(0.008)		(0.005)
Occ. Math Level		-0.009		-0.026**
		(0.011)		(0.008)
Constant	-1.868**	-1.380	-3.063**	2.925
	(0.348)	(1.018)	(0.436)	(1.494)
Clusters (Occupations)	474	429	97	88
Unique Observations	8,295,671	7,567,351	1,475,610	1,409,053
Pseudo R-Squared	0.034	0.060	0.086	0.124

Source. 2000 Decennial Census 5% PUMS and O*NET.

Notes: Standard errors clustered by occupation in parentheses. The occupation's mean age and weekly hours are calculated from the Census. “Explosive strength” "assisting and caring for others,” and “math” requirements are linear terms merged from O*NET. Some miscellaneous occupations cannot be matched in O*NET.

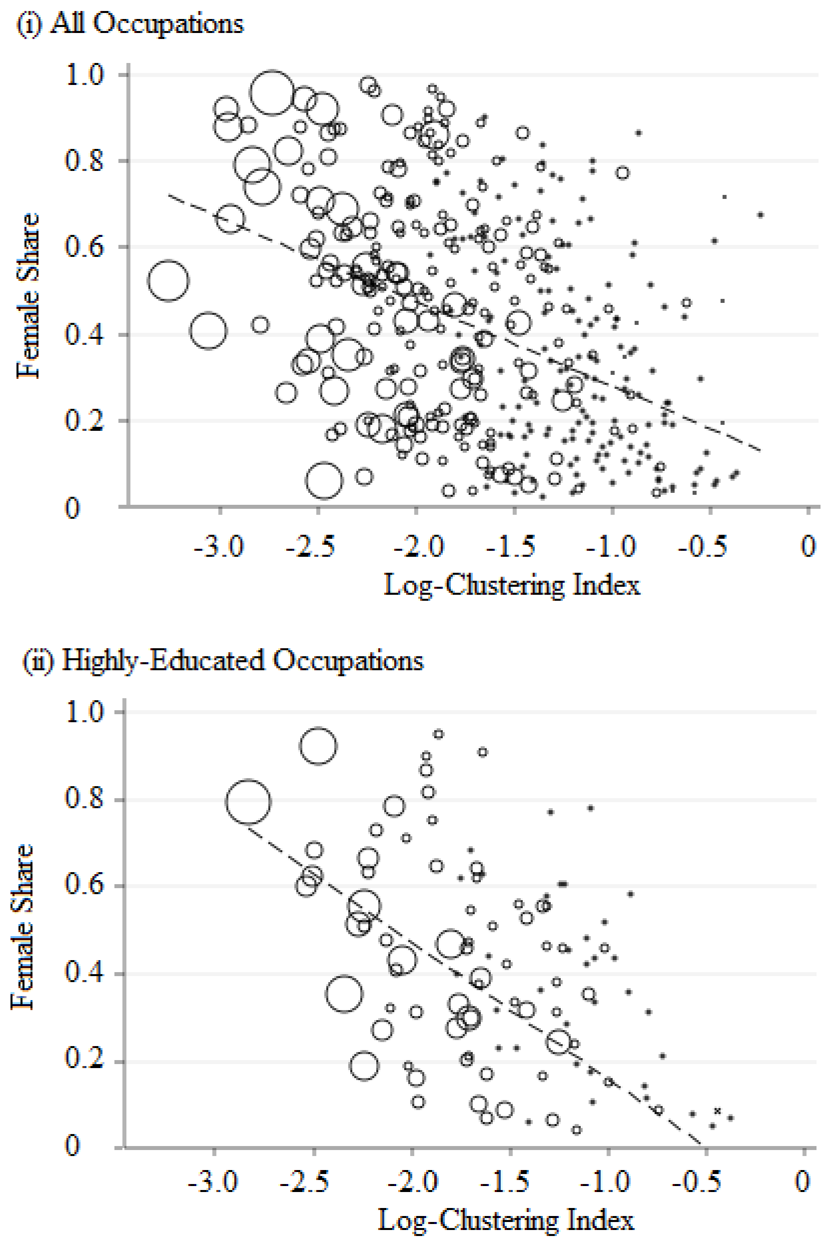
** Significant at the 1 percent level.
* Significant at the 5 percent level.

TABLE 6–Quantile Regression-Estimated Median Wages and Clustered Occupation Wage Premiums by Education, Age, and Sex

	Age 25			Age 40			Age 55		
	Men	Women	Gap	Men	Women	Gap	Men	Women	Gap
(i) Graduate									
Clustered	$16.58	$15.80	5%*	$26.39	$20.99	26%*	$27.53	$20.74	33%*
	(0.16)	(0.13)		(0.06)	(0.05)		(0.07)	(0.07)	
Dispersed	$12.24	$13.51	-9%*	$22.92	$21.09	9%*	$25.56	$23.00	11%*
	(0.21)	(0.10)		(0.07)	(0.03)		(0.07)	(0.04)	
Clust. Premium	35.3%*	16.9%*		15.1%*	-0.4%*		7.73%*	-9.8%*	
(ii) Bachelor's									
Clustered	$15.25	$13.67	12%*	$23.87	$18.40	30%*	$24.02	$17.26	39%*
	(0.06)	(0.08)		(0.03)	(0.05)		(0.04)	(0.07)	
Dispersed	$12.06	$12.26	2%*	$18.91	$16.86	12%*	$19.41	$17.50	11%*
	(0.09)	(0.05)		(0.05)	(0.03)		(0.06)	(0.03)	
Clust. Premium	26.4%*	11.4%*		26.2%*	9.10%*		23.7%*	-1.3%*	
(iii) Some Coll./A.D.									
Clustered	$10.86	$9.34	16%*	$17.04	$13.02	31%*	$17.71	$13.03	36%*
	(0.02)	(0.02)		(0.01)	(0.01)		(0.01)	(0.02)	
Dispersed	$9.74	$8.66	12%*	$14.37	$12.22	18%*	$14.60	$12.76	14%*
	(0.02)	(0.01)		(0.02)	(0.01)		(0.02)	(0.01)	
Clust. Premium	11.4%*	7.81%*		18.6%*	0.06%*		21.3%*	2.10%*	
(iv) High School/Less									
Clustered	$9.15	$7.57	21%*	$12.87	$9.33	38%*	$13.19	$9.60	37%*
	(0.01)	(0.01)		(0.00)	(0.01)		(0.00)	(0.01)	
Dispersed	$8.61	$7.22	19%*	$11.63	$9.15	27%*	$11.90	$9.60	24%*
	(0.01)	(0.00)		(0.01)	(0.00)		(0.01)	(0.00)	
Clust. Premium	6.34%*	4.88%*		10.6%*	1.94%*		10.8%*	0.03%	

Notes: Data Source: 2000 Decennial Census 5% PUMS. Median wage estimates are calculated from sixteen quantile regressions (two sexes, two occupational categories, and four educational attainments) with linear and quadratic terms for age. Standard errors are calculated by bootstrapping.
* Significant at the 1 percent level.

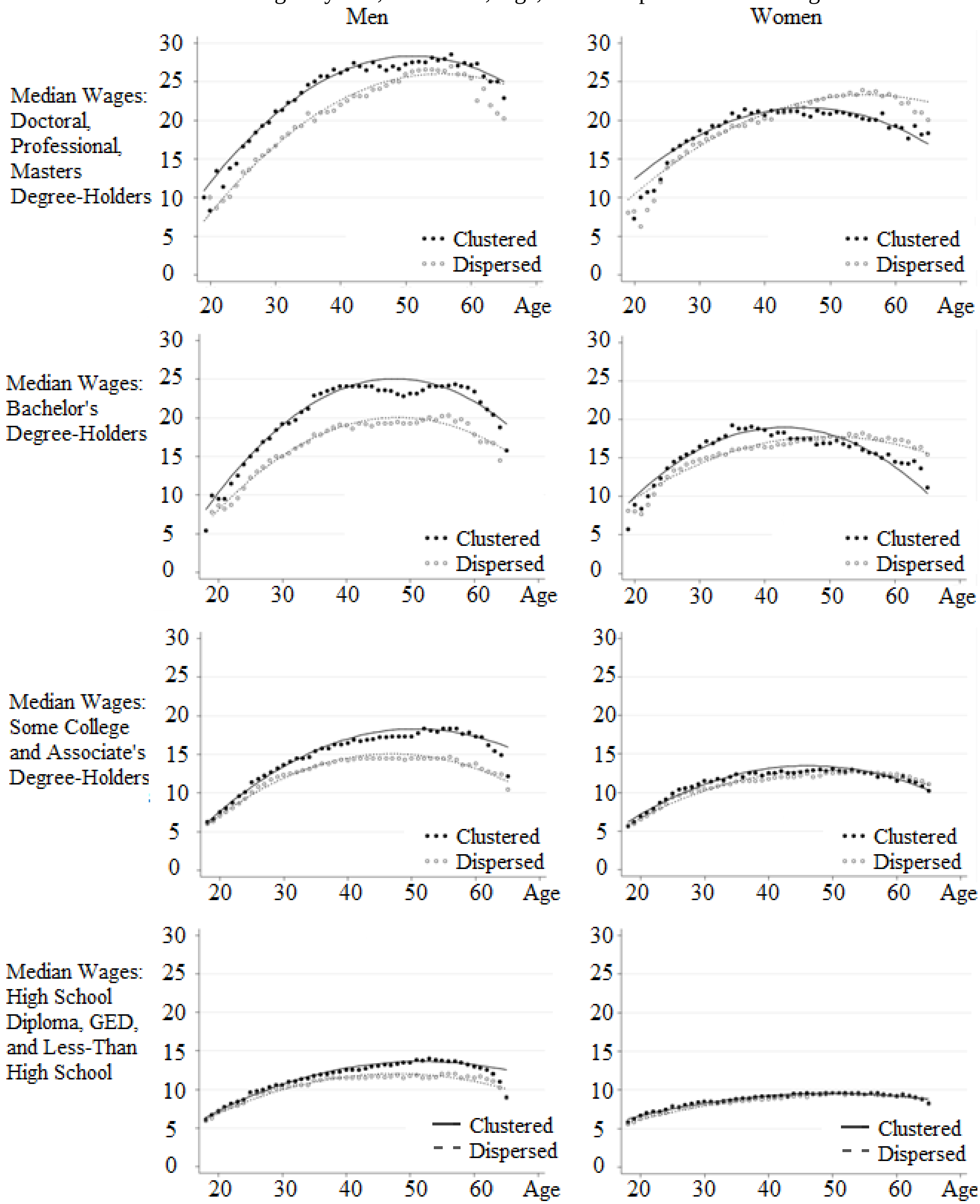
FIGURE 1
Scatterplot and OLS Fit of Female Share versus Log-Clustering Index



Notes. Circle size denotes number of workers in the occupation. Scatterplot (i) excludes occupations in which segregation might be explained due to high physical strength requirements (extraction, construction, production, farming, and military). Scatterplot (ii) excludes occupations in which the majority of workers do not hold bachelor's degrees.

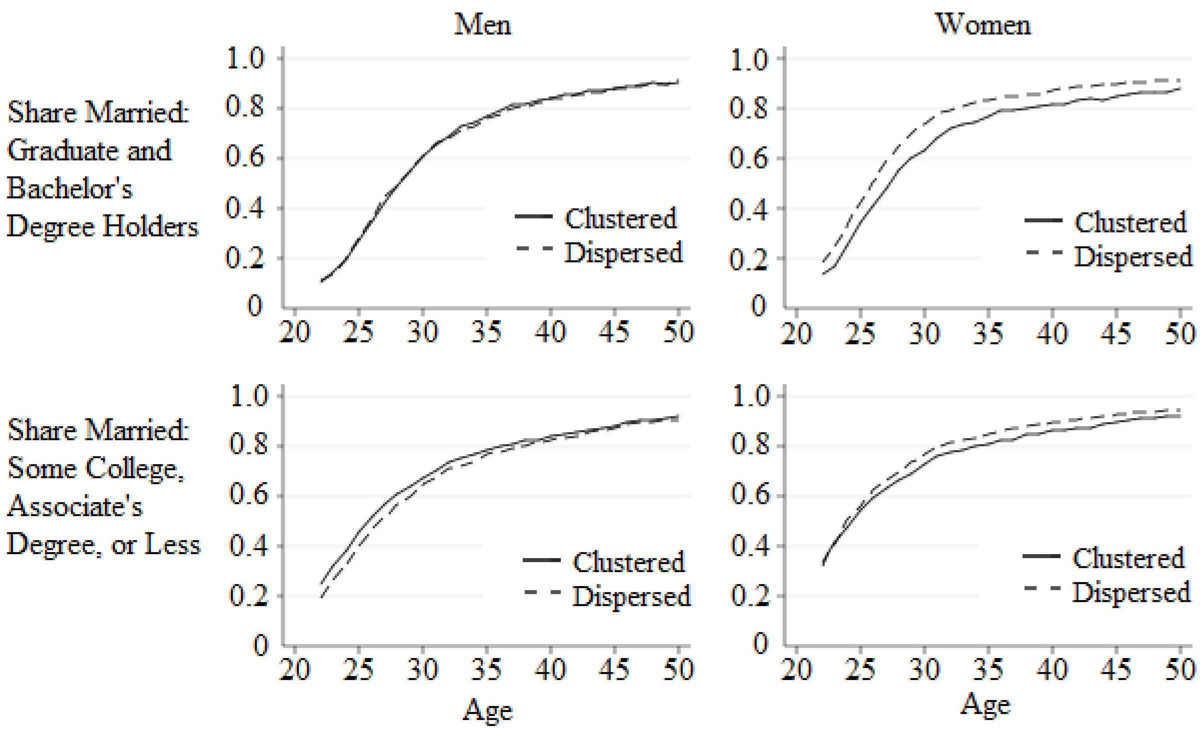
FIGURE 2

Median Wages by Sex, Education, Age, and Occupational Clustering



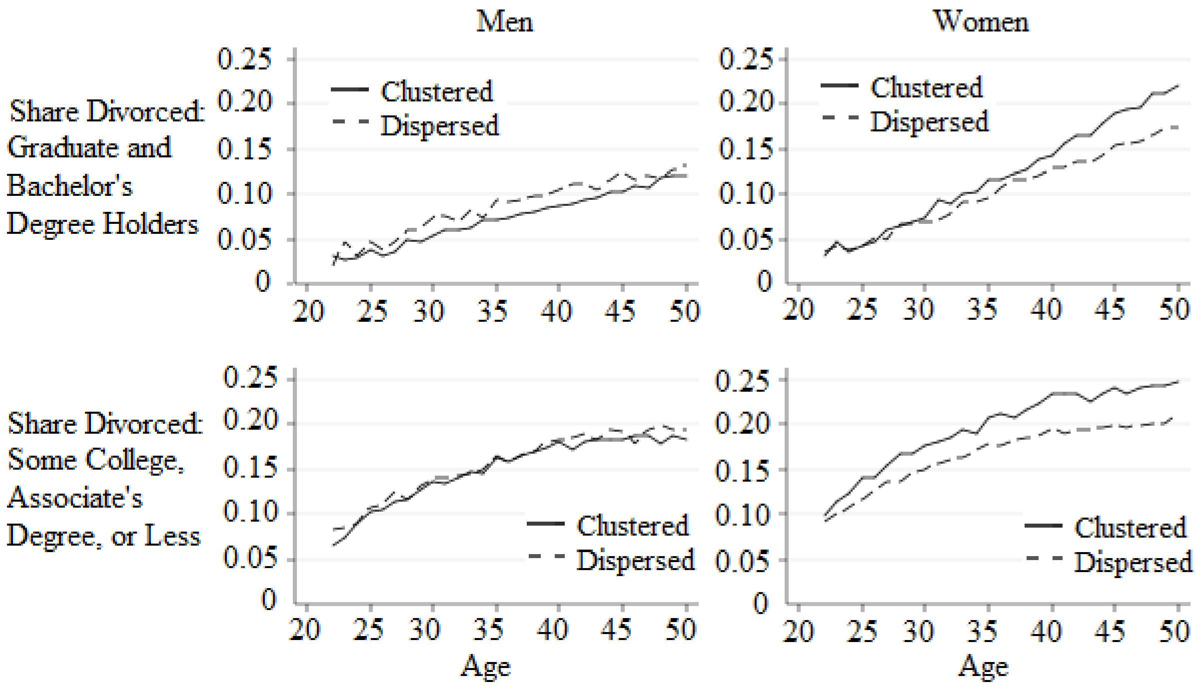
Notes. Lines represent predicted median wages using quantile regression with linear and quadratic age terms.

FIGURE 3
Delayed Marriage: Share of Men and Women
Who are Non-Never Married, by Age, Occupation Type, and Education



Notes. Shares are calculated separately at each year of age. Standard errors are at most 0.0057.

FIGURE 4
Share of Men and Women who are Divorced (and Not Re-Married),
by Occupational Clustering and Education



Notes. Shares calculated separately at each year of age. Standard errors are at most 0.0043.