NBER WORKING PAPER SERIES

GENDER-TARGETED JOB ADS IN THE RECRUITMENT PROCESS: EVIDENCE FROM CHINA

Peter Kuhn
Kailing Shen
Shuo Zhang

Working Paper 25365
http://www.nber.org/papers/w25365

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2018

This research is supported by the National Natural Science Foundation of China through Grant No. 71203188, "Impacts of Hukou, Education and Wage on Job Search and Match: Evidence Based on Online Job Board Microdata". We thank Austin Jones for careful research assistance. This paper benefited from many helpful comments from seminar and conference participants. Special thanks are due to Eliza Forsythe, Lisa Kahn, Marianna Kudlyak, Michael A. Kuhn, Ioana Marinescu and Benjamin Villena-Roldan. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Gender-Targeted Job Ads in the Recruitment Process: Evidence from China
Peter Kuhn, Kailing Shen, and Shuo Zhang
NBER Working Paper No. 25365
December 2018
JEL No. J16,J63,J71

ABSTRACT

We document how explicit employer requests for applicants of a particular gender enter the recruitment process on a Chinese job board. We find that 95 percent of callbacks to gendered jobs are of the requested gender; worker self-selection ("compliance" with employers' requests) and employer callback decisions from applicant pools ("enforcement") both contribute to this association, with compliance playing the larger role. Explicit gender requests account for over half of the gender segregation and gender wage gap observed on the board. Ad-level regressions with job title and firm fixed effects suggest that employers' explicit gender requests have causal effects on the gender mix of applications received, especially when the employer's likely gender preference is hard to infer from other contents of the ad. Application-level regressions with job title and worker fixed effects show that both men and women experience a callback penalty when applying to a gender-mismatched job; this penalty is significantly greater for women (44 percent) than men (26 percent).

Peter Kuhn
Department of Economics
University of California, Santa Barbara
2127 North Hall
Santa Barbara, CA 93106
and NBER
pjkuhn@econ.ucsb.edu

Kailing Shen
Research School of Economics
ANU College of Business & Economics
HW Arndt Building (25a)
The Australian National University
Canberra ACT 0200
Australia
kailing.shen@gmail.com

Shuo Zhang
University of California, Santa Barbara
701 Bolton Walk
Apt 101 Goleta
Goleta, CA 93117
shuo_zhang@ucsb.edu

A data appendix is available at http://www.nber.org/data-appendix/w25365

Statements in a job ad that either men or women are preferred are widely used in emerging-economy labor markets.[1] This practice, which we call gender profiling, has been studied by Kuhn and Shen (KS, 2013) in China, and by Delgado Helleseter, Kuhn and Shen (DKS, forthcoming) in China and Mexico. Based on these studies, a number of empirical regularities have been established. For example, in all the data sets that have been studied so far, gender profiling is overall quite symmetric in the sense that a roughly equal number of job ads request men and women. Profiling is also job-specific in the sense that a substantial share of the variation in requested gender occurs across jobs within the same firm. In addition, gender profiling (in favor of both men and women) is much more common in jobs requiring low skill levels, whether measured by education or experience requirements or the offered wage. Finally, there is a strong interaction between employers' stated age and gender preferences, with the mix of requests strongly favoring women at young ages and men at higher ages. Some, but not all of this 'age twist' is connected to employers' frequent requests for young, physically attractive women in helping or customer-contact positions, and for older men in managerial positions.

While the above research has provided useful new facts about gender profiling, the scope of its contribution is constrained by the type of data used: these studies are based on samples of job *ads* only. Thus, while we now know what employers *ask for* (in terms of age and gender) in a large population of jobs, we do not yet know how workers respond to such requests in their application behavior, nor how serious employers are when they make such requests. At one extreme, advertised gender requests could be hard requirements in the sense that gender-mismatched applications are always rejected, or are successful only when no workers of the requested gender apply. If so, one might also expect workers' application decisions to strongly conform with firms' stated requests. At the other extreme, advertised gender requests could just be soft suggestions that a particular gender is preferred, or even that a particular gender might prefer working in that job (for example due to the presence of same-sex co-workers or a flexible work schedule). In this 'soft' case, gender-mismatched applicants could fare quite well, or indeed just as well as gender-matched applicants when they apply.

To measure how gendered job ads interact with workers' application decisions and employers' callback behavior, this paper studies applicant and callback pools to job ads on a Chinese job board (XMRC.com) over a six-month period in 2010. A key advantage of this data is

---

[1] Appendix 1 provides examples of explicitly gendered job ads from the ten most populous countries served by Indeed.com ("the world's #1 job site"), representing 57 percent of the world's population. With the exception of the United States, gendered ads were easy to find on all the remaining platforms. A similar search on Computrabajo.com (which serves 20 Spanish-speaking countries) quickly detected explicit gender requests on all the larger platforms --including Colombia, Mexico, Colombia, Argentina, Peru and Venezuela-- with the exception of Spain and Chile.

that --in addition to knowing the characteristics of all the ads (including the requested gender, if any)-- we know the gender and qualifications of every person who applied to each ad, and of persons who were called back to a subset of the ads. This allows us to address three descriptive questions, and two causal questions, none of which to our knowledge have been addressed before.

On the descriptive side, our first goal is to measure the total amount of gender *matching* that occurs. Gender matching is a simple summary indicator of the relationship between gender requests in job ads and actual recruitment outcomes: In jobs that request a particular gender, what share of successful (i.e. called-back) applicants is of the requested gender? Having measured the amount of gender matching, we can then partition it into portions attributable to applicants' *compliance* with employers' gender requests (i.e. to applicant self-selection), and to employers' *enforcement* of their own requests when choosing workers from the applicant pool (i.e. to active selection by employers). In other words, if employers do indeed end up hiring the gender they requested, is that mostly because only workers of the requested gender apply, or because employers actively reject a large number of gender-mismatched applicants?

Our second descriptive task is to measure the total amount of gender *segregation* in successful applicant pools across occupations, firms and jobs, and to quantify the relationship between segregation and explicit gender designations in job ads. Are explicit gender labels so rare, or so weakly correlated with hiring decisions that they can play only a minor role in gender segregation among successful applicants? Or can the labels account for most of the *de facto* segregation that occurs? If so, does their role reflect mostly workers' compliance or firm's enforcement actions? Conceptually, these questions are isomorphic to quantifying the role of 'red lining' practices –the historic designation of U.S. urban neighborhoods by race -- in residential racial segregation (Aaronson et al., 2017). Abstracting from gendered job labels, we can pose an additional question about workplace gender segregation that to our knowledge has not been answered: does the observed level of gender segregation across *all* jobs –not just the gender-targeted ones-- result mostly from workers' self-sorting in deciding where to apply, or from employers' active selection among applicants? Existing studies of occupational segregation have not been able to address this question due to the absence of data on workers' application behavior.

Our third descriptive task is to measure the share of the gender *wage gap* that is associated with explicit gender requests. Specifically, suppose we knew nothing about jobs except their wage and the gender labels attached to them. What share of the market-wide gender wage gap could we account for with just this information?

On the causal side, our goals are to estimate the effects, in this labor market, of small, exogenous changes in the behavior of a single firm or worker on outcomes affecting that firm or worker.[2] For firms, we seek an answer to the following thought experiment: Holding constant the other contents of a single job ad (and of all other ads in the market), what would happen to the gender mix of applications the ad received if we exogenously switched its gender label from neutral (i.e. neither gender is specifically requested) to male or female? Answering this question reveals the extent to which the presence of one particular word in a job ad directs applicants' job search. To address this question, we regress the gender mix of applications to a job ad on indicators for explicit requests for men or women, with controls for an extensive list of ad and job characteristics, including the qualifications requested, the wage posted, firm fixed effects, and job title fixed effects. Importantly, Marinescu and Wolthoff (2017) show that job titles are more detailed and more predictive of wages and application decisions than are six-digit SOC codes. These detailed controls for job characteristics are critical because, for example, women may disproportionately apply to female-labeled jobs for a variety of reasons, including occupational preparation and working conditions, that are signaled by features of the job ad other than the requested gender.

On the worker side, imagine a worker who has submitted an application to a non-gendered job ad. Now, imagine that he or she re-directed that application to an ad that was identical in all respects (including the job title, firm, wage and requested qualifications) except that the ad requested a person of the opposite gender. How would the worker's callback probability change? Answering this question reveals how 'hard' employers' gender requests are in this market, i.e. the extent to which attaching a gender label to a job reflects a *ceteris paribus* intent by employers to enforce a particular gender preference when workers apply. To address this question, we regress an indicator of whether an application received a callback on indicators for the six possible matches between worker types (men and women) and job types (male, female, and no gender request). Included are detailed controls for firm and job characteristics, for the match between the job's requirements and the worker's qualifications, and most importantly both job title and worker fixed effects.

Controlling for worker fixed effects in this context is critical because, for example, workers who choose to apply to gender-mismatched jobs (e.g. women who apply to jobs that explicitly request men) may be highly selected. On the one hand, if most women who apply to male jobs do so because they are applying indiscriminately, those women are likely to be negatively selected (i.e. less productive than women who avoid gender-mismatched jobs).[3] In

[2] In making these comparisons, we are implicitly assuming that changing the gender label on a single job, or redirecting a single application made by one worker has no effect on aggregate behavior and expectations.
[3] To simplify the discussion, throughout the paper we refer to jobs that request men as "male jobs" or "men's jobs" and jobs that request women as "female jobs" or "women's jobs". "Gender-mismatched" applications refer to

that case, women's raw callback penalty from applying to a male job will overestimate the effect of applying to such a job on a woman of fixed ability. On the other hand, women who apply to male jobs may do so primarily because they feel they are *better* qualified on some dimension –such as education or experience—that compensates for being of the 'wrong' gender. In that case, naïve estimates will underestimate the penalty faced by equally-qualified women when they apply to men's jobs.

Our main results are as follows. First, we find that total gender *matching* is high: on a job board where 42 percent of the ads request a specific gender, 95 percent of callbacks to gendered jobs are of the requested gender. Worker compliance is also high, with 92.4 percent of *applications* to gendered job ads having the requested gender. Both of these figures are quite similar for male and female jobs, and are higher than matching and compliance rates defined analogously for employers' age, education and experience requests.[4] Firms' enforcement decisions reinforce these application patterns, but are far from lexicographic: Among applicants to explicitly female jobs, men are 80 percent as likely to get a callback as women; female applicants to male jobs are 45 percent as likely to be called back as men. Second, in an accounting sense, a large majority --74 percent-- of the gender matching between actual callback pools and firms' gender requests can be attributed to applicants' 'compliance', or self-selection into gender-targeted jobs.[5]

Third, turning from gender matching to gender *segregation*, we use noise-adjusted segregation measures (Carrington and Troske 1997) to calculate that 59 (59) [53] percent of the gender segregation across *all* jobs (firms) and [occupations] on this job board is associated with the explicit gender labels attached by employers to jobs. Like our results for gender matching, self-selection decisions by workers account for almost all of this label-linked segregation. Abstracting from gendered job labels, we find that workers' self-selection decisions can account for the vast majority –97 percent-- of overall gender segregation across the jobs, firms and occupations in our sample. The intuition is that application pools are so gender-segregated that –holding these pools fixed—even a completely gender-neutral callback policy would have little effect on overall segregation. Fourth, explicit gender labels on jobs can account for 61 percent of the gender *wage gap* among successful (called-back) applicants on this job board.

---

mismatch between employers' requests and the applicant's gender. Later, when we use a machine-learning approach to predict the likelihood that a particular job title would request men, we refer to jobs whose titles are associated with frequent requests for men (women) as "implicitly male (female)" jobs.

[4] See Section 2 for our exact definitions of matching on these dimensions. For example, in the case of age we use the share of callbacks that fall into the age range that is explicitly requested in the job ad.

[5] We emphasize the 'accounting' nature of this exercise because high self-sorting could be caused by high enforcement, as we argue later in the paper.

Fifth, regression analysis strongly suggests that explicit gender requests in job ads have causal effects on workers' application decisions. With interacted fixed effects for firms and job titles, we estimate that the presence of a male job label reduces the female share of applicants to a job by 15 percentage points, while a female label raises the female share by 25 percentage points. Both of these effects, as well as the difference between them, are highly statistically significant. We conclude that explicit gender labels in job ads do appear to convey relevant information to prospective applicants that cannot be inferred from the other contents of the ad, and that this information directs workers' applications.

To understand why female job labels have a larger estimated effect on applicant gender mix than male labels do, we use a Bayesian machine learning approach (McCallum and Nigam 1998) to identify job ads whose gender preferences can be clearly predicted from other words in the job title, and those that cannot. Consistent with the hypothesis that prospective applicants try to infer their hiring prospects from all the information contained in the ad, we find that explicit gender labels have the largest effects on applicant gender mix in jobs whose title does not suggest a clear gender preference. Essentially, among those jobs, women tend to apply only when the job explicitly requests women, while men *abstain* from applying only when the job explicitly requests women. Thus, adding a female label has a more powerful effect on applicant gender mix. This pattern –that men are more likely than women to apply when jobs aren't clearly 'for' their gender-- echoes existing findings that female job searchers are more ambiguity-averse, and more responsive to affirmative action statements than men (Gee 2018, Ibanez and Reinter 2018).

Finally, regression analysis suggests that explicit gender labels in jobs also have ceteris-paribus effects on workers' callback chances, in the sense that labels predict how the same worker's callback chances would change when applying to identical jobs making different gender requests. Specifically, controlling for both worker and job title fixed effects, a man's callback probability is estimated to fall by 2.3 percentage points (or 26 percent) if he applies to an explicitly female job compared to a nongendered job. Women's callback probability is estimated to fall by a greater amount –3.8 percentage points or 44 percent-- if she applies to an explicitly male job compared to a nongendered job. Interestingly, both these effects are smaller in magnitude than the regression-unadjusted differentials, suggesting that, if anything, workers who apply to gender-mismatched jobs are negatively selected.[6]

Our paper contributes to a number of literatures, the first of which uses the contents of job ads to study labor markets. Such ad-content studies include Hershbein and Kahn (2015) and Modestino, Shoag and Balance (2015) both of whom ask whether employers request

[6] Point estimates indicate that this negative selection is greater for men applying to female jobs than vice versa, but this difference is not statistically significant.

higher qualifications for the same jobs when local labor market conditions make workers "easier to get". Brencic and Norris (2009, 2010, 2012), and Brencic (2010, 2012) use the same type of data to study aspects of employers' recruiting strategies, including whether to post a wage and whether to adjust ad contents during the course of recruitment. Relative to this literature, a key advance of our paper is the use of internal job board data to see whether and how such changes in ad content actually matter: do they direct workers' search, and do they inform potential applicants of how employers will respond when workers who do not meet the advertised criteria apply?

Second, our paper relates to a large literature that studies racial, gender, and other differentials in callback rates using resume audit methods (Bertrand and Mullainathan 2004, Kroft et al. 2013, Neumark et al. 2015). While our estimates of callback differentials are not experimentally based, a key advantage of our job-board-based approach is that it lets us study callbacks to the entire population of jobs on offer, which vary dramatically in their gender preferences. For example, even though a roughly equal number of jobs on XMRC request women and men, 85 percent of ads for front desk personnel explicitly request women, and 88 percent of ads for security personnel explicitly request men (DKS, forthcoming). This extreme heterogeneity poses a challenge for audit studies, which typically elicit an average race or gender preference in a relatively narrow set of jobs, often selected to be race- or gender-neutral.[7] In contrast, a key parameter in our approach *is* this heterogeneity, as captured by our *mismatch penalty* parameter: how does, say, a woman's callback probability change when she redirects her application from a nongendered to a female job? Notably, our estimates of the mismatch penalty control for unobserved worker quality by using worker fixed effects, since we can observe the same worker applying to different types of jobs.

Our job-board-based approach also broadens the study of race- and gender differentials in recruiting beyond callback differentials, to workers' application decisions and their interaction with callback differentials. When we do this, as noted we find that the vast majority of gender segregation in labor markets is not connected –at least directly—with the main parameter estimated in resume audit studies: how resumes are treated when they are submitted to employers. Instead, job segregation is closely connected to workers' choices on where to apply. Uniquely, job boards provide the opportunity to study application and callback decisions simultaneously across the entire spectrum of jobs on offer, and highlight the role of

---

[7] In addition to cost, a key reason for this narrow focus is the difficulty of constructing plausible resumes for a large variety of jobs, many of which are highly specialized. Thus, for example, both Bertrand and Mullainathan (2004) and Kroft et al. (2013) restrict their attention to four occupations: sales, administrative support, clerical, and customer service. Carlsson and Rooth's (2007) study is noteworthy for studying the heterogeneity in discrimination across 13 occupations.

job ad content, placement and targeting as potentially under-researched determinants of how workers are allocated to jobs.[8]

A third related literature is a substantial body of theory on directed search in labor markets (e.g. Albrecht et al., 2006). With a few recent exceptions, this literature has not examined data on how workers actually direct their applications. These exceptions include Marinescu and Wolthoff (2015); Belot, Kircher and Muller (2017); and Banfi and Villena-Roldan (forthcoming), all of whom study the effects of the posted wage on the number and quality of applications a firm receives. Marinescu and Rathelot (2015) study the geographic scope of workers' search, and Kudlyak, Lkhagvasuren and Sysuyev (2014) study how workers re-direct their search over the course of a search spell. Finally, Flory, Leibbrandt and List (2015) and Mas and Pallais (2017) study how workers' application decisions respond to competitive work environments and non-wage job attributes respectively. Given our strong estimated effects of attaching gender label to a job ad on application behavior, our results suggest that both search theorists and empirical researchers could profit from considering the effects of job ad characteristics *other than the posted wage* on workers' application decisions.

Finally, there is a large literature on gender differentials in labor markets, but very little of it has focused on the explicit gender profiling of jobs in emerging economy labor markets like the one we study here. Understanding this practice would seem to be an essential component of understanding gender differentials in labor markets in much of the world.

1. Data

As noted, our data consist of internal records of XMRC.com, an Internet job board serving the city of Xiamen. XMRC is a private firm, commissioned by the local government to serve private-sector employers seeking relatively skilled workers.[9] Its job board has a typical U.S. structure, with posted ads and resumes, on-line job applications and a facility for employers to contact workers via the site. XMRC went online in early 2000; it is nationally

[8] An emerging concern in this regard derives from the increasing capacity to micro-target all types of online ads. For example, Verizon recently placed a job ad that was set to run "on the Facebook feeds of users 25 to 36 years old who lived in the nation's capital, or had recently visited there, and had demonstrated an interest in finance" (Angwin, Scheiber and Tobindec, 2017). In contrast to the Chinese case that we study –where all applicants can view all ads—in the Facebook case non-targeted workers were not even aware of the ad's existence.

[9] The other major local job site, XMZYJS, is operated directly by the local government. It serves private sector firms seeking production and low-level service workers. Unlike XMRC, XMZYJS does not host resumes or provide a service for workers and firms to contact each other through the site.

recognized as dominant in Xiamen, possibly due to its close links with the local government and social security bureau.[10]

To study the effect of gender profiling on application and callback patterns, we began with the universe of ads that received their first application between May 1 and October 30, 2010. We then matched those ads to all the resumes that applied to them, creating a complete set of applications. Finally, for the subset of ads that used XMRC's internal messaging system to contact applicants, we have indicators for which applicants were contacted after the application was submitted. This indicator serves as our measure of callbacks. Our primary dataset for the paper is this subset of ads for which callback information is available, which comprises 3,637/42,744 = 8.5 percent of all ads. Summary statistics for this sample are very similar to the universe of ads, shown in Appendix Table A1. In addition, Section 4 replicates our analysis of application decisions –which does not require information on callbacks—in both the larger and smaller samples, with very similar results.

Aside from being the only integrated dataset of ads, resumes, applications and callbacks we are aware of –especially in an environment that permits gendered job ads--, an important advantage of our 2010 XMRC sample is its simple and unambiguous indicator of employers' gender requests.[11] On many job boards (both in China and elsewhere), employers' gender requests must be inferred by parsing the text of the ad. In this process, the researcher needs first to identify all the different ways a firm *could* indicate a gender preference, as well as the many ways that gendered words serve a different purpose, like conveying information about the job (e.g. selling or making women's clothing), and 'inclusion' statements like "open to both men and women".[12] Sometimes judgment calls are necessary, for example in deciding whether the adjective 'beautiful' can describe both men and women. On XMRC, in contrast, employers are required to complete a 'desired gender' field (indicating male, female, or no preference) when creating a profile for each job. This datum is then made visible to workers (and can be entered by workers into a search query). Thus, our measure of whether the employer makes a gender preference statement is clear, standardized, and salient to all jobseekers on the site.

A second key advantage of our setting is the relatively simple nature of the search technology on the site: In 2010 (and still today), XMRC's site largely emulated printed job ads, where workers peruse ads using simple search filters to decide where to apply. More recently (and coming soon to XMRC), many job boards use machine learning to display suggested job

---

[10] XMRCs offices are in the same building as complementary local government offices (e.g. for social security and payroll taxation), offering employers the advantage of 'one-stop shopping' for employment-related services.
[11] Unfortunately, the callback indicator in our 2010 XMRC extract is not available for more recent years, because the firms using the site have transitioned away from using XMRC's internal messaging system as their means of contacting workers. As we discuss below, there may be other reasons to focus on data from before 2012 as well.
[12] This was the procedure used to identify gendered job ads on Zhaopin.com in Kuhn and Shen (2013).

matches to individual workers based on the worker's location, qualifications, employment history and recent searches. In these cases, the jobs a worker applies to are jointly determined by the jobs that are suggested to her by the board's algorithms *and* her choices from that set.[13] This joint determination does not apply to our data.

Third, the environment in Xiamen in 2010 was remarkably free of legal impediments to posting a gendered job ad, and free of stigma attached to employers posting such ads. While China's constitution has formally given women equal rights since 1982, these principles had few practical consequences for labor markets until July 2012, when the first lawsuit claiming gender discrimination in employment was filed. The first regulations that appear to have constrained firms' ability to post gendered job ads appeared in May 2016, when China's Ministry of Industry and Information Technology clearly specified fines for both job boards and employers posting such ads.[14] Since then, some Chinese job boards (especially those serving highly skilled workers) have responded by eliminating –or at least making it hard to find-- overtly discriminatory job ads on their sites. Others, including XMRC, continue to host gendered ads despite the new regulations. Even boards that have eliminated gendered ads, however, continue to allow indirect signals of their employers' desired gender (such as "gentleman" (绅士), "beautiful face" (面容姣好), and "little brother" (小哥哥)). Perhaps more importantly, these sites also allow recruiters to filter applications and resumes by gender, making it easy to restrict their attention to a preferred gender.[15]

In sum, while gendered recruitment by employers is still present in China's new legal environment, it is more varied in form and harder to detect because of the new incentives to hide it. XMRC in 2010 thus provides a much cleaner picture of how employers would choose to advertise jobs when unconstrained, and of how employers treat applications that do not match a measure of gender preferences that employers have few incentives to misrepresent. Arguably, our XMRC data may also provide insights for how gendered job ads work in countries where they remain largely unregulated.

In all, our primary dataset comprises 229,616 applications made by 79,697 workers (resumes) to 3,637 ads, placed by 1,614 firms, resulting in 19,245 callbacks. Thus there was an average of 63 applications per ad and 5.3 callbacks per ad. One in twelve applications received a callback, while one in four resumes received a callback. Descriptive statistics are provided in Tables 1 and 2 for ads and applications respectively. Table 1 shows that 867/3,637 = 24 percent of ads requested female applicants, 18 percent requested male applicants and the remaining 58

[13] We do not observe which ads were viewed by workers; thus our estimated effects should be viewed as incorporating workers' decisions regarding which types of jobs to search for.
[14] See Appendix 2 for additional details on China's labor laws as they apply to gender profiling in job ads.
[15] See Appendix 3 for a survey of gender targeting on Chinese job boards today.

percent did not specify a preferred gender.[16] The average years of requested education were 12.2, and were more than a year higher in jobs requesting women than men. Forty-eight percent of ads specified a preferred worker age; the mean requested age was 28. Consistent with the age twist identified in DKS, the requested age was considerably lower for jobs specifically requesting women. On average, one year of experience was requested. 58 percent of ads posted a wage; the mean posted wage was 2,446 RMB per month overall but only 2,001 RMB in jobs requesting women.

Table 2 shows that 124,275/229,616 = 54 percent of applications came from women. The typical application had 14.35 years of education, with women holding about half a year more education than men. Average applicant age was 24.0 years. Other applicant characteristics observed in our data (and used in the regression analysis) include experience, new graduate status, marital status, current wage (when provided), myopia, height, the number of experience and job spells listed, and whether an English version of the resume is available.

To provide some context for the sample of jobs and workers on XMRC, Table A2 compares the characteristics of job ads on XMRC with those of private-sector employees in Xiamen and in urban China, respectively.[17] The employment data are taken from the 2005 Chinese Census 1% microdata sample. Clearly, the ads on XMRC seek workers who are considerably younger, better educated, better paid, and more female than the employed population of Xiamen, or of a typical large Chinese city. This is as we might expect, for three reasons. The first is XMRC's explicit niche in the local labor market: to serve relatively skilled workers. Second, due to a massive recent expansion of China's education system, younger cohorts are much better educated than their parents. Thus, any job board seeking skilled workers will naturally be disproportionately seeking young workers.[18] Third, as on any job board, the ads and resumes on XMRC represent a population of vacancies and jobseekers, not of employed workers. Thus we would expect new labor market entrants (who are all looking for work) and young workers (who turn over more frequently than other workers) to be substantially overrepresented relative to the currently employed population.

Finally, the bottom panel of Table A2 attempts to compare the broad occupation distributions of XMRC ads to China's and Xiamen's urban labor force. This is challenging because of the occupational classification system used by XMRC, which uses 37 categories that

[16] This compares to 19, 18 and 63 percent in the universe of job ads. See Table A1.

[17] 'Urban China' in Table A2 and throughout this paper refers to China's largest cities –specifically the four municipalities directly under the jurisdiction of the central government (Beijing, Shanghai, Tianjin and Chongqing) plus the 15 subprovincial cities.

[18] Rapid educational upgrading since the 2005 Census also implies that Table A2 is likely to overstate the education gap between the XMRC ads and Xiamen's 2010 labor force.

were created by the website; mapping these into Census categories is a fairly subjective exercise. With these cautions in mind, Table A2 indicates that jobs in production, construction and manufacturing are under-represented on XMRC, while professional and technical jobs are highly over-represented. Again, this is consistent with XMRC's focus on skilled workers, a population we know is less subject to gender profiling than less-skilled workers.

2. Descriptive Analysis—Gender Matching

Descriptively, our first goal is to measure the extent to which the final pool of successful applicants to a job ad (i.e. the *callback pool*) reflects the employer's stated gender preferences. This concept of gender matching, G, applies only to explicitly gendered ads. We also wish to measure the relative contributions of workers' *compliance* with firms' requests and employers' *enforcement* of their own stated requirements to the total amount of gender matching that occurs. The analysis begins with some basic descriptive statistics on applications and callbacks in Table 3.

Starting with total gender matching, row 1 of Table 3 shows the share of callbacks that are female (δ) for the three job types in our data: jobs requesting women (F jobs), jobs requesting men (M jobs) and jobs that do not state a gender preference (N jobs). These statistics indicate a high congruence of the callback pool with firm's stated requests. Specifically, 94.0 percent of callbacks to F jobs are female and 100 - 3.7 = 96.3 percent of callbacks to M jobs are male. Combining F and M jobs, 94.8 percent of callbacks to gendered job ads are of the requested gender. Row 2 shows the share of *applications* to the three job types that are female (α). It suggests that applicants' compliance with employers' gender requests plays a substantial role in accounting for this high level of gender matching, since applicant pools are almost as highly sorted by gender as callback pools. Specifically, 92.6 percent of applications to F jobs are female and 100 - 7.9 = 92.1 percent of applications to M jobs are male. Combining F and M jobs, 92.4 percent of applications to gendered job ads are of the requested gender.

The remaining rows of Table 3 show that employers' enforcement of their own stated requests also helps to account for the overall amount of gender matching that occurs. Specifically, in jobs explicitly requesting female applicants, men who do apply are only 1/1.246 = 80.3 percent as likely to be called back as women. In jobs requesting men, female applicants are only 44.5 percent as likely to be called back as a man. Thus, at least in the raw data, employers' enforcement of their own gender requests is stronger against women applying to male jobs than men applying to women's jobs.

To get a better sense of the overall amount of gender matching and its components, it is useful to define the following index of gender matching:

$$G = \frac{g - g_0}{1 - g_0} \qquad (1)$$

where g is the share of gendered ads that are of the requested gender and g_0 is the share of gendered ads that *would be* of the requested gender if there was no gender matching (i.e. if we re-allocated the total population of called-back workers across all jobs --whether *F*, *N* and *M*-- so that the total number of callbacks to each job remained the same, but the gender mix of callbacks was equalized across all jobs). Thus G=1 if all callbacks to gendered jobs match the employers' request, and G=0 if the female share of callbacks (δ) equals its population average in all jobs. In our data, g =.948 and g_0 = .501, so our overall index, G = .897. In other words, on a scale where zero indicates no gender matching and 10 indicates perfect matching, the total amount of matching equals 9.

With this index in hand, we can now assess the relative contributions of compliance and enforcement to gender matching, G, using the identity:

$$\delta^J = \frac{\theta^J \alpha^J}{\theta^J \alpha^J + (1 - \alpha^J)} \qquad (2)$$

where J = *F, N,* or *M* and θ is women's relative risk of being chosen from the applicant pool, i.e. the ratio of callback rates (f/m). Equation (2) allows us to compute two counterfactual levels of g and G.[19] *Counterfactual 1* (no compliance) keeps enforcement, θ, at its actual level in each of the three job types, but sets α (the share of women in the *applicant* pool) at its population mean level in all jobs (i.e. at .541, from Table 3). *Counterfactual 2* (no enforcement) keeps compliance, α, at its actual level in each job type, but sets θ (women's relative risk of being picked from the applicant pool) at its population average .866 in all jobs. The results are reported in Table 4.

According to row 2 of Table 4, eliminating worker compliance while maintaining actual levels of enforcement would reduce the share of callbacks that are of the requested gender, g, from .948 to .617. The corresponding decline in the gender matching index, G, is from .897 to .232. Thus, workers' compliance with employers' gender requests accounts for (0.897-0.232)/0.897 = 74 percent of the gender matching in our data. According to row 3, eliminating employers' enforcement while maintaining actual levels of worker compliance would have a much smaller impact, reducing g from .948 to .921 and G from .897 to .842. Thus, active

[19] Like other indices used in this paper, the G index depends on the relative sizes of the three job types (J), as well as on the overall share of workers who are called back to each job type. Throughout the paper, we design our counterfactual thought experiments to hold both of these quantities constant, varying only the gender *mix* of workers who apply to different job types (or firms, occupations, etc.) and the gender *mix* of callbacks.

enforcement by employers of their own gender requests accounts for only (.897-.842)/.897 = 6 percent of the gender matching in our data. Because the decomposition in equation (2) is exact but nonlinear, the remaining 20 percent of gender matching is due to the interaction between compliance and enforcement.[20] We conclude that *compliance, i.e. applicants' self-sorting according to employers' gender requests in job ads, accounts for the vast majority of gender matching in gendered ads.* The intuition is straightforward: Because applicant pools are so highly gender-segregated, even completely equal treatment of male and female applicants in all job types would have only a small impact on the gender mix of callbacks to each job if application patterns are held fixed.

To put our estimates of gender-matching, compliance and enforcement in context, Table A3 presents comparable measures of those three quantities for employers' gender, age, education and experience requests, as well as for the match between the posted wage and the applicant's current wage (when reported). Thus, for example, row 2 shows the share of *called-back workers* whose age is within the ad's requested age range (e.g. 24-28), the share of *applications* whose age is in the requested range, and the share of age-mismatched applications that are rejected.[21] Interestingly, compliance, enforcement and total matching are all greater for gender than for these other four characteristics. While these differences are particularly dramatic on the worker self-selection side, substantial enforcement differences are also present: The shares of age-, education-, experience-. or wage-mismatched applicants that are called back all exceed 25.2 percent, compared to 5.2 percent of gender-mismatched applicants. Together, these statistics suggest an especially important role for gender, relative to these other characteristics, in what employers and employees consider to be a good match.

3. Descriptive Analysis—Gender Segregation and the Gender Wage Gap

In this second part of our descriptive analysis, we broaden our focus beyond the gendered jobs to all the jobs in our sample. Motivated by evidence of high levels of gender segregation across occupations (Blau et al. 2013), across firms (Card et al. 2016) and across jobs within firms (Bielby and Baron 1984), we wish to assess the contribution of the explicit job labels (*F*, *N* and *M*) to gender segregation across all these partitions of the labor market. As noted, measuring the contribution of explicit gender designations for jobs to gender segregation is mathematically analogous to the measuring the association of *red-lining* with urban residential segregation (Aaronson et al., 2017), a practice which in some cases gave

[20] By 'exact' we mean that eliminating both compliance and enforcement would reduce G to zero.

[21] Mismatch in education, experience and wages is measured by the indicators used in Table 9's callback regressions, which are based on broad categories. For example, education is measured using five categories (primary, middle, technical school, post-secondary and university) and a match occurs when the job's request and the employee's actual education fall into the same category. Additional details are provided in Table A3.

official government sanction to explicit racial categorization of neighborhoods.[22] In our case, neighborhoods that would be categorized as black, mixed, or white are directly analogous to employers' explicit designation of jobs as *F*, *N* or *M*. In the urban context, these labels presumably allocated home seekers to neighborhoods both by directing where home seekers search for housing (compliance) and via landlords' and home sellers' refusals to transact with 'race-mismatched' persons who offer to purchase or rent a home (enforcement).

In addition to measuring the total contribution of job profiling to gender segregation, this Section also decomposes that contribution into its compliance and enforcement components. Finally, we measure the connection between gendered job ads and the gender wage gap: To what extent do explicitly gendered jobs account for the gap in wages available to successful job applicants on this job board?

3.1 Measuring Segregation

To measure segregation, we use Duncan and Duncan's (1955) segregation index, applied to the set of successful applicants (i.e. callbacks) in a unit, *i*, which can be a job ad, a firm, or an occupation. The index, *S*, can be calculated from the female shares, δ_i, in those units as:

$$S = \frac{\sum_i \gamma_i |\delta_i - \Delta|}{2\Delta(1-\Delta)} \qquad (3)$$

where δ_i is the female share in unit *i*, Δ is the female share in the population, and γ_i is unit *i*'s share of the callback population. Thus, *S* is the population-weighted mean absolute deviation of the female share from its global mean, divided by its maximum attainable value, $2\Delta(1-\Delta)$.[23] Like our gender matching index *G*, Duncan and Duncan's *S* index varies between 0 and 1. It is widely used in studies of residential segregation (Cutler et al. 1999, Logan et al. 2004). Duncan and Duncan's *S* also has a well-known, natural interpretation: In our context, it gives the share of men (or women) who would have to be reassigned to a different unit (job, firm, occupation, etc.) in order for men and women to be distributed identically across units.[24]

To use Duncan and Duncan's index in our context, however, we need to address an issue that doesn't usually arise in the residential segregation context: the effect of small unit sizes.

[22] For an example of officially sanctioned residential redlining, see Section 980 (3) of the Federal Housing Association's 1938 Underwriting Manual, which recommends "Prohibition of the occupancy of properties except by the race for which they are intended" in restrictive housing covenants. (Federal Housing Association, 1938).

[23] Equivalently, *S* can be calculated via the better known formula, $S = \frac{1}{2}\sum_j \left[\frac{\phi_i}{\Phi} - \frac{\mu_i}{\mathrm{M}}\right]$, where ϕ_i is the share of callbacks in unit *i* that go to women, $\mu_i = 1 - \phi_i$ is the share of callbacks in unit *i* that go to men, and Φ and $\mathrm{M} = 1 - \Phi$ are their population equivalents.

[24] This property is independent of which group is being re-allocated and of the relative size of the two groups (Zoloth 1976). Notably, however, the counterfactual reallocation of residents underlying this interpretation does not preserve the total populations of the units.

This effect is most important when we wish to measure and decompose segregation across individual job ads, since the average size of the callback pool to an ad in our data is 5.3 workers. Thus, purely random variation in where workers send their resumes and in which resumes are picked from the application pool could generate a considerable amount of *de facto* segregation.[25] To adjust our segregation index for the effects of random matching, we extend the one-stage sample-shuffling approach developed by Carrington and Troske (1997) to reflect the fact that the allocation of workers to jobs is actually the outcome of two random processes: the allocation of applicants to jobs, and the selection of successful applicants from applicant pools. In addition to incorporating that fact, our two-stage approach allows us to conduct counterfactual exercises that quantify the roles of compliance and enforcement in the segregation process.

More specifically, Carrington and Troske estimated the amount of racial segregation across Chicago workplaces we would expect if we took as given total employment at each workplace, and then imagined that the actual population at each workplace was a random draw from a binomial distribution whose mean black share was the population average. Simulating the Duncan-and-Duncan segregation index over multiple replications, then taking the mean of the resulting indices gave them an estimate of the amount of segregation we'd see if workers were allocated to jobs in a race-blind way. Here, we take as given the total number of applications and callbacks at every job ad. We then simulate the amount of segregation we would expect if the gender mix of applications to each ad, *and* of callbacks to each ad was the result of a random draw from binomial distributions with parameters derived from the population mean levels of α and θ. The idea is to hold fixed the total number of applications men and women make, the number of applications arriving at each job, and the total number of 'interview slots' (callbacks) available for each job. With these 'structural' features of the labor market fixed, we then assume that workers direct their applications randomly and that firms select candidates randomly. How much gender segregation would we expect to see?

In more detail, recall that the overall mean of α, $\bar{\alpha}$ =.541 and consider an ad that received 80 applications and issued 5 callbacks. We first simulate the number of female and male applications to that ad (a^f and a^m) as a random draw of 80 applications from a pool with population parameter.541, i.e. $a^f \sim B(n,p) = B(80,.541)$, $a^m = 80 - a^f$, and B indicates the binomial distribution. Next, taking this randomly-generated application pool as given (say, 51 women and 29 men), we simulate the number of male and female callbacks (c^f and c^m) as a random draw of 5 callbacks from a pool with population parameter given by:

[25] To see the point, note that if each firm calls back only one worker, segregation will always be complete: every firm's callback pool will be entirely male or entirely female.

$$p^c = \frac{\bar{\theta} a^f}{\bar{\theta} a^f + a^m} \qquad (4)$$

where $\bar{\theta}$ = .866 is the overall mean of women's relative callback risk. Thus, $c^f \sim B(n,p)$ = B(5, p^c); c^m = 5 - c^f. Doing this for every job, then calculating the realized segregation index, *S*, completes a single iteration.

Figure A1 plots the distribution of realized *S* values from 1000 iterations in this baseline scenario where there is no systematic variation across jobs in either application or callback behavior. It shows a surprisingly concentrated distribution with a mean of .317 and all values falling between .30 and .34. Thus, while random matching can generate a high level of measured segregation, the amount of segregation it generates is tightly constrained by the distribution of applicant pool sizes and callback pool sizes and the overall share of men and women in the population.

Finally, to remove the effects of this randomness, we follow Carrington-Troske by defining a *noise-adjusted* segregation measure, $\tilde{S}$, as:

$$\tilde{S} = \frac{S - S_0}{1 - S_0} \qquad (5)$$

where *S* is the unadjusted segregation index from equation (3) and $S_0 = .317$ is the mean level of segregation expected from noise in matching. Since *S* = .732, the noise-adjusted index of gender segregation across jobs in our data is given by $\tilde{S} = \frac{.732 - .317}{1 - .317} = .607$. Interestingly, this level essentially coincides with Cutler et al.'s (1999) threshold of 0.6 for defining a U.S city as having a residential ghetto.

3.2 Counterfactual Segregation Indices

Having developed a noise-adjusted measure of gender segregation across jobs, we next ask how much of this segregation is associated with gender profiling, and how much of that 'label-linked' segregation, in turn, is associated with compliance versus enforcement. To that end, we use different assumptions on α and θ to generate five counterfactual $\tilde{S}$ indices. Of these, *counterfactual A* measures the total contribution of the three job *types* (the equivalent of 'redlining' in the residential segregation context) to gender segregation across jobs. Here, instead of a common α and θ for all ads, we simulate *S* allowing both α and θ to take three levels, one for each job type. *Counterfactuals B and C* parse counterfactual A into portions related to active selection by employers versus self-selection by workers, by letting only one of α and θ vary across job types. Finally, *counterfactuals D and E* ignore the job labels (*F*, *N*, and *M*) and divide the total amount of gender segregation across jobs into a worker self-selection component and an active employer selection component. In the former case, we simulate *S* using the actual numbers of applications received by each ad, while imposing the same relative risk, θ, for all jobs. Counterfactual *E* is the mirror image of this case.

The mean levels (across 1000 iterations) of noise-adjusted segregation from all the above simulations are displayed in Table 5.[26] According to counterfactual A, if the parameters α and θ differ *only* across the three job categories (*F*, *N* and *M*), mean noise-adjusted segregation, $\tilde{S}$, would equal .359, which is 59.2 percent of all the gender segregation across jobs. Thus, about 60 percent of the total gender-segregation in the populations of successful job applicants across individual job ads on this job board is associated with the explicit gender labels employers attach to ads. To express this result more concretely, it may help to re-state the thought experiment underlying it, which imagines that there are only three types of jobs on XMRC (*F, N* and *M*) in this economy. The three job types differ in their tendency to attract female applicants (α) and in employers' propensity to select women from applicant pools; within each job type all jobs are identical. How much gender segregation would there be, compared to what we actually see? The answer is 60 percent. The remaining 40 percent (much of it within the *N* jobs) is not associated with employers' explicit requests, and is presumably similar to the type of segregation that prevails in countries that do not practice explicit gender profiling in jobs.

Turning to the contributions of compliance and enforcement to the 'label-linked' segregation identified above, B and C show that essentially all of the label-linked job segregation is due to self-sorting: allowing only α to differ across the three categories leads to a level of noise-adjusted segregation that is 57.8 percent of the actual level, while allowing only θ to differ generates only 6.6 percent of actual noise-adjusted segregation. Thus, in the analogy to residential segregation, home seekers' (job seekers') compliance with the designations of three neighborhood (job) types accounts for 57.8 percent of the census-block level (job-level) segregation in the city (labor market). These results mirror the dominant role of self-selection in accounting for gender-matching (*G*) which we have already identified.

Finally, counterfactuals D and E abstract completely from the gender labels attached to job ads and simply ask what share of noise-adjusted sex segregation in the successful applicant pools across individual job ads is associated with men's and women's differential application patterns, versus their differential success rates conditional on applying. Together these two counterfactuals show that *self-sorting* (both directed and undirected) *accounts for 97 percent of all the systematic gender segregation* across jobs in our data.[27]

[26] The distributions of *S* values across these counterfactual simulations are also highly concentrated, similar to the baseline, 'noise-only' simulation.

[27] Enforcement alone –without any self-selection—can account for as much as 18 percent. The enforcement and compliance shares now add up to more than 100 percent because (in contrast to Table 4) these counterfactuals

The preceding methods for computing actual and counterfactual noise-adjusted segregation indices across jobs can also be applied to segregation across other labor market units, including occupations, firms, and occupation*firm cells. The results of these calculations are summarized in Table 6. Notice first that --as one might expect-- the impact of noise-adjustment on the estimated level of segregation diminishes as the unit size increases (from jobs through occupations). In fact it is minimal in the case of occupations, where the unit size is the largest, but it plays non-negligible roles in the remaining cases. At $\tilde{S}$= .561, Table 6 shows that gender segregation is almost as high across firm*occupation cells as across individual job ads, and that explicit gender profiling accounts for 59 percent of that segregation. Segregation across firms and occupations is lower, though it is interesting to note that $\tilde{S}$ is slightly higher across the 36 occupation categories on the XMRC website than across the much larger number of firms in our sample. This is consistent with a long literature documenting the importance of occupational sex segregation and with DKS's finding that a large share of the variance in employers' gender requests occurs *within* firms. Column 4 of Table 6 shows that explicit gender profiling accounts for 59 percent of the gender segregation across firms and for 53 percent of the gender segregation across occupations.

A final perspective on the contribution of explicit gender labels to gender segregation examines the amount of gender segregation that is present within the 58 percent of our job ads that are not explicitly gendered: If explicit labels are epiphenomena that do not affect the allocation of workers to firms, we might expect to see just as much segregation within the nongendered ads as in the entire sample. Perhaps employers' gender preferences and workers' job preferences are just as highly gendered even when –as in the United States-- no public gender labels are attached to the jobs. We perform these calculations in Appendix Table A4, and find much less gender segregation within sample of nongendered ads than in our sample overall. Compared to .607 overall, noise-adjusted gender segregation within the sample of nongendered jobs is only .417. Occupational segregation by gender is .446 overall, but only .300 in non-gendered jobs. These figures suggest, but do not prove, that gendered job ads have real effects on the allocation of labor in China.

Summarizing our descriptive analysis of gender segregation in callback pools, we find that jobs on XMRC are highly gender-segregated, with a noise-adjusted Duncan and Duncan segregation index of .607. In other words, 60.7 percent of either men or women would have to change jobs to equate the gender ratio across all jobs. Explicit gender profiling in turn accounts for 59 percent of that job-level segregation. Gender profiling also accounts for 59 percent of

add enforcement and compliance, in turn, into a baseline scenario where neither is present, rather than subtracting them from a scenario where both are present. Analogous to adding regressors to an equation sequentially, the share explained by the first factor considered is larger in the absence than in the presence of a control for the second.

the gender segregation across firms, and for 53 percent of the gender segregation across occupations. Finally, the vast majority of the association of gender profiling with all the outcomes studied in this Section is through workers' compliance with firms' advertised requests when deciding where to send their resumes, rather than through active denials of callbacks to gender-mismatched applicants.

3.3 Actual and Counterfactual Gender Wage Gaps

With some minor modifications, the methods developed in Sections 3.1 and 3.2 can be adapted to measure the association between gendered job ads and the gender wage gap. To that end, we begin by estimating the gender wage gap among successful job applicants on XMRC, then decompose that gap into portions associated with gendered job ads, and (within that portion) to compliance versus enforcement processes. While a measure of the gender gap in *advertised* wages is readily available from our sample of job ads, most existing estimates of gender wage gaps refer to the gap in wages earned by employed workers. To approximate this measure more closely, we use a worker-based approach. Essentially, we treat callbacks as job offers and measure a jobseeker's wage as the highest posted wage he or she was offered, regardless of the type of job (*F*, *N* or *M*) it was from.[28] Calculating mean wages this way yields a gender wage gap, ω, of 0.146 log points.[29]

How much of this gender wage gap is associated with explicitly gendered job ads? To answer this question, we create the same 1000 simulated, counterfactual assignments of applications to callbacks we used in Section 3.2: recall that these simulations assumed that all assignments were random except for the fact that there were three types of jobs (*F*, *N* or *M*), each with its own gender mix of applicants, α, and its own relative propensity to pick women from the applicant pool, θ. However, instead of calculating a segregation index for each assignment, we calculate an economy-wide gender wage gap, ω by assigning each callback the mean wage for that *job type*, $\overline{w}^F, \overline{w}^N,$ or $\overline{w}^M$. Finally, the gender wage gap that is associated with the three job types is just the mean gap across these 1000 iterations. Re-doing these simulations while allowing *only* α or θ to differ across the three job types shows the relative contribution of compliance and enforcement to that gap.

The results of these calculations are reported in Table 7. Overall, the log wage gap implied by the three explicit job types equals .089. In other words, the fact that the three explicit job types –each of which has its own α and θ-- pay different mean wages can account for 60.8 percent of the overall gender gap in wages. As in our analysis of gender matching and

---

[28] Thus, for workers who did not receive any callbacks, or whose callbacks did not post any wages, we do not observe a wage.

[29] Interestingly, this turns out to be very similar to the gender gap *advertised* wages on XMRC, of 0.172 log points. Both gaps are, however, considerably smaller than the.280 log point gender wage gap reported in Table A2 for the city of Xiamen as a whole. Recall that our XMRC sample is much younger than Xiamen's employed workforce; gender wage gaps tend to be much smaller early in workers' careers.

gender segregation, decomposing this 'explained' portion of the gender wage gap into compliance and enforcement effects attributes the vast majority to compliance, for the simple reason that the types of callbacks workers get are determined almost exclusively by where they choose to apply.

4. Regression Analysis—Compliance

Our analysis so far paints the first statistical portrait of how gendered job ads enter into the recruitment process. As a descriptive exercise, however, it does not identify a causal effect of job profiling on either workers' application behavior or employers' selection behavior. In this Section, we focus on application behavior --compliance-- and define a causal effect of gender profiling on workers' application decisions as the outcome of the following thought experiment: Imagine that the observed patterns of job profiling and application decisions in our data constitute a labor market equilibrium in the sense that employers' advertising and selection decisions are optimal given workers' application behavior, and vice versa. In this equilibrium, row 2 of Table 3 indicates that *F*, *N* and *M* job ads attract applicant pools that are 92.6, 44.7 and 7.9 percent female respectively. Now suppose we exogenously switch the explicit gender label attached to just one of the many *N* jobs to *F* or to *M*, keeping everything else –including the labels on all the other jobs-- unchanged.[30] What will happen to the share of applicants to that job that are female? If this share does not change, then the large differences in the gender mix of the three job types in Table 3 are not causal, in the sense that the gender labels do not actually direct workers' application decisions. Instead, the labels may simply be standing in for other features of the job (such as the occupation) that tend to attract applicants of different genders.

Accordingly, our econometric attempts to isolate a causal effect of gender labels on application behavior will focus on controlling as tightly as possible for other characteristics of jobs (or job ads) that might also explain why different ads attract different mixes of men and women. We take two complementary approaches. In the first, in addition to a detailed list of skill requirements and other desiderata in in the job ad, we control for firm fixed effects and job title fixed effects. Job titles are the main heading in every job ad. They provide a brief description of the job and can run up to 18 words in XMRC. For example, here is a random sample of ten (translated) job titles on the XMRC website: front desk administration assistant, project engineer, quality control, shift leader, customer service maintenance specialist,

[30] Because it fixes the application and callback rates in all other jobs, note that this thought experiment implicitly holds fixed the *beliefs* of prospective applicants about their callback chances in the three job types (*F*, *N* and *M*). In essence, if attaching a male job label on an XMRC job signals to female applicants that their chances of getting a callback are X percent lower than they would have been otherwise, our regression estimates incorporate the effects of those applicant beliefs on women's application strategies.

administration, ME product engineer, experienced two-dimension designer, customer service engineer, and front desk clerk. Job titles provide considerably more relevant information about the type of work than even the most granular standardized occupational classification systems. For example, Marinescu and Wolthoff (2015) found that job titles on Careerbuilder.com were much more predictive of advertised wages than 6-digit SOC codes, and were essential controls for identifying the effect of advertised wages on the number and quality of applications an ad received. Thus, in this approach we will be comparing two observationally identical ads for a very narrowly defined type of work, holding constant the identity of the firm advertising the job.

In our second approach, we replace the job title fixed effects in the above analysis by indicators of the predicted, or *implicit* 'maleness' or 'femaleness' of the job derived from a machine learning analysis of the words in the titles. Essentially, we use the words in the title to predict whether a person reading it can infer whether the job is likely to request men, or to request women. While these two predicted probabilities (*Mp* and *Fp*, respectively) absorb less variation in job characteristics than the full set of title fixed effects, they provide a simple structure that helps us see precisely where –i.e. in which types of jobs—inserting a gender label into a job ad has the largest impact on application behavior. Notably, in both approaches, we use the entire sample of job ads available to us, not just the subset for which callback behavior is observed. To check for robustness, we replicated both analyses for the 'callbacks' subsample with very similar results.

4.1 Approach 1: Job Title Fixed Effects

As noted, here we run regressions in our entire sample of 42,744 ads, where the dependent variable is the share of applications that are female (α). The regressors of interest are the labels attached to the ad (*F*, *N* or *M*). In more detail, we estimate:

$$\alpha_j = a + bF_j + cM_j + dX_j + e_j \quad (6)$$

where *j* indexes jobs (ads), *F* (*M*) is a dummy for whether the job requests women (men) and *N* is the omitted job type. In column 1, we include no controls (X_j). Column 2 adds controls for the following job characteristics: requested education, experience, and age; the advertised wage; a dummy for whether a new graduate is requested; the number of positions advertised; plus dummies for missing education, age, wage and number of positions. Columns 3-5 in turn add occupation, job title and firm fixed effects, and column 6 interacts these job title and firm fixed effects. Thus, column 6 compares applicant pools across ads posted by the same firm for the same detailed job title, but with different gender requests. The extent to which the *b* and *c* coefficients attenuate as we add these controls captures the extent to which explicit gender labels are correlated with other features of job ads (such as a typically male occupation or job title) that allow applicants to infer the ad's desired gender even in the absence of an explicit gender request.

Table 8 shows that, as expected, the unadjusted effects of both the *M* and *F* job labels attenuate substantially --from -35 to -15 percentage points for *M* labels and 50 to 25 percentage points for *F* jobs-- as we add detailed controls for job characteristics, including interacted firm and job title fixed effects. Thus, a substantial share of the correlation between jobs' gender labels and the gender mix of applicants reflects the fact that men and women apply to different types of jobs, regardless of whether those jobs are explicitly targeted at their gender. Still, the estimated effects of the gender labels remain economically large and highly statistically significant, even when comparing the same job title in the same firm with different gender labels attached. [31] This suggests that an explicit gender request in a job ad may have substantial causal effects on the gender mix of applications it will receive. In other words, employers' gender requests appear to direct workers' applications.[32]

4.2 Approach 2-- Implicit Maleness and Femaleness

To better understand the source of the apparent compliance effect identified in Table 8, we now try to identify the types of jobs in which making an explicit gender request has the largest effects on application mix. If prospective applicants are using gender labels and other features of the job ad to predict whether a person of their gender would have a good chance of receiving a callback, we would expect explicit requests to have the largest impact on applications *in jobs where it is difficult for workers to infer the employer's gender preferences from the other contents of the ad.* To formalize this notion, we now replace the job title fixed effects in Table 8 by predicted probabilities that the job requests men (women), calculated from the words that appear in the title. Treating each ad's job title as a document, we calculate the implicit maleness and femaleness of each job using the Bernoulli naïve Bayes classifier (McCallum and Nigam 1998) for document classification; classifiers of this type are widely used in predicting whether a document is of a given type, for example a spam email.

---

[31] As column 6 indicates, our estimates with title*firm fixed effects are identified by 1,448 job ads for which a firm*job title cell made different gender requests at different times. Importantly, these estimates are not driven by a single large firm, job title or title*firm cell: these 1,448 ads represent 416 distinct job titles posted by 505 different firms, and comprise 686 title*firm cells. Histograms of estimates that leave out one job title at a time in Figure A2 are highly concentrated: all estimates of the request-male effect are between -.155 and -.136 and statistically significant (p<.01). All estimates of the request-female effect are between .238 and .265 and highly statistically significant (p<.01).

[32] Table 8 is replicated for the subsample of jobs for which we have callback information in Table A5, with similar results: Here, the most saturated specification we can estimate replicates column 5, where firm and job title fixed effects are entered separately. In that specification, the estimated effects of male and female labels are -.103 and -.240 respectively; both are highly statistically significant.

Briefly –exact details are available in Appendix 6-- for each word, *w*, that appears in our entire set of job titles, we first estimate the probability of observing that word in the title of a job that requests men, Prob (observe word *w* | job requests men) using empirical frequencies. Next, treating job titles as 'baskets of words' which appear independently, we can compute the probabilities of observing a given job title, *k*, given the job requests men, Prob (observe title *k* | job requests men) from its constituent words. Finally, using Bayes formula plus an assumption about workers' prior beliefs, we can compute the predicted maleness of each job title based on the words it contains.[33] Using the same procedure to predict each title's femaleness yields the two continuous variables,

$$M_p \equiv Prob(job\ explicitly\ requests\ men|\ job\ title\ k) \qquad (7)$$

$$F_p \equiv Prob(job\ explicitly\ requests\ women|\ job\ title\ k) \qquad (8)$$

which we use in our empirical analysis to represent the information contained in the job title about whether the job is likely to request men or women. Overall, *Mp* and *Fp* are quite predictive of employers' actual requests, with correlations of .411 and .402 with actual requests for men and women respectively. As we might expect, they identify what we might think of as stereotypically male and female jobs: the five 'most female' job titles (starting with the highest) are "front office desk staff", "administration office staff", "office staff", "cashier" and "administration assistant". The five 'most male' are "driver", "technician", "warehouse managing staff", "warehouse manager, and "production manager".[34] These indices of implicit maleness or femaleness allow us to estimate the effect on application behavior of adding an explicit gender request to jobs that 'look the same' in terms of an employer's likely gender preference, and to see in which types of jobs the effect of explicit requests on application behavior is the greatest.

More specifically, we now regress the female share of applicants to a job, α_j, on employers' explicit gender requests (*F* and *M*), plus all the control variables used in column 5 of Table 8 (other than the job title fixed effects) plus quartics in the implicit maleness or femaleness of the job that workers could infer from the job's title (*Fp* and *Mp*). In addition, each of these quartics is interacted with the three explicit job types, *F*, *N* and *M*. These interactions allow, for example, the effect of an explicit request for women to be either stronger or weaker in jobs that are stereotypically male (based on the words that appear in the job title) than in jobs whose titles do not convey an obvious gender preference.

[33] We adopt the naïve prior that the unconditional chances a job requests men equals 50 percent. This simplifies the computations and reflects the idea that individual jobseekers may not have access to good summary statistics on the share of jobs of different types available to them.

[34] Additional examples of job titles at different levels of *Fp* and *Mp* are provided in Appendix Table A6.

Predicted male and female applicant shares from these regressions are shown in Figure 1. Part (a) of the Figure shows the predicted female applicant share as a function of the predicted femaleness of the job based on the words in the job title, separately for the three types of jobs (*F*, *N*, and *M*). Predicted maleness is held fixed at its mean. Part (b) is the parallel figure for male applicant shares as a function of perceived maleness, holding predicted femaleness at its mean. Finally, part (c) shows the effects of encountering a request for a particular gender (relative to a non-gendered job) on the share of that gender in the applicant pool, with 95 percent confidence bands. These are the distances between the top two curves in parts (a) and (b).

Figure 1 shows, first of all, that explicit requests for male and female applicants have stronger effects on the gender mix of applications when the words in the job title do not send strong signals about whether the employer is likely to prefer men or women (i.e. when *Mp* and *Fp* are low). For example, when *Fp* is near zero, the predicted effect on the female applicant share of inserting an explicit request for women into an *N* job is about 53 percentage points. This effect diminishes to about 26 percentage points when *Fp* equals 0.7. A similar pattern is present for men, though it is less pronounced.

Second, there is a subtle but interesting gender difference regarding *when* explicit requests matter. In 'not-obviously-female' (low *Fp*) jobs, women comprise a relatively large share of applicants *only when* the job explicitly requests women. In 'not-obviously-male' (low *Mp*) jobs, men comprise a relatively large share of applicants both when men are explicitly requested, *and* when the job does not have a gender label. Together these patterns help us understand the much larger impact of *F* labels than *M* labels on the applicant mix in Table 8. Essentially, the main gender difference in application behavior occurs in jobs that –based on their title—are neither stereotypically male nor female. If we think of applying for jobs as entering a competition to get hired, these patterns are evocative of well-known gender differences in entry into competition (Niederle and Vesterlund 2007), and of less-well known gender gaps in the propensity to apply for jobs in the presence of ambiguity (Gee, forthcoming).[35]

We conclude our discussion of compliance effects with a reminder that our substantial estimated effects are consistent with at least two underlying mechanisms. One is that job labels communicate information about a worker's chances of getting a callback; in this view, women avoid male jobs because they know they have a much lower chance of getting those jobs if they apply. The second mechanism is that –much like labels on men's and women's

[35] To probe robustness to functional form, Figure A3 forces predicted applicant shares to be between zero and one by changing the dependent variable from α (the female share) to log ($\alpha/(1-\alpha)$), and replacing the quartics in *Fp* and *Mp* by linear terms (still interacted with *F*, *N* and *M*). In both cases, our main conclusions --including the larger effects of *F* labels than *M* labels on applicant mix-- continue to hold.

clothing—job labels communicate information about whether the worker is likely to want the job, without conveying any reluctance by the firm to transact with the worker. In this mechanism, women avoid male jobs because women dislike certain job attributes –perhaps competitive pay policies, long and inflexible hours, or even the absence of female co-workers—associated with those jobs. Assessing the relative importance of these two mechanisms requires an analysis of how gender-mismatched applications are treated when they are made, which is our goal in the next Section.

5. Regression Analysis—Enforcement

As noted, our goal in this Section is to estimate the effect of exogenously re-directing a single worker's *application* from a non-gendered (*N*) job to an identical gender-mismatched job, i.e. to a job that explicitly requests the opposite gender from the worker's. If little or nothing happens to the worker's chances of receiving a callback, then employers' advertised gender preferences are 'soft' preferences, in the sense that job labels may convey information about workplace characteristics that men and women may evaluate differently, but gender-mismatched applicants are evaluated on the same basis as the other applications that arrive. If instead there is a large gender *mismatch penalty*, gendered job labels are hard requirements imposed by employers.

To set the stage for our analysis, we begin by considering the naïve estimates of men's and women's mismatch penalties that emerge directly from our descriptive analysis. According to Table 3, women's apparent callback penalty from applying to an *N* versus an *M* job is 8.7 - 4.3 = 4.4 percentage points, or a 51 percent reduction in the chances of getting a callback. Similarly, men's penalty equals 9.0 – 5.8 = 3.2 percentage points, or a 36 percent reduction. Together, these numbers suggest that –while hardly absolute—employers' enforcement of their advertised gender requests is moderately 'hard'. Our main goal in this Section is to empirically distinguish two distinct scenarios that could account for these numbers. The difference between the scenarios hinges on the types of people who choose to apply to gender-mismatched jobs.

Consider scenario 1, in which selection into gender mismatch is positive. Here, applications that are made to gender-mismatched jobs are, on average, better qualified, and better matched (on dimensions other than gender) than other applications. This makes sense, for example, if the women who choose to apply to jobs requesting men are better qualified on dimensions like education, experience, and unobserved ability that the applicants hope will compensate for being of the 'wrong' gender. Related, when applying to explicitly male jobs, women may restrict their attention to the jobs fit their qualifications most closely. In this scenario, Table 3's raw mismatch penalties will underestimate the adverse effects of gender mismatch on the callback rate (because the people who cross-apply are better-qualified and better matched than those who do not). Adding controls for worker qualifications and job-worker match should increase the magnitude of the estimated penalty towards its true, larger value.

Now consider scenario 2, where selection is negative. Here, applications that are made to gender-mismatched jobs are, on average, less qualified, and worse matched (on dimensions other than gender) than other applications. This makes sense, for example, if the women who apply to jobs requesting men are primarily people who apply to jobs indiscriminately, for example because they have low application costs, are highly motivated to find a job, or are simply careless. In all of these cases, applicants who ignore explicit gender requirements might also ignore education, experience and other important requirements, so their applications are more poorly matched on average. In this case, Table 3's 4.4 percentage point mismatch penalty for women will overestimate the callback penalty associated with applying to an *M* job for a woman of fixed ability. Adding controls for worker qualifications and job-worker match should attenuate the magnitude of the estimated penalty towards its lower, true value. Indeed, if negative selection is strong enough, the true mismatch penalty could be zero. Here –since our best estimates of worker compliance are large—the most natural interpretation of our data would be one where advertised gender labels communicate job attributes that men and women care differently about, as in the example of men's and women's clothing.

To distinguish these two scenarios empirically, we run linear probability regressions in a sample of applications, where the dependent variable is an indicator for whether the worker received a callback. In doing so, we try to control as tightly as possible for other aspects of match and worker quality that might affect callback rates. Of particular note, we control for unobserved worker ability by using worker fixed effects—i.e. we will compare the callback rates of the same worker who sends her resume to two observationally-identical jobs that differ only in their explicit gender label. We control for the detailed type of work using job title fixed effects. To account for the fact that people who apply to jobs requesting the 'other' gender might be better- or worse matched to the job on dimensions other than gender, we also include detailed controls for matching on a variety of characteristics.

In more detail, we estimate the following linear probability model:

$$Callback_i = \alpha + \beta_1 FtoF_i + \beta_2 FtoM_i + \beta_3 MtoF_i + \beta_4 MtoM_i + \delta Mworker_i + \varphi X_i + \varepsilon_i \quad (9)$$

where *i* indexes *applications*. Of the six possible application types, women applying to nongendered jobs (*FtoN*) is the omitted type. In this specification, β_1 and β_2 give the effect on women of applying to *M* and *F* jobs (relative to nongendered jobs), while β_3 and β_4 give the effect on men of applying to *M* and *F* jobs (again, relative to nongendered jobs). The parameter

δ gives the callback gap between men and women applying to nongendered jobs. Our main focus will be on the *gender mismatch penalties* associated with applying to a job that is targeted at the 'other' gender, β_2 and β_3.

Column 1 of Table 9 estimates equation 9 without controls, replicating the unadjusted gaps in Table 3. Column 2 adds controls for the job's requested level of education, experience and age; the advertised wage; and an indicator for whether a new graduate is requested. Also included are indicators of the match between the applicant's characteristics and those requirements, including indicators for whether the applicant's education, age and experience are below or above the requested level, the match between the advertised wage and the applicant's current or previous wage, and the match between requested and actual new-graduate status. Column 3 adds controls for the following worker (CV) characteristics: whether he/she attended a technical school; the applicant's *zhicheng* rank; whether an English CV is available; the number of schools attended, experience spells and certifications reported.[36] Indicators for applicant height, myopia and marital status are also included, all interacted with the applicant's gender.[37]

Column 4 adds fixed effects of the occupation of the advertised job, using XMRC's occupational categories. Column 5 adds job title fixed effects plus two indicators of the amount of competition for the job: the number of positions advertised and the number of persons who applied to the ad.[38] Column 6 adds a full set of worker fixed effects; in this case, the effects of fixed applicant characteristics ("detailed cv controls" and the main gender effect) are not identified. Interactions between applicant gender and job type, which are our main coefficients of interest, however, remain identified. In effect, column 6 compares the outcomes of *the same worker* who has applied to observationally identical jobs that differ only according to the gender label (*F*, *N* or *M*) attached to the job, while allowing for this effect to differ according to the applicant's gender.

Before discussing our main coefficients of interest, it is worth noting that whenever they are statistically significant, observable indicators of the match between worker qualifications and job requirements are of the expected signs in Table 9: workers who have less education or experience than requested, or are older than requested are less likely to be called back. Finally,

---

[36] *Zhicheng* is a nationally-recognized worker certification system that assigns an official rank (from one through six) to workers in almost every occupation. Ranks are based on education, experience and in some cases nationwide or province-wide exams.

[37] These 'detailed CV controls' introduced in column 3 are not requested in job ads very often, so it is not practical to construct variables summarizing their match with the job's requirements.

[38] These 'queue length' or 'submarket tightness' controls account for the possibility that overall competition for callbacks might be systematically stiffer in some job types than others. For example, callback rates in jobs that request women might be lower for *all* applicants if women 'crowd into' those jobs more than men crowd into jobs that request men (Sorensen 1990).

the job competition controls (not shown) are always highly statistically significant, verifying that the ratio of applicants to open positions in any ad has strong effects on the chances of being called back. Also of some interest, workers with *more* education than the job requests also experience a statistically significant callback penalty in all specifications but one (Shen and Kuhn 2013).

Turning to the mismatch penalties, both men's and women's penalties attenuate somewhat as we add covariates in Table 9. This consistent pattern supports scenario 2 – in which gender mismatched applicants are negatively selected and Table 3's raw mismatch penalty overstates the true penalty—over scenario 1 in which gender-mismatched applicants are positively selected. That said, the mismatch penalty remains both economically and statistically significant in the presence of worker fixed effects (column 6). For a woman, applying to a job requesting men reduces her callback chances by 3.8 percentage points, only a little less than the unadjusted effect (4.4 percentage points). For men, the attenuation is more pronounced –from 3.2 to 2.3 percentage points—suggesting a greater amount of negative self-selection into gender-mismatched applications among men.

In sum, our preferred estimates in Table 9 (column 6) imply that both men and women face substantial callback penalties when they apply to jobs that request the 'other' gender. These large and significant callback penalties point away from the extreme 'clothing labels' version of scenario 2 –where women's reluctance to apply to explicitly male jobs is due purely to the applicants' own tastes—as the best representation of our data. In addition, there is some gender asymmetry in our estimated callback penalties, with women's penalty for applying to explicitly male jobs (3.8 percentage points, or 44 percent) exceeding men's when applying to female jobs (2.3 percentage points, or 26 percent). This gender difference is highly statistically significant.

Two potential concerns with the above estimates are the possibility of gender misclassification and the effects of luck in the application process. Concerning gender misclassification, if some workers' genders are miscoded in their XMRC profiles we would expect to see the misclassified workers to apply to an unusually large number of apparently gender-mismatched jobs. To see if this could be driving our results, in Appendix 5 we exclude from the estimations the very small number of workers who direct more than half of their applications to opposite-gender jobs, with very little change in the results.[39]

[39] Miscoding of the *requested* gender is not a concern since our data are the exact record of requested gender that workers observe on the job board when deciding where to apply. See Appendix 5 for additional discussion of how gender is coded on the job board and on how we construct our "gender misclassification-robust" subsample of applications.

Concerning luck, our results could overstate employers' openness to gender-mismatched applicants if a significant number of wrong-gendered applicants are being called back only because no candidates of the preferred gender applied to the job (Lang, Manove and Dickens 2005; Lazear, Shaw and Stanton 2018). While our job competition controls capture some of these effects, a more direct test is to look directly at applicant pools containing zero applicants of the requested gender. As it happens, none of the 666 male jobs in our dataset received zero male applicants. We did find five female jobs that received no female applicants, and these jobs did call back some men. However, these jobs constitute less than 0.6 percent of the 867 female jobs in our sample.

We conclude this Section with two important caveats regarding the interpretation of our enforcement estimates. The first is that the substantial and robust mismatch penalties in callback rates we estimate here do not in themselves constitute evidence for any particular form of discrimination, such as taste-based or statistical discrimination. Indeed, mismatch penalties are consistent with a number of underlying processes, including gender differences in productivity (both real and imagined), and the tastes of employers, recruiters, co-workers and customers, with the important proviso that any such productivity or taste differences must be highly *job-specific* to explain the patterns in our data. To distinguish among these possible sources of mismatch penalties, research needs to examine the precise types of jobs in which they occur. For example, to assess the role of job-specific productivity differences one could look at tasks where there is established evidence of gender differentials in performance (Baker and Cornelson 2016, Cook et al, 2017). Customer tastes could be isolated by looking at jobs involving customer contact, and at whether gender requests are accompanied by requests for a photograph or for applicant beauty. More detailed parsing of the words in job titles can also provide useful clues. Indeed, DKS (forthcoming) parse job titles in their study of employers' *decisions to post* gendered job ads, finding some support for a customer-tastes explanation of a significant share of these posts. Specifically, they find a large group of ads requesting young, attractive women in customer-contact jobs.

A second caveat concerns treatment effect heterogeneity. Specifically, while we have a number of controls for the quality of the match between the worker and the job, it is important to remember that our estimates still represent *treatment-on-the-treated* effects on the sample of applications people choose to make to gender-mismatched jobs. If –as we might expect-- workers disproportionately apply to the gender-mismatched jobs where they know their personal gender mismatch penalty (i.e. their treatment effect) is small, our estimates in Table 9 would underestimate the callback penalty associated with a randomly-selected gender-mismatched job. Of course, if selection on treatment effects is negative (such that the workers who apply to gender-mismatched jobs are the most harmed by doing so), the mismatch penalty for a randomly-selected job would be smaller than the one we estimate here.

6. Summary and Discussion

We believe that this is the first paper to study the (proximate) *consequences* for workers and firms of a common practice in the world's labor markets: job advertisements that request workers of a specific gender. Using internal application and callback information from a Chinese Internet job board, we find a high degree of correspondence between the gender employers request in a job ad and actual outcomes of the recruiting process: nearly 95 percent of successful applications (callbacks) to explicitly gendered jobs are of the requested gender. We then partition this high level of gender-matching into two components—the propensity of workers to apply where they are requested (compliance) or employers' active rejection of gender-mismatched applicants (enforcement). Of these, compliance, or worker self-selection, plays the dominant role. Gendered job ads also account for a substantial share of the gender segregation observed in our data across jobs, firms and occupations, mostly through worker self-selection as well. Intuitively, since so few workers apply to gender-mismatched jobs, total gender segregation would change very little in a counterfactual world where employers ignored gender in all their callback decisions, but application patterns were held constant. In addition, gendered job ads, which comprise 40 percent of the ads in our data, can account for 60 percent of the gender wage gap in the data.

In addition to painting this statistical portrait of how gendered job ads enter the recruitment process, we have attempted to answer two causal questions: In this labor market, what would be the effect on an employer's application pool of adding an explicit request for male or female applicants to a job ad, with no other changes to the ad? And what would be the consequences for a worker's chances of getting a callback, of redirecting his (her) application from a nongendered job ad to an identical ad that requested the opposite gender? Using firm and job title fixed effects in the former case, and worker and job title fixed effects in the latter, we find that these consequences appear to be substantial: explicit gender requests are not just superfluous information that can be inferred from other contents of the job ad. Instead, gendered job ads appear to direct workers' application decisions *and* to predict how employers will treat applicants of the 'wrong' gender.

While we believe that our analysis has increased our understanding of the role of gendered job ads in the recruitment process, an important caution is that our results are not directly informative about what would happen if a country successfully banned such ads, as the United States and Austria did in 1974 and 2004 respectively.[40] One reason is the fact that employers might respond to a ban by crafting other, legal signals that direct the same workers

[40] In 1973, gendered job ads were prohibited by the U.S. Supreme Court. (*Pittsburgh Press Co. v. Pittsburgh Commission on Human Relations et al*). In 2004, the Austrian government instituted a 360 Euro per-ad fine on gendered job ads as part of the Austrian Equal Treatment Act. See Walsh et al. (1975, chapter 5) for a fascinating study of gendered job ads in the United States prior to the 1973 prohibition.

to the previously-labeled jobs, resulting in little or no change to application and callback patterns.[41] A second reason relates to the endogeneity of workers' application decisions (the α's in our decompositions) and of employers' callback decisions (the θs) to a gendered-ad ban. To see this, imagine first --as seems likely-- that a ban increased the number of workers applying to (formerly) gender-mismatched jobs, because all jobs now appear equally 'open' to both men and women. If women's relative callback rates (θ) remained unchanged in all jobs after such a ban, the ban's primary first-order effect would be to increase labor market frictions (because it now will be harder for workers to avoid jobs where their gender is dispreferred). On the other hand, if employers' openness to (formerly) gender-mismatched applicants (the θs) also changes when a ban is introduced, a ban's effects on labor market frictions --and on other outcomes like gender segregation and gender wage gaps-- could conceivably go in either direction. Further research, perhaps drawing on the natural experiments associated with historical bans, would be needed to identify these more complex effects.

[41] For example, as discussed in Appendix 3, recent attempts to discourage gendered job ads in China have led some employers to use code words to avoid detection. In addition, job boards have made it easy to filter resumes by gender, both within the applicant pool and when a recruiter is searching through resumes posted on the site.

References

Aaronson, Daniel, Daniel Hartley, and Bhashkar Mazumder. 2017 "The Effects of the 1930s HOLC "Redlining" Maps. Chicago Federal Reserve working paper no. 2017-12

Albrecht, J., P. Gautier and S. Vroman. 2006. "Equilibrium directed search with multiple applications". *Review of Economic Studies* 73: 869—91

Angwin, Julia, Noam Scheiber and Ariana Tobindec. 20, 2017 "Facebook Job Ads Raise Concerns About Age Discrimination" *New York Times*, December 20, 2017

Banfi, Stefano, and Benjamin Villena-Roldan. Do High-Wage Jobs Attract more Applicants? Directed Search Evidence from the Online Labor Market *Journal of Labor Economics,* forthcoming.

Baker, Michael and Kirsten Cornelson (2016) "Gender Based Occupational Segregation and Sex Differences in Sensory, Motor and Spatial Aptitudes" NBER Working Paper No. 22248.

Belot, Michele, Philipp Kircher and Paul Muller (2017). "How Wage Announcements Affect Job Search Behaviour - A Field Experimental Investigation", European University Institute, working paper.

Bielby, W. T., and J. N. Baron, ''A Woman's Place Is with Other Women: Sex Segregation within Organizations,'' in Sex Segregation in the Workplace: Trends, Explanations, Remedies, B. F. Reskin, ed. (Washington, DC: National Academies Press, 1984).

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

Brenčič, Vera. 2010. Do Employers Respond to the Costs of Continued Search? *Oxford Bulletin of Economics and Statistics*, 72(2), 211-245.

Brenčič, Vera. 2012. Wage Posting: Evidence from Job Ads, *Canadian Journal of Economics*, 45(4), 1529-1559.

Brenčič, Vera and John B. Norris. 2009. Employers' Online Search: An Empirical Analysis, *Industrial Relations*, 48(4), 684-709.

Brenčič, Vera and John B. Norris. 2012. Employers' Online Recruitment and Screening Practices, *Economic Inquiry*, 50(1), 94-111.

Brenčič, Vera and John B. Norris. 2010. Do Employers Change Job Offers in Their Online Job Ads to Facilitate Search? Economics Letters, 108(1), 46-48.

Card, David, Ana Rute Cardoso, and Patrick Kline. forthcoming."Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women" *Quarterly Journal of Economics*.

Carrington, William J. and Kenneth R. Troske (1997) "On Measuring Segregation in Samples With Small Units" *Journal of Business & Economic Statistics*, 15:4, 402-409.

Cook, Cody, Rebecca Diamond, Jonathan Hall, John A. List, and Paul Oyer. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Uber Drivers " unpublished paper, Stanford University, December 20, 2017

Cutler, David M., Edward L. Glaeser and Jacob L. Vigdor. 1999. "The rise and decline of the American ghetto" *Journal of Political Economy* 107(3): 455-506

Delgado Helleseter, Miguel, Peter Kuhn and Kailing Shen (forthcoming) "Age and Gender Profiling in the Chinese and Mexican Labor Markets: Evidence from Four Job Boards", *Journal of Human Resources*.

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

Federal Housing Administration. 1938. *Underwriting Manual: Underwriting and Valuation Procedure Under Title II of the National Housing Act With Revisions to February, 1938* (Washington, D.C.), Part II, Section 9, Rating of Location. Available at: http://wbhsi.net/~wendyplotkin/DeedsWeb/fha38.html (visited May 10, 2016).

Flory, Jeffrey A., Andreas Leibbrandt, John A. List; Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions, *The Review of Economic Studies*, Volume 82, Issue 1, 1 January 2015, Pages 122–155.

Gee, Laura (2018) "The More You Know: Information Effects on Job Application Rates in a Large Field Experiment", *Management Science*. https://doi.org/10.1287/mnsc.2017.2994

Hirschbein, Brad and Kahn, Lisa. "Is College the New High School? Evidence from Vacancy Postings" unpublished paper, W. E. Upjohn Institute, 2015.

Human Rights Watch. 2018. *Only Men Need Apply: Gender Discrimination in Job Advertisements in China*

Ibanez, Marcela and Gerhard Reinter. "Sorting through Affirmative Action: Three Field Experiments in Colombia" *Journal of Labor Economics* 2018.

Kuhn, Peter and Kailing Shen. “Gender Discrimination in Job Ads: Evidence from China” *Quarterly Journal of Economics* 128(1) (February 2013a)

Kuhn, Peter and Kailing Shen. “Do Chinese Employers Avoid Hiring Overqualified Workers? Evidence from an Internet Job Board” *Research in Labor Economics* 37 (2013b): 1-30.

Kuhn, Peter and Kailing Shen. “Do Employers Prefer Migrant Workers? Evidence from a Chinese Job Board” *IZA Journal of Labor Economics* 4:22 (2015) DOI: 10.1186/s40172-015-0038-0

Kuhn, Peter. “Employer Investments in Employee Assessment” unpublished paper, UC Santa Barbara, March 2016.

Kudlyak, Marianna, Damba Lkhagvasuren and Roman Sysuyev. “Systematic Job Search: New Evidence from Individual Job Application Data”

Lang, Kevin, Michael Manove, and William T. Dickens. 2005. "Racial Discrimination in Labor Markets with Posted Wage Offers." *American Economic Review*, 95 (4): 1327-1340.

Lazear, Edward P., Kathryn L. Shaw, and Christopher T. Stanton. 2018. “Who Gets Hired? The Importance of Competition among Applicants” *Journal of Labor Economics* 36(S1).

Leibbrandt, A., & List, J. A. (2014). “Do women avoid salary negotiations? Evidence from a large-scale natural field experiment”. *Management Science*, 61(9), 2016-2024.

Logan, John R., Brian J. Stults and Reynolds Farley. “Segregation Of Minorities In The Metropolis: Two Decades Of Change” *Demography*, 41(1), February 2004: 1–22

Makinen, Julie. “China's women begin to confront blatant workplace bias” February 28, 2014. Available at: http://www.latimes.com/world/asia/la-fg-china-women-work-20140228-story.html#page=1. Visited April 8, 2015.

Marinescu, Ioana and Roland Rathelot “Mismatch Unemployment and the Geography of Job Search” unpublished paper, University of Chicago 2015. [pdf]

Marinescu, Ioana and Ronald Wolthoff “Opening the Black Box of the Matching Function: the Power of Words”, unpublished paper, Wharton, 2017.

Mas, Alexandre and Amanda Pallais. “Valuing Alternative Work Arrangements”. *American Economic Review*, Dec 2017

McCallum, Andrew and Kamal Nigam (1998). “A comparison of event models for Naive Bayes text classification” *AAAI-98 workshop on learning for text categorization.* 752

Modestino, Alicia Sasser, Daniel Shoag and Joshua Balance . "Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?" unpublished paper, Northeastern University.

Neumark, David Ian Burn and Patrick Button. "Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment" NBER Working Paper No. 21669, October 2015

Niederle, Muriel, and Lise Vesterlund. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122(3), 1067-1101. August 2007.

Pittsburgh Press Co. vs Pittsburgh Commission on Human Relations et al., U.S. Supreme Court decision argued March 20. 1973, decided on June 21, 1973.

Shen, Kailing and Peter Kuhn. "Do Chinese Employers Avoid Hiring Overqualified Workers? Evidence from an Internet Job Board" *Research in Labor Economics* 37 (2013): 1-30.

Sorensen, Elaine. "The Crowding Hypothesis and Comparable Worth" *Journal of Human Resources* Vol. 25, No. 1 (Winter, 1990), pp. 55-89

Walsh, John, Miriam Johnson and Marged Sugarman (1975) *Help Wanted: Case Studies of Classified Ads*. Salt Lake City, Utah: Olympus Research Corporation (funded by U.S. Department of Labor research contract no. 2141-73-28).

Zoloth, B.S. 1976. *Alternative Measures of School Segregation*. Madison: Institute for Research on Poverty, University of Wisconsin.

Tables and Figures

Table 1: Descriptive Statistics: Ad Sample

	Ad Requests Women (*F* jobs)	Gender not specified (*N* jobs)	Ad Requests Men (*M* jobs)	All Ads
Education specified?	0.961	0.899	0.925	0.919
Education Requested (years), if specified	12.70	12.31	11.25	12.21
Tech School Requested?	0.301	0.165	0.206	0.207
Desired Age Range specified?	0.638	0.390	0.566	0.481
Desired Age, if Requested (midpoint of interval)5r	25.91	28.81	29.47	28.03
Experience Requested (years)	0.785	0.997	1.215	0.987
New Graduate Requested?	0.069	0.023	0.030	0.035
Wage Advertised?	0.638	0.557	0.556	0.576
Wage, if advertised (yuan/month, midpoint of interval)	2,001	2,658	2,439	2,446
Number of positions specified?	0.964	0.923	0.971	0.941
Number of positions, if specified	1.915	2.249	2.033	2.130
Number of applicants	79.49	62.56	46.55	63.66
Sample Size	867	2,104	666	3,637

Table 2: Descriptive Statistics: Application Sample

	Applications from Women	Applications from Men	All Applications
Education (years)	14.56	14.11	14.35
Completed Tech School?	0.155	0.164	0.159
Age (years)	23.24	24.86	23.99
Experience (years)	2.674	3.886	3.230
New Graduate?	0.210	0.155	0.185
Current wage listed?	0.688	0.702	0.694
Current wage, if listed (yuan/month)	2,090	2,462	2,263
Married (if marital status listed)	0.140	0.215	0.174
Occupational Qualification (*Zhicheng*)[1]	1.086	1.403	1.231
Myopic	0.328	0.268	0.301
Height (cm)	160.6	171.5	165.6
English CV available?	0.145	0.104	0.126
Number of Schools listed	0.312	0.279	0.297
Number of Experience Spells	2.678	2.606	2.645
Number of Certifications	1.462	0.886	1.198
Sample Size	124,275	105,341	229,616

Notes:

1. *Zhicheng* is a nationally-recognized worker certification system that assigns an official rank (from one through six) to workers in almost every occupation. Ranks are based on education, experience and in some cases nationwide or province-wide exams.

Table 3: Application and Callback Patterns by Job Type

	Ad Requests Women (*F* jobs)	Gender not specified (*N* jobs)	Ad Requests Men (*M* jobs)	All Ads
1. Share of callbacks that are female (δ)	0.940	0.437	0.037	0.505
2. Share of applications that are female (α)	0.926	0.447	0.079	0.541
3. women's callback rate (f)	0.072	0.087	0.043	0.078
4. men's callback rate (m)	0.058	0.090	0.096	0.090
5. ratio of callback rates ($\theta = f/m$)	1.246	0.958	0.445	0.866
N (ads)	867	2,104	666	3,637
N (callbacks)	4,859	11,569	2,817	19,245
N (applications)	68,638	130,266	30,712	229,616

Table 4: Actual and Counterfactual Gender-Matching Rates

	Share of callbacks that are of the requested gender (g) (1)	Gender-matching index (G)[3] (2)
Baseline: Actual values	0.948	0.897
Counterfactual 1-- no compliance: Equal female share in applications, α, across all jobs[1]	0.617	0.232
Counterfactual 2:-- no enforcement: Equal female callback advantage (θ) in all jobs[2]	0.921	0.842

Notes:

1. Applies the population female applicant share (α) (.541) to all three job types.
2. Applies the population female risk ratio (θ) (.866) to all three job types.
3. $G = \frac{g - g_0}{1 - g_0}$ and g_0 = .501.

Table 5: Actual and Simulated Noise-Adjusted Segregation Indices across Jobs (Ads)

	Noise-Adjusted Segregation Index ($\tilde{S}$)	Share of noise-adjusted segregation explained ($\tilde{S}$ simulated/$\tilde{S}$ actual)
ACTUAL	**0.607**	**1.000**
SIMULATIONS:		
Effects of job categories (*F*, *N* and *M*) on segregation:		
A. Total effect of job categories: both α and θ vary across job categories	0.359	0.592
B. Effect of self-sorting across the three job categories: α varies across job categories, θ is the same in all ads	0.351	0.579
C. Effect of enforcement in the three job categories: θ varies across job categories, α is the same in all ads	0.040	0.066
Effects of applicant self-sorting and employer choice on segregation:		
D. Effect of self-sorting across all jobs: each job has its own α, all jobs have the same θ	0.588	0.969
E. Effect of employer choice within all jobs: each job has its own θ, all jobs have the same α	0.107	0.176

Table 6: Actual and Counterfactual Segregation across Job Titles, Occupations and Firms

Gender Segregation across:	Raw segregation index (S) (1)	Noise-adjusted segregation ($\tilde{S}$) (2)	Noise-adjusted segregation associated with job profiling (Counterfactual A) (3)	Share of noise-adjusted segregation associated with job profiling (3/2) (4)
Jobs (from Table 5)	0.732	0.607	0.359	0.592
Firm*Occupation cells	0.662	0.549	0.322	0.587
Firms	0.506	0.395	0.234	0.592
Occupations	0.405	0.385	0.204	0.531

Table 7: Actual and Simulated Gender Wage Gaps

	Gender Wage Gap (ω actual)	Share of gender wage gap explained (ω simulated/ ω actual)
ACTUAL	0.146	1.000
SIMULATED WAGE GAPS:		
A. Total effect of job categories: both α and θ vary across job categories	0.089	0.608
B. Effect of self-sorting across the three job categories: α varies across job categories, θ is the same in all ads	0.089	0.608
C. Effect of enforcement in the three job categories: θ varies across job categories, α is the same in all ads	0.011	0.076

Table 8: Effects of Employers' Gender Requests on the share of female applications received (α)

	(1)	(2)	(3)	(4)	(5)	(6)
Ad requests men (*M*)	-0.3547***	-0.3226***	-0.2459***	-0.1222***	-0.1203***	-0.1462***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.025)
Ad requests women (*F*)	0.4954***	0.4519***	0.3736***	0.2263***	0.2339***	0.2462***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.028)
Primary School		0.0247**	0.0095	-0.0019	-0.0057	-0.0292
		(0.011)	(0.009)	(0.005)	(0.006)	(0.026)
Middle School		-0.0627***	-0.0507***	0.0036	-0.0055	-0.0343
		(0.011)	(0.011)	(0.006)	(0.007)	(0.033)
Tech School		0.0673***	0.0477***	0.0004	-0.0014	-0.0415*
		(0.008)	(0.007)	(0.005)	(0.005)	(0.024)
Post-secondary		0.1159***	0.0639***	-0.0016	-0.0061	-0.0408
		(0.008)	(0.007)	(0.004)	(0.005)	(0.027)
University		0.1203***	0.0499***	-0.0137**	-0.0125*	-0.0189
		(0.010)	(0.008)	(0.006)	(0.007)	(0.045)
Number of positions advertised		-1.7400***	-0.9615***	-0.1220	-0.1338	-0.5756
		(0.164)	(0.121)	(0.124)	(0.130)	(0.571)
Occupation Fixed Effects			Yes	Yes	Yes	Yes
Job Title Fixed Effects				Yes	Yes	
Firm Fixed Effects					Yes	
Title*Firm Fixed Effects						Yes
N (ads)	42,744	42,744	42,744	42,744	42,744	42,744
"Effective" *N*	42,744	42,744	42,744	25,438	23,819	1,448
R^2	0.554	0.590	0.721	0.925	0.950	0.974

Standard errors in parentheses, clustered by firm. *** p<0.01, ** p<0.05, * p<0.1

Notes: in addition to the covariates shown, columns 2-5 also control for the following job ad characteristics: requested experience level (quadratic), requested age level (quadratic in midpoint of range), advertised wage (quadratic in midpoint of bin; 8 bins), dummy for whether new graduate requested, number of positions advertised, plus dummies for missing education, age, wage and number of positions. All regressions are weighted by the total number of applications received. 'Effective' *N* excludes job titles, firm IDs, and title*firm cells that only appear in one ad in columns 4, 5 and 6 respectively.

Table 9: Effects of Job Labels (*F*, *N* and *M*) on Callback Rates

	(1)	(2)	(3)	(4)	(5)	(6)
Female Worker * Female Job	-0.0149***	-0.0106***	-0.0103***	-0.0103***	-0.0147***	-0.0163***
	(0.009)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Female Worker * Male Job	-0.0440***	-0.0432***	-0.0431***	-0.0425***	-0.0333***	-0.0377***
	(0.013)	(0.004)	(0.004)	(0.004)	(0.006)	(0.008)
Male Worker * Female Job	-0.0328***	-0.0273***	-0.0274***	-0.0214***	-0.0229***	-0.0230***
	(0.010)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Male Worker * Male Job	0.0054***	0.0015	0.0016	0.0037*	-0.0064	-0.0164***
	(0.009)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)
Male Worker	0.0038**	0.0004	-0.0030	-0.0065***	-0.0172***	
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	
Education less than requested		-0.0054**	-0.0049*	-0.0076***	-0.0087***	-0.0115***
		(0.002)	(0.003)	(0.003)	(0.002)	(0.004)
Education more than requested		-0.0042***	-0.0070***	-0.0044**	0.0012	0.0061**
		(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Age less than requested		-0.0004	-0.0017	-0.0019	-0.0034*	-0.0019
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age more than requested		-0.0330***	-0.0309***	-0.0283***	-0.0203***	-0.0213***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Experience less than requested		-0.0063***	-0.0067***	-0.0081***	-0.0095***	-0.0071***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Experience more than requested		0.0005	0.0021	0.0014	-0.0011	0.0012
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Wage below advertised		-0.0009	-0.0007	-0.0017	0.0000	-0.0013
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Wage above advertised		0.0009	0.0007	0.0002	-0.0060***	-0.0046
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Detailed CV controls			Yes	Yes	Yes	
Occupation Fixed Effects				Yes	Yes	Yes
Competition Controls					Yes	Yes
Job Title Fixed Effects					Yes	Yes
Worker Fixed Effects						Yes
N (ads)	229,616	229,616	229,616	229,616	229,616	229,616
'Effective' *N*	229,616	229,616	229,616	229,616	229,590	192,681
R^2	0.001	0.005	0.005	0.016	0.198	0.387

Standard errors in parentheses, clustered by worker. *** p<0.01, ** p<0.05, * p<0.1

Notes to Table 9:

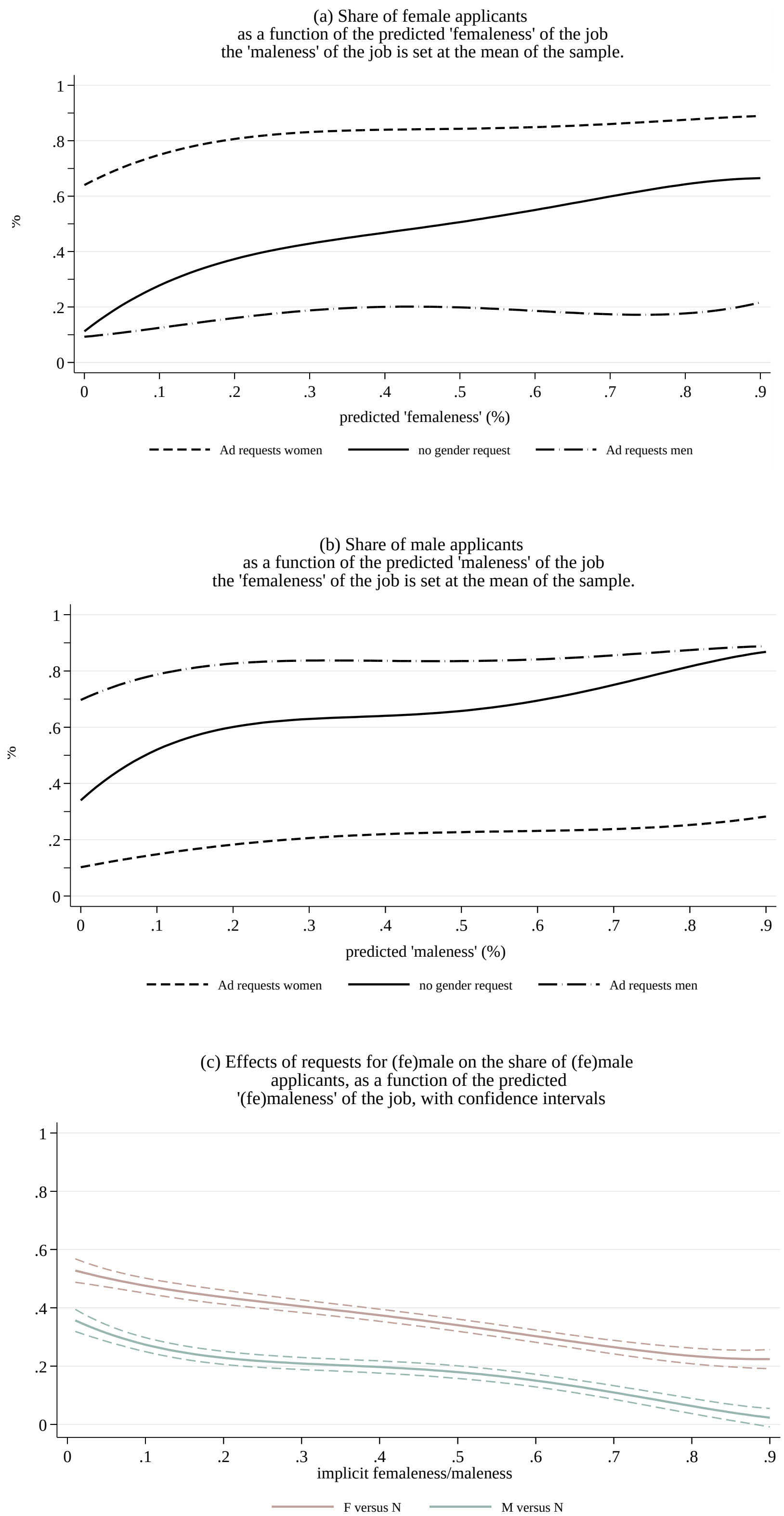
In addition to the covariates shown, columns 2-6 include the following controls for *ad characteristics*: requested education (5 categories), experience (quadratic), age (quadratic), the advertised wage (quadratic in midpoint of bin; 8 bins) and a dummy for whether a new graduate is requested. Columns 2-6 also include a dummy for whether the applicant's new graduate status matches the requested status, plus indicators for missing age and wage information for either the ad or the worker.

"Detailed CV controls" (used in columns 3-6) are an indicator for attending technical school; the applicant's *zhicheng* rank (6 categories); an English CV indicator; the number of schools attended, job experience spells and certifications reported; and the following characteristics interacted with gender: height, myopia, and marital status (interacted with applicant gender)

Occupation fixed effects control for the 36 categories used on the XMRC website.

'Effective' *N* excludes job titles-and worker IDs that only appear in one ad in columns 5 and 6 respectively.

Figure 1: Effects of Gender Requests (*F* and *M*) and Predicted Gender (*Fp* and *Mp*) on the Female Share of Applicants



Notes to Figure 1:

Figures represent predicted values of the female share of applicants (α) from a specification identical to column 5 in Table 8, where the job title fixed effects are replaced by quartics in *Fp* and *Mp*, each interacted with explicit job type (*F*, *N* and *M*). Predictions in part (a), which shows the effect of implicit femaleness (*Fp*), hold *Mp* at its mean. Predictions in part (b), which depicts the implicit maleness (*Mp*), hold *Fp* at its tenth mean. All other characteristics are set at their means. The regression is weighted by the number of applications to each ad, and standard errors are clustered at the occupation(firm) level.

Part (c) shows the predicted effects of attaching an explicit male (female) label to a job ad (relative to an *N* label) at different levels of implicit maleness (femaleness), with 95 percent confidence bands. Notably, both effects are larger in jobs whose title does not convey a clear preference for the applicant's gender. In addition, the effects of explicit requests for women on application behavior are significantly larger (both economically and statistically) than the effects of explicit requests for men.

Predictions for values of *Fp* or *Mp* greater than 0.9 are imprecise and not shown; only 2,462 ads have values in this range, comprising .0377 and .0330 of the sample respectively.