NBER WORKING PAPER SERIES

DO WORKERS VALUE FLEXIBLE JOBS? A FIELD EXPERIMENT

Haoran He
David Neumark
Qian Weng

Working Paper 25423
http://www.nber.org/papers/w25423

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2019

The experiment described in this paper was registered on the AEA RCT registry website (AEARCTR-0002645) in December 2017, prior to data collection. We are grateful to the job board on which we ran our experiment. We are grateful for helpful comments from Xuan Chen, Yan Chen, Daniel Houser, Peter Kuhn, Paul Oyer, Marie Claire Villeval, Fei Wang, Hui Xu, Xiangquan Zeng, participants at many conferences, and seminar participants at Beijing Jiaotong University, Central University of Finance and Economics, East China Normal University, Jinan University, Nanjing Audit University, Shanghai University of Finance and Economics, South China Normal University, and Southwestern University of Finance and Economics. We thank Shiwen Chen, Yiqiang Guo, Wanyu Jiang, Kelin Lu, Yue Yin, Zhuo Zhang, and especially Shuqing Luo, for research assistance. Financial support from the National Natural Science Foundation of China (71973016 and 71503255), the Beijing Natural Science Foundation (9192013), and Ministry of Education in China Project of Humanities and Social Sciences (18YJA790032) is gratefully acknowledged. Any remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Haoran He, David Neumark, and Qian Weng. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Do Workers Value Flexible Jobs? A Field Experiment
Haoran He, David Neumark, and Qian Weng
NBER Working Paper No. 25423
January 2019, Revised July 2020
JEL No. J01

ABSTRACT

We explore workers' valuation of job flexibility, using a field experiment conducted on a Chinese job board. Our experimental job ads differ randomly in offering jobs that are flexible regarding when one works (time flexibility) or where one works (place flexibility), and offering different salaries. Application rates are higher for flexible jobs, conditional on the salary offered, providing evidence that workers value job flexibility. Moreover, under some plausible conditions our evidence is informative about job seekers' willingness to pay for flexible jobs of the types offered in the experiment, and points to fairly high valuation of the most flexible jobs.

Haoran He
Beijing Normal University
No 19, Xinjiekouwai St.
Haidian District
100875 Beiing
China
haoran.he@bnu.edu.cn

Qian Weng
Renmin University of China
No. 59 Zhongguancun Street
Haidian District
Beijing 100872
China
qian.weng@ruc.edu.cn

David Neumark
Department of Economics
University of California, Irvine
3151 Social Science Plaza
Irvine, CA 92697
and NBER
dneumark@uci.edu

A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/2645

1. Introduction

Non-monetary job characteristics, in addition to earnings, have long been recognized as potential factors affecting workers' job choices. In this paper, we present evidence on how workers value job flexibility. One could study many job amenities, but job flexibility is of considerable interest. Employers may be interested in offering more flexible jobs to recruit workers or to reduce office costs, and governments may be interested in promoting job flexibility to increase labor supply and reduce congestion. These motivations presume that workers actually value flexible jobs. But do they?

We contribute field experiment evidence stemming from a collaboration with a Chinese online job board on which workers search for jobs and employers look for workers. We generate random variation in job flexibility offered to job seekers in invitations to apply for jobs offered by a single firm. Our job flexibility conditions vary with respect to when one works (time flexibility, i.e., the ability to schedule one's work within a workday) and where one works (place flexibility, i.e., the ability to telecommute). We also vary the pay of the jobs. We compare application rates across job flexibility conditions conditional on pay, and estimate workers' valuation of job flexibility by contrasting application rates across various job flexibility and pay combinations.

In our view, this field experiment approach improves upon the other two approaches used to estimate workers' valuation of job characteristics. Non-experimental studies try to infer these valuations by regressing wages on job characteristics of realized job choices. But estimates often appear to be incorrectly signed, possibly due to correlations between job amenities and either unmeasured worker heterogeneity (Brown, 1980;[1] Kniesner et al., 2012) or unmeasured job or firm characteristics (Woodcock, 2008), or to endogenous sorting of workers across jobs (e.g., Bonhomme and Jolivet, 2009; Sullivan and To, 2014).[2] Some recent research addresses these challenges by using

[1] Brown (1980) also focused on potential measurement error in job characteristics.

[2] For example, workers may vary in unmeasured productivity, and more productive workers may use their higher income to "buy" more pleasant job characteristics (assuming positive job amenities are a normal good). This would lead to a positive bias in the estimated effect of positive job amenities on pay, in contrast to the predicted negative effect of positive amenities on pay from the theory of compensating differentials. Online Appendix A shows that we get the same non-intuitive results from non-experimental data collected from the same job board we use in our field experiment.

matched employer-employee data to try to account for firm or job-match heterogeneity (Taber and Vejlin, 2016; Lavetti, 2017; Lavetti and Schmutte, 2017; Sorkin, 2018). More recently, researchers have used the stated preference approach to avoid problems with non-experimental data. This approach estimates average workers' preferences based on their choices between pairs of exogenously assigned hypothetical jobs with different combinations of amenity levels and pay (Eriksson and Kristensen, 2014; Mas and Pallais, 2017; Wiswall and Zafar, 2017; Maestas et al., 2018).[3]

We see two key advantages of using a natural field experiment (Harrison and List, 2004) that offers real jobs to real job seekers. Like the stated preference approach, the field experiment estimates workers' valuation free of potential biases from omitted variables and sorting. But in contrast to the stated preference approach where job-choice decisions are not incentivized, our subjects are actually searching for jobs, and hence have incentives to respond in ways most likely to get them the jobs they want.[4] The incentive compatibility of our experimental estimates, plus the fact that the estimates come from real job seekers who did not know they were under scrutiny in a scientific study, should make them more internally and externally valid. On the other hand, we look at a narrow set of jobs (and at one employer), so the results may not generalize to different types of jobs and the workers searching for them.

We focused on job flexibility for a few reasons. First, job flexibility is common and appears likely to increase.[5] Second, workers have interest in job flexibility to make it easier to integrate work

[3] For example, Mas and Pallais (2017) use a discrete choice experiment in hiring for a U.S. call center to estimate willingness to pay (WTP) for alternative work arrangements relative to traditional office positions. They find that most workers are not willing to pay for scheduling flexibility (choosing when to work), but are willing to give up 20% of their wage to avoid a schedule set by an employer on short notice, and are willing to pay 8% for the option to work from home. Maestas et al. (2018) use hypothetical job profiles to estimate WTP for non-monetary job characteristics (including schedule flexibility and telecommuting opportunities) in a representative sample of U.S. workers, and find that setting one's own schedule and telecommuting are equivalent to 9.0% and 4.1% wage increases, respectively.

[4] In contrast, for example, Mas and Pallais (2017) "did not tell applicants that these were the actual positions" and "assured applicants that we would not look at their choices before making hiring decisions" (2017, p. 3,729).

[5] In the United States, 79% of employers offered some degree of flexible working hours to their employees in 2008 (Galinsky et al., 2008). In the European Union, 55% of firms with 10 or more employees used policies with flexible starting and ending times in 2009 (Riedmann et al., 2010). About 18% of the labor force in the United States telecommuted at least once per week in 2008 (MSNBC, 2008), and a quarter of European employees mostly worked from places other than the office in 2010 (European Foundation for the Improvement of Living and Working Conditions, 2012). Moreover, in the United States, the proportion of employees who primarily work from home has more than tripled over the past 30 years (Mateyka et al., 2012).

and family obligations (Perlow, 1997), and firms may regard job flexibility as a means to increase motivation and commitment (Caillier, 2012; Eaton, 2003).[6] Third, governments in many countries are trying to increase labor supply in response to population aging, and increased job flexibility may help achieve this goal.[7] Fourth, with the spread of laptops and cellphone connectivity, telecommuting is becoming increasingly common in developing countries,[8] and may help these countries manage growing congestion and urbanization. Finally, to conduct our experiment on the valuation of job flexibility, we needed cooperation from an interested employer; the employer with which we worked was particularly interested in how offering more flexible jobs would affect recruitment (and potentially wages as well).

While there may be a strong prior expectation that workers value job flexibility, there may be downsides to flexibility. Working from home may reduce the chance of promotion due to less on-the-job training and less face time in the office (Kossek and Dyne, 2008; Elsbach et al., 2010; Bloom et al., 2015). Flexible work schedules may not only allow workers to choose when to work, but may also entail requirements to work longer or at irregular hours (Goldin, 2014) – consistent with the findings of Mas and Pallais (2017) regarding workers' negative valuation of employers' ability to set schedules on short notice. Given both the potential importance of workers' valuation of job flexibility, and the possibility that *a priori* expectations are wrong, our evidence – aside from deploying methods to obtain more reliable evidence on workers' valuation of job flexibility – is substantively important.

Our experimental evidence points consistently to workers valuing job flexibility. Subjects were much more likely to apply for flexible jobs, conditional on the salary offered.

We also develop a framework to guide the interpretation of our findings. We start by showing that there are assumptions under which the estimates accurately measure job seekers' willingness to

[6] There is a good deal of research on the impacts of job flexibility on employers and workers. See, Delaney and Huselid (1996), Huselid et al. (1997), and Konrad and Mangel (2000) for effects on firms; Caillier (2012) and Possenriede et al. (2014) for effects on individual outcomes such as work motivation; Eaton (2003) and Lyness et al. (2012) for impacts on organizational commitment and perceived productivity; and Eldridge and Pabilonia (2008, 2010) for effects on working hours.

[7] For example, Chen et al. (2019) examine the benefits of flexible work arrangements for Uber drivers, and conclude that drivers would reduce work hours supplied by more than two-thirds if required to supply labor inflexibly at prevailing wages.

[8] The share of managers allowing telecommuting in many developing countries is estimated to be 10-20% (Bloom et al., 2015).

pay for job flexibility. We then relax these assumptions to clarify conditions under which these estimates remain informative by providing lower bounds on how much workers are willing to pay for job flexibility – as well as conditions under which they may not be informative. Under the conditions for which the experimental evidence can be used to compare jobs seeker's valuation of job flexibility and salary, our evidence implies fairly high valuation of the most flexible jobs.

2. Experimental design

Our field experiment is designed to provide estimates of workers' valuation of job flexibility free of the biases from job and firm heterogeneity and sorting in estimates using non-experimental data, as well as potential problems from using hypothetical scenarios in the stated preference approach. In the experiment, we exogenously varied job flexibility conditions in posted jobs, and the salary offered. The flexibility conditions included schedule or time flexibility (when to work) and place flexibility (the ability to work from home, or off-site). We collected information on applications to these jobs to obtain evidence on workers' valuation of job flexibility.[9]

Experimental setting – job board, employer, positions, and targeted job seekers

Our field experiment was run on a very large nationwide online job board in China, which posts tens of millions of job openings per year, with over 100 million job seekers in a year of whom millions are active each day. The job board specializes in white-collar, high-education jobs and job seekers are highly educated.

Designing job ads for the experiment was simplified because the ads are produced using a standard template capturing information about the job and employee requirements. Importantly, the job information includes fringe benefits – among them flexible working conditions[10] – and the range of pre-tax monthly salary offered.[11]

We collaborated with a start-up company in the information technology (IT) industry. The company had real recruitment demand for several positions and was interested in exploring how

[9] We transferred the resumes of the job applicants to the company (as explained below), and its HR department contacted selected applicants for further recruitment procedures. The company contacted 87 applicants for interviews and the rest of them received a rejection letter via their account on the job board.
[10] An employer can choose up to eight pre-specified fringe benefits. The list of potential fringe benefits appears in Appendix Table A2 (in Online Appendix A).
[11] Online Appendix B provides examples of randomly selected job ads posted on the job board.

varying the flexibility of working conditions offered would affect its recruitment. IT jobs are common on the job board (about one-fifth of jobs, based on the non-experimental data we collected, described in Online Appendix A), and companies in this industry frequently allow flexible working conditions (about one-quarter of the above-mentioned jobs). Using a small (20-99 employees) and not well-known company helped minimize the effect of our intervention on the market. The company is located in the northwestern part of Beijing where there is a cluster of IT companies.

The job positions used in the experiment were dictated by the demands of the company, and the suitability of positions for offering job flexibility. Among the positions for which the company planned to recruit, we selected five that were suitable for setting up independent remote tasks that could be performed with an internet connection from different locations or at different times. The positions were Java engineer, financial executive, human resource manager, marketing executive, and sales executive, which could be classified in the following broader occupations that could be listed on the job board: software, finance, personnel, marketing, and sales management.

Job seekers need to first register and provide individual information to construct a standardized resume. The required information includes, among other things: gender, birth date, education and work/internship experience, type of job sought, current employment status, and generic monthly salary expectation ranges (pre-tax).[12] After registering, job seekers can access job ads listed on employers' webpages, or use a search bar to search for jobs. When they click on an ad, they see a full-page description, and they can then click the "apply" button in a job ad to apply for that job and send their generated resume to the employer. The employer then receives the generated resume along with contact information. Most employers choose to contact applicants offline outside of the job board's communications system, by telephone or email, so the job board does not capture the final outcomes of the search process (callbacks, job offers, and pay).[13]

We define the population of eligible job seekers targeted for our experiment based on the

[12] The ranges are selected from a drop-down menu, but job seekers can choose not to reveal this to potential employers.
[13] This information, and other information about how the job board works in practice, was provided to us in conversations with staff at the job board.

following criteria: (1) residence in Beijing at the time of the experiment; (2) college degree or higher; (3) active in job search (having logged into their job board account within one month from when we first extracted their resumes for consideration for inclusion in the experiment);[14] and (4) a match between any of the "intended occupations" chosen by the job seeker (they could indicate up to three) and our chosen occupations.[15] We drew the population of job seekers for the study, which was all job seekers registered on the job board fulfilling these criteria, one week before each set of experimental contacts was made (as explained below).

Treatments

For each job position, we implemented a 2 × 2 between-subjects factorial design in which we varied whether the job ad posted included time or place flexibility. There were four variants:

- *NoFlex*: The job did not offer time or place flexibility. Employees needed to work in the office Monday through Friday, for 8 hours between 9 AM and 6 PM.
- *TimeFlex*: On Monday, the conditions were the same as in the *NoFlex* treatment. For Tuesday through Friday, employees had to work in the office, but could choose their 8 hours starting between 7 AM and 10 AM and ending between 4 PM and 7 PM.
- *PlaceFlex*: On Monday, the conditions were the same as in the *NoFlex* treatment. For Tuesday through Friday, employees could work wherever they liked for 8 hours between 9 AM and 6 PM by logging into the company's online working system.
- *FullFlex*: The same conditions for Monday applied. For Tuesday through Friday both the time flexibility in *TimeFlex* and the place flexibility in *PlaceFlex* applied.[16]

Similar kinds of job flexibility conditions are referenced in data on worker preferences for job flexibility.[17] Moreover, flexible conditions are common in some of the experimental occupations,

[14] The one-month cutoff was chosen to correspond to the usual definition of unemployment. In the United States a worker is defined as unemployed if she searched for a job in the past month (see https://www.bls.gov/cps/cps_htgm.htm#unemployed, viewed August 15, 2018). In China, the criterion is three months (see http://www.stats.gov.cn/tjsj/tjzd/gjtjzd/201807/t20180717_1610135.html, in Chinese, viewed August 18, 2018). We chose the more restrictive U.S. standard to have a more active sample of job searchers.
[15] For a job seeker with multiple "intended occupations" matching our occupations, for the experimental intervention we randomly assigned one of the positions (using equal probabilities).
[16] In all cases, our job ads indicated that these flexible work arrangements could be used after the first month on the job.
[17] For example, according to a 2017 flexible work options survey of over 5,500 U.S. professionals on

including software, sales, and marketing, but less so for finance and personnel.[18]

The time ranges for *TimeFlex* were based on actual practices to alleviate traffic congestion that were commonly observed in other job ads on the job board, and met the needs of the collaborating company. Between 7 AM and 9 AM, and again between 5 PM and 7 PM, one lane is designated for "buses only" on main roads in Beijing. The average one-way commuting time in Beijing was 53 minutes in 2017 (DIDI, 2018). Our flexible time setting thus allowed commuters to avoid the peak hours by finishing or starting commuting before or after peak travel hours. We also verified, from thousands of job ads collected from the job board during the summer of 2017, that our *TimeFlex* conditions were among the most common that appeared. To try to rule out the possibility that time flexibility was interpreted as requiring employees to work overtime or at irregular hours, we explicitly mentioned that working hours were 8 hours chosen by employees. Our *PlaceFlex* setting might be considered as most relevant to reducing commuting time.

The job flexibility treatment conditions were presented in three places in the job ads – and other than the treatment conditions, the job ads were identical for each position. First, the title of the job ad at the top included the treatment conditions. For the *NoFlex* treatment, nothing appeared after the job title. For the *TimeFlex*, *PlaceFlex*, and *FullFlex* treatments, respectively, the additional text appearing was "with flexible work time," "with flexible work place," and "with both flexible work time and place." Second, information on the treatment appeared in boxes below the job ad title listing fringe benefits. For the *TimeFlex*, *PlaceFlex*, and *FullFlex* treatments the box indicated "flexible working conditions," while this did not appear among the fringe benefits for the *NoFlex* treatment.[19] Third, the "Work Arrangements" section of the ad listed the details of the job flexibility conditions

flexjobs.com, telecommuting and flexible scheduling are considered as the preferred types of job flexibility by 81% and 70% of respondents, respectively – far higher percentages than for other types of flexibility such as part-time schedule (46%), alternative schedule (44%), and freelance contract (39%); see Reynolds (2017).

[18] Among all the jobs posted on *flexjobs.com*, during January to September, 2017, computer and IT, sales, and accounting and finance are listed as the three out of the top 10 career fields for flexible jobs (Reynolds, 2017). In the non-experimental data we collected (see Online Appendix A), 25.0%, 24.7%, and 21.0% of the jobs in software, sales, and marketing, respectively, listed "flexible job" in the fringe benefits, whereas 13.8% and 15.4% of the jobs in finance and personnel, respectively, listed "flexible job" in the fringe benefits.

[19] All the other fringe benefits were identical across treatments. The other fringe benefits included five insurance funds and one housing fund, double pay at year end, performance-based bonus, communication allowance, paid annual leave, regular health examination, and holiday benefits.

outlined above. These are the three places where information on job amenities commonly appeared in other job ads.[20]

In the job ads, the required educational level was set to college degree,[21] which was the most common education requirement on this job board. The required years of work experience was set to 5-10 years.[22] We chose this experience range so that, given the college degree requirement, targeted job seekers would tend to be in an age range when there are family responsibilities associated with small children, which could make job flexibility more salient and valuable.[23] Online Appendix C provides an example of the job ad for the financial executive position in the *FullFlex* treatment.

The job flexibility treatments were also presented in the job ad emails and app messages that subjects received. The email pushing and message pushing – through which job invitations are sent to potentially suitable candidates based on matching algorithms – are two services the job board provides to employers to facilitate matches between jobs and workers.[24] Email pushing entails sending a brief description of selected job openings via email to candidates.[25] Message pushing plays a similar role. A job seeker typically receives 1-2 emails or app messages per day during the time she frequently logs onto her account or searches for jobs. In the experimental emails, the treatment condition appeared in both the subject line of the emails (as in the titles of the job ads), and under "Work Arrangements" (in the same manner as specified in the job ads). The app messages simply contained information on the job title and the treatment condition for *TimeFlex*, *PlaceFlex* and

[20] As in all such experiments, in manipulating job flexibility conditions (or pay, discussed below), the assumption is that the subjects do not perceive the jobs as differing along other dimensions. A potential concern is that job seekers may use the advertised job characteristics as signals to infer other unobserved job characteristics. However, this is unlikely to influence our analysis for two reasons. First, our experimental job ads (and most ads on the job board we use) include many job details that may affect application incentives, including the job description, job responsibilities, and employer information. Second, we control for any differences in the descriptions in our regression analysis.

[21] Employers could select from no degree requirement, below college, college, bachelor, or master or above.

[22] Employers could select from no experience requirement, or 0, 0-1, 1-3, 3-5, 5-10, or >10 years.

[23] However, we did not restrict our targeted eligible job seekers to have work experience of 5-10 years.

[24] The matching algorithms can rely on information in the resume (e.g., the salary offered in the job matches the monthly salary expectation of the job seeker), job seekers' previous job search behavior (e.g., the occupation or industry of the job matches the occupation or industry in which the job seeker searched in the past), or application behavior on the job board (e.g., the occupation or industry of the job matches the occupation or industry in which the job seeker has previously applied) on the job board. Employers can define some parameters of the matching algorithm for their postings (e.g., pushing ads to those with a college degree, or working in a particular industry).

[25] According to Horton (2017), this is a common service on online job boards.

FullFlex treatments. Both the job ad emails and app messages included a link to the job ads.

To obtain evidence on willingness to pay (WTP) for job flexibility, in the job ad emails and app messages we varied the pre-tax monthly salary offered across three ranges – 10,000-15,000 CNY ("*Low*"), 15,000-20,000 CNY ("*Medium*"), and 20,000-25,000 CNY ("*High*").[26] These ranges were derived from the distribution of salary ranges in over 8,000 real job ads we collected from the job board during the summer of 2017, which were recruiting for our five experimental occupations and required at least college education and at least 5-10 years of work experience.[27] The use of salary ranges rather than specific salary amounts in the emails and messages coincided with the "negotiable" setting for monthly salary in the job ads.[28] For job ad emails, these salary ranges were presented in the subject line and right after the job title in the main body of the emails. For the job ad app messages, these salary ranges were presented in the message right after the job title. Online Appendices D and E present examples of the job ad email and the app message for the financial executive position in the *FullFlex* treatment.

Experimental procedure

In the experiment, the company published the job ads on the job board, and then the job board pushed the corresponding job ads via both emails and app messages to the population of eligible job seekers, with one job seeker receiving one ad.[29] Since all active job ads – with the corresponding flexibility conditions – had to simultaneously appear on the employer's webpage on the job board, we posted the job ad for each treatment sequentially, one treatment at a time for a week, for each of the five job positions. We posted ads for the five positions simultaneously. To control for potential

[26] We used a monthly salary instead of an hourly wage in the experiment because it was natural for full-time jobs of similar types (almost all the ads on the job board) to set a fixed monthly salary. We also explicitly mentioned in the job ads that the work day was 8 hours, to try to preclude higher salary being viewed as compensation for longer working hours.

[27] We separately computed the 25th, 50th, and the 75th percentiles for the lower and upper limit of the salary range, which were 10,001, 15,001, and 20,001 CNY for the lower limit, and 15,000, 20,000, and 30,000 CNY for the upper limit, respectively. To keep the width of the ranges comparable, we used 25,000 as the upper limit for the High range.

[28] We incorporated experimental variation in salary offers in the emails and app messages only, and simply presented pay as "negotiable" in the job ads published on the employer's webpage (discussed below), because it would have been far more complex and taken far longer to also coordinate the webpage job ads to vary both flexibility conditions and pay. However, the word "negotiable" is not inconsistent with the pay ranges pushed to experimental subjects.

[29] As discussed later (and in Online Appendix F), there was also a survey stage to our experiment after the application stage.

temporal confounds due to the order in which the treatments appeared, we randomized the order of the treatments by position and week, as shown in Table 1. The application stage lasted for four weeks. In each week we published the job ad for each position and each treatment on Thursday morning and stated in the ads and emails that the application deadline would be 9 AM the following Wednesday. Applying for the experimental jobs serves as the primary measure of interest in our job.

We could have simply drawn the population of job seekers (meeting the criteria described above) before the start of the entire experiment, randomly divided the population (for each position) into four equal-sized parts, and randomly assigned one to each week. But job seekers might be less active in later weeks. To ensure that eligible job seekers across weeks and treatments had a similar degree of activeness in job search, we drew the population of eligible job seekers, for each position, one week prior to each week during which the experiment was administered.[30]

We divided our selected population for each position in each week into two parts: 4/5 were randomly split into three groups of equal size to be sent job ad emails and app messages with different monthly salary offers in the three ranges described above; the remaining 1/5 were included as an additional control group (besides those assigned the *NoFlex* treatment). This additional control group was not contacted by email or app message, enabling us to consider Hawthorne effects for those contacted in the *NoFlex* condition. The treatment design is summarized in Table 2.

For the job seekers who were sent emails and app messages, applying for a job involves two steps. First, they needed to click the link to the job ad, which directed them to the corresponding job ad on the employer's webpage on the job board. Second, they needed to click the "apply" button on the webpage. The redirection from the email or app system to the webpage was standard for all email or message recipients contacted through these means, and applications taken on the job board could be recorded. Any job seeker, whether or not they received an email or an app message, could search the job board and find and apply for the experimental jobs. Throughout the experiment, there were no

[30] We employed the following approach to determine how and whom to select in each week for each position. In week 1, we randomly selected 1/4 of the eligible job seekers for each position. In weeks 2-4, we first excluded those who had already been selected, and those who had applied for the experimental jobs without being contacted in previous week(s), and then randomly selected 1/3, 1/2, and all of the remaining eligible job seekers, respectively, for each position. In this manner, we used up the entire population of eligible job seekers with comparable population size across weeks.

communications between the job seekers and the experimenters, except for the carefully scripted job ad emails and app messages sent in the application stage.

Compared to the procedure adopted by Flory et al. (2015) and Hedblom et al. (2019), in which randomized treatment conditions were revealed only after job seekers had expressed interest in the job by emailing their resumes, our treatment conditions regarding job flexibility were presented to job seekers in the job ads at first contact. In our view, this procedure had two advantages: it preserved the normal way of presenting key job conditions on the job board we used; and it allowed us to collect data on responses to treatment conditions in one stage.

3. Experimental data analysis

The experiment was conducted in January and early February of 2018. We randomly assigned eligible job seekers into various job flexibility conditions and gave the subsample for each treatment to the job board.[31] The key data collected from the experiment are the individual-level data on who applied for the experimental jobs.

Do flexible jobs attract more applications?

The total number of job seekers included in the experiment is 123,988, with 23,000 to 26,000 in each treatment (and the additional control group) depending on the size of the population of eligible job seekers in each job position and each week. Since the job ad emails and app messages with different combinations of treatments, salary offers, and positions had separate links, the job board could monitor each link for how many emails and app messages had been successfully sent, how

[31] The mean differences for almost all individual characteristics are small across treatments. We also ran pairwise (across four treatments and the additional control group) Kolmogorov-Smirnov tests of equality of distributions of these characteristics. We find differences significant at the 10% level for 35 out of 150 comparisons – 20 more than we would expect by chance (see Appendix Table G1 in Online Appendix G). There could be two reasons. First, our randomization did not stratify on any individual characteristics, since the resume items were so numerous and different characteristics had different numbers of missing values. Second, we have a very large sample compared to most experiments, which makes it more likely that small differences are statistically significant at conventional levels. Our regressions control for these differences in characteristics across treatments, but our estimated treatment effects do not vary with whether we include these controls.

many of them had been opened, and how many resulted in an application.[32,33]

Table 3 reports on the joint distributions of the job flexibility conditions in the jobs subjects applied to and the job flexibility conditions with which they were treated, for the full sample (in the top panel), and then by salary level. Recall that we posted the experimental job ads on the company's webpage, so subjects could have applied for jobs with different flexibility conditions than those with which they were treated. However, most numbers are on the main diagonal, indicating compliance (job seekers applying for the treatment they received), and infrequent applications stemming from experimental job seekers simply finding the ads on the webpage.[34]

The application rates reported in the last column are computed as the number of applications divided by the number of job seekers who were sent job ad emails and app messages. For the full sample, the application rates for all treatments are below 0.5%. Application rates are low because there are thousands of job openings posted every day, presumably with many job descriptions similar to ours, so that getting an email or an app message as part of the experiment would not be expected to generate a large number of applications.[35] Nonetheless, the application rates are 62%-92% higher for flexible jobs than for non-flexible jobs. For example, the application rate in the *NoFlex* treatment is 0.24%, compared to 0.45% for *FullFlex*. The bottom three panels show that application rates increase with the offered salary level. More important from our perspective, within each salary level, the application rate is higher in the treatments with job flexibility.

Next, we turn to regression analysis using the individual-level data. Table 4 reports the analysis of the differential effects of the job flexibility treatments on the application decision, with

[32] Not every email was successfully sent because email addresses in some resumes were invalid. Nor were app messages always successfully sent because not every job seeker downloaded and installed the app or agreed to receive messages pushed to them. It seems highly likely that missing data for these two reasons was random with respect to preferences for job flexibility.

[33] We found no impact of the flexibility treatment on whether job seekers opened the job ad email or app message, likely because the information available when deciding to open an email or message has only a brief mention of flexibility conditions in the subject line or next to the job title. (See Appendix Tables G2 and G3 in Online Appendix G.) In contrast, once the email or app message link is opened, a job seeker sees the details about the job flexibility conditions.

[34] The timing of the applications in the off-diagonal cells indicates that these observations are a mix of applying with a delay, and "non-compliance" due to the ads also appearing on the company's webpage on the job board.

[35] Personal conversations with the staff at the job board indicated that a 0.5% rate of application for job ads pushed to job seekers is typical.

NoFlex as the reference group. The dependent variable is an indicator for whether the job seeker applied for our job.[36] We present marginal effects from a probit model in the top panel, and tests of estimated differences of marginal effects in the bottom panel. Column (1) includes only the treatment dummy variables. We find evidence that all three job flexibility treatments significantly boost application rates relative to the *NoFlex* treatment, with effects ranging from 0.13 to 0.24 percentage point – with the effect largest for the *FullFlex* condition. These estimates should be compared to a baseline application rate of 0.24 percentage point for the *NoFlex* treatment (Table 3, top panel), indicating roughly a doubling of application rates. Column (2) adds the salary offer level and position dummy variables; the estimates are slightly smaller, but statistically significant in all cases. Column (3) adds job seekers' individual characteristics;[37] the estimates are little changed from those in column (2) and continue to indicate that more flexible job conditions boost application rates significantly. The bottom rows of the table indicate that there is one significant difference between the different job flexibility treatments – with a significantly higher application rate (by 0.08-0.11 percentage point) for *FullFlex* vs. *PlaceFlex*.[38]

More importantly, the marginal effects of flexibility conditions and high salary level have similar magnitudes; the *p*-values from tests of the equality of marginal effects of *TimeFlex* and *High*, *PlaceFlex* and *High*, and *FullFlex* and *High* all exceed 0.10. This evidence suggests that job seekers value flexibility by amounts in the same ballpark as having a monthly salary that is higher by 10,000 CNY (i.e., from the salary range 10,000-15,000 CNY to 20,000-25,000 CNY), which would imply

[36] Applications made to treatments or positions other than those applicants were sent are not included. Thirty-two job seekers made in total 39 applications to other treatments; there are two more applications than the number of off-diagonal elements because of applications made to different positions than the ones sent, but with the same flexibility treatments.

[37] In the self-reported resume data on individual characteristics, there are sometimes inconsistencies related to the dates of events reported, such as a birth date later than dates of other events like the start of highest degree education or a first job, or start dates for specific spells (such as education) later than ending dates. There are also some less clear inconsistencies, such as completing education at too young an age (e.g., completing college before age 18 or university before age 20). We clean the data to eliminate these kinds of inconsistent cases whenever individual characteristics are considered. This results in excluding fewer than 1.6% of observations. In addition, since filling in some characteristics such marital status is optional, including them also excludes a significant proportion of observations.

[38] Appendix Table G6 reports the results from the analysis of Hawthorne effect by including the additional control group. We find a positive and significant Hawthorne effect, and similar effects of the different flexibility treatments relative to each other.

that flexibility is highly valued.[39]

To provide more detailed evidence on the monetary valuation of job flexibility, column (4) adds the interactions of the treatment dummies and the salary offer dummies.[40] The significant estimated difference in application rates between *FullFlex* low salary and *NoFlex* medium salary jobs, and the insignificant estimated difference in application rates between *FullFlex* low salary and *NoFlex* high salary jobs, suggests that job seekers value the option to work flexibly both in terms of time and place by 5,000-10,000 CNY a month. The significant estimated difference in application rates between *FullFlex* medium salary and *NoFlex* high salary jobs reinforces the previous result that the valuation for the full flexibility condition holds irrespective of salary levels.

The high WTP for *FullFlex* jobs could arise because our sampled job seekers are likely to have family responsibilities; their average age is 28, and 58% are female.[41] In addition, since the average commuting time both ways (106 minutes) accounts for approximately 25% of working time (eight hours a day), it can be quite valuable to save this time and associated costs.

Overall, the evidence on job application rates indicates that job seekers are more likely to apply for flexible jobs. This provides evidence that job flexibility is a positive job amenity. However, there may be variation in preferences for job flexibility across different types of job seekers. We next present evidence on this variation.

Heterogeneity across job seekers in applications for flexible jobs

We explore differences in responses to flexible job offers by gender and marital status, which can be related to family responsibilities. We do not have information from job seekers' resume data on whether they have children, but we presume that married people are considerably more likely to have children and consequently might particularly value job flexibility, especially women. Either place or time flexibility could make it easier to juggle the demands of work and family.

To address this question, we introduce into the models of the treatment effects on applications (from Table 4) the two-way interactions between treatment dummies, gender, and marital status, as

[39] Below, we discuss conditions under which this interpretation is likely valid.
[40] This analysis was not included in our pre-analysis plan because it was suggested by a reviewer.
[41] According to China's 2010 National Population Census (Issue 6-1), the median age of first birth is 27 for women in urban areas.

well as three-way interactions, and the estimation results are reported in Table 5.

The main effects of the flexibility treatments are now reflective of treatment comparisons for unmarried males. They show that flexible jobs do not increase application rates significantly for this group. In the bottom panel of the table, rows (i)-(iii) show that unmarried females are also not more likely to apply for flexible jobs. Rows (iv)-(vi) indicate that married males respond more strongly to the full flexibility treatment, compared to the no flexibility treatment. And rows (vii)-(ix) indicate that married females respond more strongly to all types of flexible jobs.

The evidence that married job seekers seem to prefer both types of flexibility is consistent with workers, at least in a large, congested city like Beijing, placing a positive value on being able to economize on commuting time by avoiding the peak travel hours or by working from home. Married women may value flexibility more because of greater responsibilities for housework, child care, and caring for parents and parents-in-law. Given the above findings, we look in particular at evidence on workers' valuation of job flexibility for married females, using the same specification as in Table 4, column (4). Despite the smaller sample, the estimated differences in Table 6 provide some statistical evidence that married women are willing to trade off full flexibility jobs for 5,000-10,000 CNY a month (based on the {*FullFlex* × Medium} vs. {*NoFlex* × High} comparison at the bottom of Table 6). In addition, this estimated differential for married females (0.0034) is larger than the estimate for the sampled job seekers as a whole (0.0020, in the last row of column (4), in Table 4), indicating that married females particularly value full flexibility.

4. Interpreting the estimates

Our experimental evidence points consistently to workers applying more to jobs offering greater flexibility, even when the pay is lower. In this section, we develop a framework for interpreting our data to better understand assumptions under which the estimates likely provide accurate information about job seekers' willingness to pay for job flexibility, and weaker assumptions under which these estimates remain informative by providing lower bounds on how much workers are willing to pay for job flexibility. In each case, we consider the implications of different assumptions for the data-generating process, the marginal effects we can estimate in terms of the underlying parameters, and what these marginal effects can tell us about the relative utility of job flexibility and

higher pay.

We start with the simplest formulation. Let jobs be distinguished by the wage (w_j) and flexibility (f_j), where $w_j = 1$ for high-wage jobs and $w_j = 0$ otherwise, and $f_j = 1$ if the job is flexible and $f_j = 0$ otherwise.[42] Applying to a job has cost *c*. We assume a job seeker is not constrained in how many jobs they can apply to, and given how the job board operates, applying for one job does not preclude applying for other jobs. But given that we find low job application rates in our experiment, the cost of applying must be non-negligible. Thus, we can think of a job seeker as simply receiving notification of one job at a time, and deciding whether to apply.

We make two assumptions to start: first, a job seeker does not account for variation across jobs in the likelihood of getting an offer – including possibly how this likelihood might vary with how many others apply; and second, all workers value wages and job flexibility similarly, up to a random taste component ε.[43]

We assume the utility of job j for individual *i* is

(1) $U_i(w_j, f_j) = \tau w_j + \theta f_j + \varepsilon_{ij}$,[44]

with ε_{ij} i.i.d. and orthogonal to w_j and f_j. Denoting by p the probability that applying results in the job seeker getting the job, the job seeker applies only if (dropping subscripts)

$$p \cdot (\tau w + \theta f + \varepsilon) - c > 0 \ ,$$

or

(2) $p(u + \varepsilon) > c$,

where $u = \tau w + \theta f$. If we assume a normal distribution for ε, with mean zero and standard deviation one (a normalization), then the probability of applying (y) is equal to

(3) $y = \Phi(u - \frac{c}{p})$.

If we estimate a probit model, we identify the coefficients τ and θ, and hence the relative

[42] We have multiple flexibility and wage conditions in our experiment, but they can always be boiled down to comparisons between how job seekers respond to particular combinations of salaries and job flexibility conditions. Hence here we work with the simplest case.
[43] Variation in application costs across individuals can be folded into the random taste component.
[44] We assume throughout that $\tau > 0$ and $\theta > 0$.

valuations of the different job characteristics.[45]

The marginal effects of *w* and *f* on the probability of applying, treating w and f as continuous, are

(4) $dy/dw = \tau\phi(u - \frac{c}{p})$

(5) $dy/df = \theta\phi(u - \frac{c}{p})$,

so the relative marginal effects identify the relative valuations of the different job characteristics.[46]

We consider two complications and what they imply for the interpretation of our estimates. First, we allow p to differ across jobs, with potential applicants accounting for both the supply of such jobs (m) and the potential number of people who apply (y), so

(6) $p = \frac{m}{y}$.

We assume both m and y are defined on a continuum of mass 1, and $p < 1$. In this case, a person applies if $\frac{m}{y}(u + \varepsilon) > c$, or

(7) $\varepsilon > -u + \frac{c}{m}y$.

We assume Nash equilibrium, so that y equals the proportion of workers who apply, implying

(8) $\Phi(u - \frac{c}{m}y) = y$.

This expression makes sense. A higher m implies the job is more plentiful, raising the probability of applying, and a higher *y* means more people apply, reducing the probability of applying. Interpreting estimated marginal effects in this case is more complicated because each individual's application decision depends on what all applicants do. Equation (8) defines an implicit

[45] Formally, we identify the relative values of τ and θ (because we normalized the standard deviation of ε).

[46] If instead we define the marginal effects based on differences between the cumulative distribution functions, which is technically the more appropriate calculation for dichotomous independent variables, then these marginal effects are

$dy/dw = \Phi(\tau - \frac{c}{p}) - \Phi(-\frac{c}{p})$

$dy/df = \Phi(\theta - \frac{c}{p}) - \Phi(-\frac{c}{p})$.

For non-trivial differences in τ and θ, these expressions approximate the relative valuations. For example, if we set $c = 0$ (so that the problem is more akin to a choice of transportation mode where there is not a cost of the choice), and set τ = 0.5 and θ = 1, then these two expressions equal 0.19 and 0.34. The discrepancy goes to zero in the limit as θ approaches τ, as then the two calculations become the same, and the implied utility is proportional to the true utility.

function of y. Substituting for u, defining the implicit function $g(y) = \Phi\left(\tau w + \theta f - \frac{c}{m}y\right) - y = 0$, totally differentiating, and setting $dg = 0$, we obtain the marginal effects

(9) $$dy/dw = \frac{\tau\phi\left(u-\frac{c}{m}y\right)}{1+\frac{c}{m}\phi\left(u-\frac{c}{m}y\right)} \ ,$$

(10) $$dy/df = \frac{\theta\phi\left(u-\frac{c}{m}y\right)}{1+\frac{c}{m}\phi\left(u-\frac{c}{m}y\right)} \ .$$

Given that these equations are the prior marginal effects (equations (4) and (5)) divided by a positive number greater than 1, they are smaller than before, which makes sense because when applicants take account of other people (with the same tastes) applying, they respond more weakly to job characteristics that they value positively. Nonethelss, the ratio of these marginal effects still equals τ/θ, implying that we still accurately estimate the relative values of job flexibility and wages.

Next, we add the additional complication that m depends on job characteristics, so it becomes $m = m(w, f)$. Assume (ignoring the dichotomous nature of the job characteristics) that $m_w < 0$ and $m_f < 0$ (both first derivatives are negative), consistent with lower labor demand when jobs offer higher wages or more job flexibility.[47] Now the partial derivatives become

$$dy/dw = \frac{\left(\tau+\frac{cy\cdot m_w}{m^2}\right)\cdot\phi\left(u-\frac{c}{m}y\right)}{1+\frac{c}{m}\phi\left(u-\frac{c}{m}y\right)} \ ,$$

$$dy/df = \frac{\left(\theta+\frac{cy\cdot m_f}{m^2}\right)\cdot\phi\left(u-\frac{c}{m}y\right)}{1+\frac{c}{m}\phi\left(u-\frac{c}{m}y\right)} \ .$$

Since $m_w < 0$ and $m_f < 0$, both partial derivatives are reduced, because p falls for more desirable jobs. However, the ratio of these expressions is no longer equal to τ/θ, but instead to

(11) $$\frac{dy/dw}{dy/df} = \frac{\tau+\frac{cy\cdot m_w}{m^2}}{\theta+\frac{cy\cdot m_f}{m^2}} \ ,$$

which implies that the estimated marginal effects no longer capture the relative valuations of the different job characteristics.

We can, however, establish a condition under which the ratio of estimated marginal effects in

[47] We could think of job seekers assuming it is less likely a firm will hire when jobs offer these characteristics, since their hiring threshold is higher, or fewer positions being available, for the same ad, when jobs offer these characteristics.

equation (11) gives a lower bound for the relative valuation of job flexibility; in this case, if the evidence from applications indicates that the flexible job is preferred to the higher-wage job, we know the worker is willing to pay at least the wage differential associated with the higher-wage job for the flexible job. This lower-bound result holds if

$$(12) \quad \frac{\tau+\frac{cy\cdot m_w}{m^2}}{\theta+\frac{cy\cdot m_f}{m^2}} > \frac{\tau}{\theta} \ .$$

We obtain a more-interpretable expression if we are willing to assume that $\frac{cy\cdot m_w}{m^2}$ and $\frac{cy\cdot m_f}{m^2}$ are sufficiently small that higher-wage and more flexible jobs attract more applicants – i.e., these terms do not outweigh the utility parameters τ and θ, so both the numerator and denominator of the left-hand side of equation (12) are positive. This assumption is consistent with the experimental evidence that both job flexibility and higher pay attract more applicants. $(\tau + \frac{cy\cdot m_w}{m^2}) > 0$ and $(\theta + \frac{cy\cdot m_f}{m^2}) > 0$, along with equation (12), imply that

$$(13) \quad \frac{m_w}{m_f} < \frac{\tau}{\theta}.$$

Equation (13) establishes that as long as the value of $|m_f|$ relative to $|m_w|$ exceeds the value of θ relative to τ, our estimates will understate the relative valuation of job flexibility. The condition (equation (13)) for the lower-bound result for the relative valuation of job flexibility (equation (12)) to hold is intuitive. In particular, the supply of jobs offering flexibility has to diminish sufficiently rapidly compared to the supply of jobs offering higher wages (when compared to the relative valuation of flexibility and wages) so that the estimated marginal effects are not dominated by workers applying to flexible jobs not because they value them more, but because they are more likely to get them. As long as that is true, our estimated marginal effects will if anything understate the relative valuation of flexible jobs.[48]

[48] A simple formulation that makes this clear is to assume $m = m(w + qf)$, in which case equation (13) becomes $\frac{1}{q} < \frac{\tau}{\theta}$. As long as this condition holds – that is, q (the cost of flexibility to the firm) is sufficiently large – then jobs offering flexibility are sufficiently scarce (relative to jobs offering higher wages) that workers respond less positively to the more-preferred flexible job offer, leading us to understate the relative valuation of flexible jobs.

Although we cannot directly measure q, we can perhaps learn something about q from observing the pay difference in the broader job ad data for jobs similar to those in our experiment (the data used in the analysis of

The lower-bound result is significant. In our experiment, we find that flexible jobs, and especially the most flexible (*FullFlex*) jobs, attract more applications than higher-paying jobs. As long as the lower-bound result holds, then given that the marginal effects understate the relative valuation of job flexibility, the evidence that workers are willing to trade off the pay differences we offered for more flexible jobs would be even stronger if we could recover the utility parameters.[49] Moreover, even if the condition (equation (13)) for the lower-bound result does not hold, we can see from equation (11) that the bias in the estimated relative valuation of job flexibility is small as long as m_w and m_f are small.

The second complication we consider is heterogeneous tastes. To fix ideas, we suppose there are two types of workers: type *h* who place positive value on flexibility (*h* to capture "likes to work from home"), and type *o* who do not (*o* to capture "likes to be at office"); *h* and *o* are used as superscripts in the following discussion. Thus, the utility functions (ignoring subscripts) are

(14) $U^h(w, f) = \tau w + \theta f + \varepsilon^h$,

(15) $U^o(w, f) = \tau w + \varepsilon^o$.

To isolate the issue of heterogeneous tastes, we revert to the simple case of a fixed probability of getting a job offer p. Let H denote the proportion of type *h*, and assume that ε^h and ε^o have the same standard normal distribution. Then the data-generating process for the response of job seekers to the job ads yields the share applying

valuation of flexibility based on employer salary offers in Appendix Table A1 in Online Appendix A). To parallel our experimental jobs, we restricted attention to jobs requiring college education or above, 5-10 years of work experience, and offered by firms of similar size and operating in similar industries as those in our experiment. Flexibility in these data is measured by whether the pre-specified fringe benefit "flexible working conditions" appears in the job ad, which is quite vague. For these jobs, the average midpoints for the pay ranges for non-flexible and flexible jobs, are 18,894 and 20,983 CNY, respectively. The slightly *higher* pay for flexible jobs suggests that there is not a cost disadvantage for offering flexible jobs, in which case it is possible our lower-bound condition does not hold and we overstate the valuation of flexible jobs. However, we do not think one can draw strong conclusions, for three reasons. First, these are not identical jobs differing only in flexibility. Second, the pay difference is not within a company (in contrast to the experiment). And third, the measure of flexibility we can use to study ads on the job board is not specific; the flexibility conditions we offer may indeed be more costly for employers.

[49] We noted earlier that our evidence indicated that workers preferred the most flexible (*FullFlex*) jobs to higher-pay jobs (based on, e.g., the estimates in columns (3) and (4) of Table 4 – with the difference significant in the latter case). If the lower-bound result holds, so that our estimates understate the utility of flexibility relative to pay, then our estimates could also imply that workers value *TimeFlex* and *PlaceFlex* at least as much or more than the pay difference between the high-pay and low-pay jobs. For example, in column (3), Table 3, the estimated coefficients on *TimeFlex* and *PlaceFlex* are 0.0011 and 0.0010, and the estimate coefficient on High is 0.0016.

(16) $y = H\Phi\left(\tau w + \theta f - \frac{c}{p}\right) + (1-H)\cdot\Phi\left(\tau w - \frac{c}{p}\right).$

In this case, the average marginal effects are

(17) $dy/dw = H\tau\phi\left(\tau w + \theta f - \frac{c}{p}\right) + (1-H)\cdot\tau\phi\left(\tau w - \frac{c}{p}\right)$,

(18) $dy/df = H\theta\phi(\tau w + \theta f - \frac{c}{p})$.

Then

$$\frac{dy/df}{dy/dw} = \frac{H\theta\phi(\tau w+\theta f-\frac{c}{p})}{H\tau\phi(\tau w+\theta f-\frac{c}{p})+(1-H)\tau\cdot\phi\left(\tau w-\frac{c}{p}\right)} = H\frac{\theta}{\tau}\frac{\phi(\tau w+\theta f-\frac{c}{p})}{H\phi(\tau w+\theta f-\frac{c}{p})+(1-H)\phi\left(\tau w-\frac{c}{p}\right)} .$$

It is easily shown that the derivative of $\frac{dy/df}{dy/dw}$ with respect to *H* is positive. Therefore the ratio is maximized at *H* = 1. But, at *H* = 1, $\frac{dy/df}{dy/dw} = \frac{\theta}{\tau}$, and thus for *H* < 1, the estimate of $\frac{\theta}{\tau}$ from this ratio of marginal effects is strictly downward biased. Thus, the lower-bound result still holds.

Now consider the more complex case where both types value wages and flexibility, but type *h* values flexibility more, which we capture by keeping τ the same for both types, but $\theta^o < \theta^h$. Then the share that applies is

$$y = H\Phi\left(\tau w + \theta^h f - \frac{c}{p}\right) + (1-H)\cdot\Phi\left(\tau w + \theta^o f - \frac{c}{p}\right),$$

implying

(19) $dy/dw = H\tau\phi\left(\tau w + \theta^h f - \frac{c}{p}\right) + (1-H)\cdot\tau\phi\left(\tau w + \theta^o f - \frac{c}{p}\right)$,

(20) $dy/df = H\theta^h\phi(\tau w + \theta^h f - \frac{c}{p}) + (1-H)\cdot\theta^o\phi\left(\tau w + \theta^o f - \frac{c}{p}\right)$.

We cannot make a definitive statement that the marginal effects of higher wages are overstated for type *h* workers (who value flexibility more), and the marginal effects of flexibility are understated for them, because the two densities in each equation are not equal. For example, in equation (20), $\theta^h > \theta^o$ makes the first term larger than the second (ceteris paribus), but that could be offset if $\phi\left(\tau w + \theta^o f - \frac{c}{p}\right) > \phi\left(\tau w + \theta^h f - \frac{c}{p}\right)$. Clearly, though, if the two terms on each side of this inequality are sufficiently close (i.e., we are evaluating these derivatives at similar parts of the

distribution), then the statement still holds.

There are four additional points. First, with heterogeneous tastes there is no single relative valuation of flexibility versus wages to compare to the ratio of marginal effects. However, under the same condition – that $\phi\left(\tau w + \theta^{o} f - \frac{c}{p}\right) \cong \phi\left(\tau w + \theta^{h} f - \frac{c}{p}\right)$ – the ratio of marginal effects, using the ratio of equation (19) to equation (20), approximates the average relative valuation of flexibility versus wages.[50]

Second, were we to simultaneously introduce the first set of considerations regarding responding to other applicants and the supply of jobs declining with w and f, the same lower-bound results would apply to each type of worker, and hence to these ratios of marginal effects.

Third, a potentially more problematic consideration is if, instead of getting one application at a time and deciding whether to apply, job seekers weigh applying against past openings from which they are awaiting a response or future vacancies to which they expect to apply.[51] In this case, it is possible that if flexible jobs are rarely advertised, and workers know from experience that, as a result, the probability of getting a flexible job offer is very low, workers who value job flexibility may apply at very high rates to increase the likelihood of getting an offer. In this case the application rate to flexible jobs may overstate the valuation of job flexibility. However, accounting for this alternative search/job-application process would substantially increase model complexity.

Fourth, it is important to note that even when the lower-bound result does not hold, the experiment provides unbiased predictions of how job applications will respond to manipulating these features of job offers. This is relevant to, for example, the policy question of whether encouraging flexible jobs might reduce congestion.

Finally, as part of our experiment, we also surveyed job applicants to collect data on salary expectations for the experimental job, and for an otherwise identical hypothetical job with different flexibility conditions, to measure their valuation of job flexibility based on within-individual

[50] If we instead using the ratio of equation (17) to equation (18), then the condition is $\phi\left(\tau w + \theta f - \frac{c}{p}\right) \cong \phi\left(\tau w - \frac{c}{p}\right)$.

[51] For example, Lang and Majumdar (2004) adopt this setup to model market imperfections when compensation is multidimensional and workers' tastes are heterogeneous.

comparisons. We were particularly interested in whether the evidence from this exercise matches the experimental evidence – which it does, broadly speaking (see Online Appendix F). Since this survey uses a within-person design, it can identify how individuals are willing to trade off wages for flexibility, for those who applied for particular jobs, although, as with the study by Mas and Pallais (2017), this kind of experimental design necessarily has to be presented to participants as contrived. Although our survey led to a small and non-random sample, the fact that we found similar large relative valuations of job flexibility suggests that – despite the potential challenges discussed in this section – we obtain reliable evidence on how workers value job flexibility.

5. Conclusions

We explore evidence on workers' valuation of the flexibility of jobs offered by employers, using data from a field experiment on a Chinese job board. In the experiment, we generate random variation across job seekers in invitations to apply for jobs that differ in terms of job flexibility conditions regarding both when one works and where one works. The experimental data allow the estimation of workers' valuation of job flexibility free of the potential biases that underlie most existing research using non-experimental data, as well as potential biases from using hypothetical scenarios in the stated preference approach.

Our experiment provides strong evidence that workers value job flexibility. Across the experimental variation in job flexibility conditions that our subjects receive, they are much more likely to apply for flexible jobs, conditional on the salary offered, and job seekers are willing to pay more (in the form of accepting lower pay) for flexible jobs. Moreover, we have shown that under some conditions that might not be regarded as too restrictive, our evidence is informative about job seekers' willingness to pay for flexible jobs of the types offered in the experiment, and more specifically indicates that workers value some of the flexible job conditions more than the higher salaries we offered them. The methods we develop and use in this paper can likely be applied fruitfully to understanding workers' valuation of job amenities generally.

References

Ai, C., and Norton, E. C. (2003). Interaction terms in logit and probit models. Economics Letters, 80(1): 123-129.

Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2015). Does working from home work? Evidence from a Chinese experiment. Quarterly Journal of Economics 130(1): 165-218.

Bonhomme, S., and Jolivet, G.. (2009). The pervasive absence of compensating differentials. Journal of Applied Econometrics 24(5): 763-95.

Brown, C. (1980). Equalizing differences in the labor market. Quarterly Journal of Economics 94(1): 113-34.

Caillier, J. G. (2012). The impact of teleworking on work motivation in a U.S. federal government agency. American Review of Public Administration 42(4): 461-80.

Chen, M. K., Rossi, P. E., Chevalier, J. A., and Oehlsen, E. (2019). The value of flexible work: Evidence from Uber drivers. Journal of Political Economy 127(6): 2735-94.

Delaney, J. T., and Huselid, M. A. (1996). The impact of human resource management practices on perceptions of organizational performance. Academy of Management Journal, 39(4): 949-69.

DIDI. (2018). 2017 China Urban Travel Report.

Eaton, S. C. (2003). If you can use them: Flexibility policies, organizational commitment, and perceived performance. Industrial Relations: A Journal of Economy and Society, 42(2): 145-167.

Eldridge, L. P., and Pabilonia, S. W. (2008). Are those who bring work home really working longer hours? Productivity Measurement and Analysis, OECD, 179-209.

Eldridge, L., and Pabilonia, S. W. (2010). Bringing work home: implications for BLS productivity measures. Monthly Labor Review, 133(12): 18-35.

Elsbach, K. D., Cable, D. M., and Sherman, J. W. (2010). How passive 'face time' affects perceptions of employees: Evidence of spontaneous trait inference. Human Relations, 63(6): 735-60.

Eriksson, T., and Kristensen, N. (2014). Wages or fringes? Some evidence on trade-offs and sorting. Journal of Labor Economics, 32(4): 899-928.

European Foundation for the Improvement of Living and Working Conditions. (2012). 5th European Working Conditions Survey. Publications Office of the European Union.

Flory, J. A., Leibbrandt, and A., List, J. A. (2015). Do competitive workplaces deter female workers? A large-scale natural field experiment on job-entry decisions. Review of Economic Studies, 82(1): 122-55.

Galinsky, E., Bond, J.T., Sakai, K., Kim, S.S., and Giuntoli, N. (2008). 2008 National Study of Employers. New York, NY: Families and Work Institute.

Goldin, C. (2014). A grand gender convergence: Its last chapter. American Economic Review, 104(4): 1091-119.

Harrison, G. W., and List, J. A. (2004). Field experiments. Journal of Economic Literature, 42(4): 1009-55.

Hedblom, D., Hickman, B. R., and List. J. A. (2019). Toward an understanding of corporate social responsibility: theory and field experimental evidence. National Bureau of Economic Research Working Paper No. 26222.

Horton, J. J. (2017). Price floors and employer preferences: evidence from a minimum wage experiment. CESifo Working Paper Series No. 6548.

Huselid, M. A., Jackson, S. E., and Schuler, R. S. (1997). Technical and strategic human resources management effectiveness as determinants of firm performance. Academy of Management Journal, 40(1): 171-88.

Konrad, A. M., and Mangel, R. (2000). Research notes and commentaries the impact of work-life programs on firm productivity. Strategic Management Journal, 21(12): 1225-37.

Kossek, E. E., and Dyne, L. V. (2008). Face-time matters: A cross-level model of how work-life flexibility influences work performance of individuals and groups. In Korabik, Lero, and Whitehead, eds., Handbook of Work-Family Integration, 305-30, Amsterdam: Elsevier.

Kniesner, T., K. Viscusi, and Woock, C. (2012). The value of a statistical life: evidence from panel data. The Review of Economics and Statistics, 94(1): 74-87.

Lang, K., and Majumdar, S. (2004). The pricing of job characteristics when markets do not clear: Theory and policy implications. International Economic Review, 45(4): 1111-28.

Lavetti, K. (2017). The estimation of compensating wage differentials: Lessons from the deadliest catch. Ohio State University, Working Paper.

Lavetti, K., and Schmutte, I. (2017). Estimating compensating wage differentials with endogenous job mobility. Ohio State University, Working Paper.

Lyness, K. S., Gornick, J. C., Stone, P., and Grotto, A. R. (2012). It's all about control: Worker control over schedule and hours in cross-national context. American Sociological Review, 77(6): 1023-49.

Maestas, N., Mullen, K. J., Powell, D., von Wachter, T., and Wenger, J. (2018) The value of working conditions in the United States and implications for the structure of wages. NBER Working Paper 25204.

Mas, A., and Pallais, A. (2017). Valuing alternative work arrangements. American Economic Review, 107(12): 3722-59.

Mateyka, P. J., Rapino, M., and Landivar, L. C. (2012). Home-based workers in the United States: 2010. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.

MSNBC. (2008). Gas prices encourage telecommuting. http://www.msnbc.msn.com/id/25007346/, accessed on February 2, 2017.

Perlow, L. A. (1997). Finding Time: How Corporations, Individuals, and Families Can Benefit from New Work Practices. Ithaca, NY: Cornell University Press.

Possenriede, D., Hassink, W., and Plantenga, J. (2014). Does face-time affect your career? Discussion Paper Series/Tjalling C. Koopmans Research Institute, 14(10).

Reynolds, B. W. (2017). The top 10 career fields for flexible jobs. Flexjob.com, October 6, 2017. https://www.flexjobs.com/blog/post/top-10-career-fields-for-flexible-jobs/, accessed on September 10, 2018.

Riedmann, A., van Gyes, G., Roman, A., Kerkhofs, M., and Bechmann, S. (2010). European company survey 2009: Overview. European Foundation for the Improvement of Living and Working Conditions.

Sorkin, I. (2018). Ranking firms using revealed preference. Quarterly Journal of Economics, 133(3): 1331-93.

Sullivan, P., and To, T. (2014). Search and non-wage job characteristics. Journal of Human Resources, 49(2): 472-507.

Taber, C., and Vejlin, R. (2016). Estimation of a Roy/search/compensating differential model of the labor market. National Bureau of Economic Research Working Paper No. 22439.

Wiswall, M., and Zafar, B. (2017). Preference for the workplace, investment in human capital, and gender. Quarterly Journal of Economics, 133(1): 457-507.

Woodcock, S. (2008). Wage differentials in the presence of unobserved worker, firm, and match heterogeneity. Labour Economics 15: 771-793.

Table 1. Order of treatments by position and week, application stage of experiment

	Week 1	Week 2	Week 3	Week 4
Java engineer	*TimeFlex*	*FullFlex*	*PlaceFlex*	*NoFlex*
Financial executive	*NoFlex*	*PlaceFlex*	*FullFlex*	*TimeFlex*
Human resource manager	*PlaceFlex*	*NoFlex*	*TimeFlex*	*FullFlex*
Marketing executive	*FullFlex*	*TimeFlex*	*NoFlex*	*PlaceFlex*
Sales executive	*PlaceFlex*	*NoFlex*	*FullFlex*	*TimeFlex*

Table 2. Summary of the experimental treatments

		Place flexibility	
		Regular work place	Flexible place
Time flexibility	Regular work time	*NoFlex*	*PlaceFlex*
	Flexible time	*TimeFlex*	*FullFlex*
	Additional control	No job ad email or app message sent	

Notes: For each job position in each treatment, one third of the subjects were randomly assigned monthly salary offer ranges of 10,000-15,000 CNY, 15,000-20,000 CNY, or 20,000-25,000 CNY.

Table 3. Applications by treatment

Full sample

	Treatment of job ads applied for					
	NoFlex	*TimeFlex*	*PlaceFlex*	*FullFlex*	Total	Application rate
Treatment of job ad emails and app messages sent						
NoFlex	45	4	2	5	56	0.24%
TimeFlex	1	98	3	4	106	0.40%
PlaceFlex	2	7	74	5	88	0.38%
FullFlex	1	4	1	113	119	0.45%
Not treated	69 (4)	103 (3)	93 (5)	108 (3)	373	(0.06%)
Total	118	216	173	235	742	

For each offered salary level

Low salary	*NoFlex*	*TimeFlex*	*PlaceFlex*	*FullFlex*	Total	Application rate
Treatment of job ad emails and messages sent						
NoFlex	14	2	1	2	19	0.24%
TimeFlex	0	22	0	0	22	0.25%
PlaceFlex	0	0	24	2	26	0.34%
FullFlex	0	2	0	30	32	0.37%
Total	14	26	25	34	99	0.30%
Medium salary	*NoFlex*	*TimeFlex*	*PlaceFlex*	*FullFlex*	Total	Application rate
Treatment of job ad emails and app messages sent						
NoFlex	12	0	0	0	12	0.15%
TimeFlex	1	29	2	2	34	0.39%
PlaceFlex	1	5	19	2	27	0.35%
FullFlex	1	1	1	45	48	0.55%
Total	15	35	22	49	121	0.37%
High salary	*NoFlex*	*TimeFlex*	*PlaceFlex*	*FullFlex*	Total	Application rate
Treatment of job ad emails and app messages sent						
NoFlex	19	2	1	3	25	0.32%
TimeFlex	0	47	1	2	50	0.57%
PlaceFlex	1	2	31	1	35	0.46%
FullFlex	0	1	0	38	39	0.45%
Total	20	52	33	44	149	0.45%

Notes: Application rate is computed as the number of applications divided by the number of sampled job seekers who were sent job ad emails or app messages. Since we cannot track who successfully received the email or app message at the individual level, we use all job seekers intended to be treated as the denominator. There are 373 applications coming from job seekers who did not receive email or app message but saw the job ad on the job board, including 15 from the additional control group, whose application number and rate are indicated in parentheses. We consider these applications as coming from job seekers who were "not treated."

Table 4. Treatment effect on applications

Dependent variable: Apply or not	(1)	(2)	(3)	(4)
TimeFlex	0.0018***	0.0012***	0.0011**	0.0002
	(0.0005)	(0.0004)	(0.0004)	(0.0007)
PlaceFlex	0.0013***	0.0011**	0.0010**	0.0010
	(0.0005)	(0.0004)	(0.0005)	(0.0008)
FullFlex	0.0024***	0.0019***	0.0019***	0.0012
	(0.0005)	(0.0004)	(0.0005)	(0.0008)
Medium		0.0003	0.0005	-0.0002
		(0.0004)	(0.0004)	(0.0007)
High		0.0012***	0.0016***	0.0010
		(0.0004)	(0.0004)	(0.0008)
TimeFlex × Medium				0.0010
				(0.0010)
TimeFlex × High				0.0021*
				(0.0012)
PlaceFlex × Medium				-0.0002
				(0.0011)
PlaceFlex × High				0.0001
				(0.0013)
FullFlex × Medium				0.0020*
				(0.0012)
FullFlex × High				0.000004
				(0.0012)
Finance		0.0043***	0.0041***	0.0040***
		(0.0005)	(0.0006)	(0.0006)
HR		0.0023***	0.0027***	0.0027***
		(0.0004)	(0.0005)	(0.0005)
Marketing		0.0022***	0.0014***	0.0013***
		(0.0005)	(0.0004)	(0.0004)
Sales		0.0029***	0.0012***	0.0012***
		(0.0005)	(0.0004)	(0.0004)
Job seeker characteristics	No	No	Yes	Yes
Number of job seekers	99,146	99,146	58,151	58,151
Estimated differences				
(i) *PlaceFlex* − *TimeFlex*	-0.0005	-0.0001	-0.0002	
	(0.0005)	(0.0005)	(0.0005)	
(ii) *FullFlex* − *TimeFlex*	0.0006	0.0007	0.0008	
	(0.0005)	(0.0005)	(0.0005)	
(iii) *FullFlex* − *PlaceFlex*	0.0011*	0.0008*	0.0009*	
	(0.0005)	(0.0005)	(0.0005)	
(iv) *FullFlex* × Low − *NoFlex* × Medium				0.0015**
				(0.0007)
(v) *FullFlex* × Low − *NoFlex* × High				0.0003
				(0.0009)
(vi) *FullFlex* × Medium − *NoFlex* × High				0.0020**
				(0.0010)

Notes: This table reports marginal effects from a probit model. Estimates for control variables are not reported; the full regression results are shown in Appendix Table G4. The marginal effects for the interactions are computed using difference-in-differences involving the cumulative normal distribution functions expressions, to recover the correct marginal effects with interactions in a probit model (Ai and Norton, 2003). In the "Estimated differences" panel, we report differences between marginal effects. The sample includes only sampled job seekers who were sent job ad emails and app messages. All applications not made to the sent treatments or positions are excluded. Job seekers whose individual characteristics have inconsistencies or missing values are also excluded. Robust standard errors allowing for heteroskedasticity are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Appendix Table G5 reports the summary statistics for the individual characteristics of job seekers and the treatments for the sample.

Table 5. Heterogeneity in treatment effect on applications by gender and marital status

Dependent variable: Apply or not	Marginal effects	Standard errors
TimeFlex	0.0017	(0.0015)
PlaceFlex	0.0012	(0.0015)
FullFlex	0.0006	(0.0014)
Female	-0.0014	(0.0011)
Female × *TimeFlex*	-0.0017	(0.0016)
Female × *PlaceFlex*	-0.0011	(0.0016)
Female × *FullFlex*	-0.0003	(0.0015)
Married	0.0021	(0.0014)
Married × *TimeFlex*	0.0002	(0.0022)
Married × *PlaceFlex*	-0.0016	(0.0021)
Married × *FullFlex*	0.0036	(0.0023)
Female × Married	-0.0020	(0.0015)
Female × Married × *TimeFlex*	0.0016	(0.0024)
Female × Married × *PlaceFlex*	0.0042*	(0.0023)
Female × Married × *FullFlex*	-0.0009	(0.0025)
Medium	0.0005	(0.0005)
High	0.0016***	(0.0005)
Finance	0.0040***	(0.0006)
HR	0.0026***	(0.0005)
Marketing	0.0013***	(0.0004)
Sales	0.0011***	(0.0004)
Job seeker characteristics	Yes	
Number of job seekers	58,151	
Estimated differences		
(i) Unmarried female: *TimeFlex* − *NoFlex*	-0.00003	(0.0005)
(ii) Unmarried female: *PlaceFlex* − *NoFlex*	0.0001	(0.0006)
(iii) Unmarried female: *FullFlex* − *NoFlex*	0.0003	(0.0005)
(iv) Married male: *TimeFlex* − *NoFlex*	0.0019	(0.0016)
(v) Married male: *PlaceFlex* − *NoFlex*	-0.0004	(0.0014)
(vi) Married male: *FullFlex* − *NoFlex*	0.0042**	(0.0018)
(vii) Married female: *TimeFlex* − *NoFlex*	0.0018**	(0.0007)
(viii) Married female: *PlaceFlex* − *NoFlex*	0.0027***	(0.0008)
(ix) Married female: *FullFlex* − *NoFlex*	0.0030***	(0.0008)

Notes: This table reports marginal effects from a probit model. Estimates for control variables are not reported; the full regression results are shown in Appendix Table G7. The marginal effects for the interactions are computed using difference-in-differences (or triple differences) involving the cumulative normal distribution functions expressions, to recover the correct marginal effects with interactions in a probit model (Ai and Norton, 2003). Standard errors are computed using the delta method. The sample includes only sampled job seekers who were sent job ad emails and app messages. All applications not made to the sent treatment or positions are excluded. Job seekers whose individual characteristics have inconsistencies or missing values are also excluded. Robust standard errors allowing for heteroskedasticity are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6. Treatment effect on applications for married females

Dependent variable: Apply or not	
TimeFlex	0.0002
	(0.0007)
PlaceFlex	0.0017
	(0.0011)
FullFlex	0.0003
	(0.0007)
Medium	0.00004
	(0.0007)
High	0.0004
	(0.0009)
TimeFlex × Medium	0.0010
	(0.0011)
TimeFlex × High	0.0027*
	(0.0015)
PlaceFlex × Medium	0.0000
	(0.0016)
PlaceFlex × High	0.0011
	(0.0018)
FullFlex × Medium	0.0035**
	(0.0016)
FullFlex × High	0.0025
	(0.0016)
Finance	0.0021**
	(0.0010)
HR	0.0012
	(0.0009)
Marketing	-0.0005
	(0.0009)
Sales	-0.0008
	(0.0009)
Job seeker characteristics	Yes
Number of job seekers	18,706
Estimated differences	
(i) *FullFlex* × Low − *NoFlex* × Medium	0.0003
	(0.0007)
(ii) *FullFlex* × Low − *NoFlex* × High	-0.0001
	(0.0009)
(iii) *FullFlex* × Medium − *NoFlex* × High	0.0034**
	(0.0015)

Notes: This table reports marginal effects from a probit model. Estimates for control variables are not reported; the full regression results are shown in Appendix Table G8. The marginal effects for the interactions are computed using difference-in-differences (or triple differences) involving the cumulative normal distribution functions expressions, to recover the correct marginal effects with interactions in a probit model (Ai and Norton, 2003). Standard errors are computed using the delta method. The sample includes only sampled married female job seekers who were sent job ad emails and app messages. All applications not made to the sent treatments or positions are excluded. Job seekers whose individual characteristics have inconsistencies or missing values are also excluded. Robust standard errors allowing for heteroskedasticity are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.