Labor Market Discrimination against Family Responsibilities:

A Correspondence Study with Policy Change in China

Haoran He[†], Sherry Xin Li[‡], Yuling Han[*]

Abstract

China shifted its controversial one-child policy (1979–2015) to a two-child policy in 2016. We take advantage of this unexpected policy change and the heterogeneities in the pre-change environment to investigate labor market discrimination against expected family responsibilities. In a two-wave correspondence study before and after the policy change, we sent 8,848 fictitious resumes aged 22 to 29 in response to online job advertisements. Their gender and only-child/siblinged status were systematically varied. We find that women, but not men, are subject to labor market discrimination for expected family responsibilities. This discrimination worsens with the increase in women's reproductive age.

Keywords: labor market discrimination, parenthood penalty, correspondence study, field experiment, one-child policy, China

JEL Classification: C93, J71, M51

Acknowledgment: We are grateful to the two anonymous reviewers for their very constructive suggestions and comments. We thank Siwei Cao, Amy Farmer, Fabio Galeotti, Jie Gong, Elizabeth Hoffman, Andrew Horowitz, Raja Kali, Chuliang Luo, Charles Noussair, David Neumark, Maria Micaela Sviatschi, Marie Claire Villeval, Qian Weng, and seminar participants at the Beijing Normal University, Central University of Finance and Economics, Renmin University of China, Tsinghua University, University of Arkansas, Iowa State University, Florida State University, Fudan University, Rensselaer Polytechnic Institute, the 6th SEBA-GATE international workshop, the 2016 international ESA, the 2017 North America ESA, the 2018 Asia-Pacific ESA, the 2018 NBER Law and Economics Summer Institute, and the 2018 Methods of Experimental Economics Research Workshop at Texas A&M University for helpful comments and discussions. We also thank Wanyu Jiang, Lunwen Qi, Jingwen Xia, Huikun He, Zhenglin Sun, Xingyu Zhong, Lijia Wang, Yumei Peng, Han Huang, Wenqi Guan, Bin Wei, Xuehan Yang, Yuqi He, Jiaqian Song, Wenxin Yan, Zhen Huang, Tenghui Wang, Jing Pan, and Chao Wang for excellent research assistance. Financial support from the National Natural Science Foundation of China (Projects No. 71973016 & 72131003) and the Beijing Natural Science Foundation (9192013) is gratefully acknowledged.

[†] Business School, Beijing Normal University, Beijing 100875, China. E-mail address: haoran.he@bnu.edu.cn; Tel.: 86-10-58807874.

[‡] Department of Economics, University of Arkansas, Fayetteville, AR 72701, USA. E-mail address: SLi@walton.uark.edu. Tel.: 1-479-5756222.

[*] Business School, Beijing Normal University, Beijing 100875, China. E-mail address: summer.han@mail.bnu.edu.cn.

Labor Market Discrimination against Family Responsibilities: A Correspondence Study with Policy Change in China

1 Introduction

A large and fruitful body of labor economics research uses audit or correspondence study methods and encompasses various aspects of labor market discrimination, including race, ethnicity, gender, age, unemployment status, and disability (Riach and Rich, 2002; Rich, 2014; Bertrand and Duflo, 2017; Neumark, 2018). However, one important area that has remained understudied is labor market discrimination based on expected family responsibilities or parenthood. Employers may engage in labor market discrimination against workers of childbearing age, most likely women, in the expectation that they are more likely to leave their jobs due to childbearing and childcare responsibilities.

Family responsibilities or childbearing intentions are often unobservable by employers in the early hiring stages. Researchers use indirect and direct approaches to tackle this challenge. The indirect approach uses applicants' other related characteristics, e.g., marital status (Maurer-Fazio and Wang, 2018), age (Petit, 2007; Helleseter et al., 2020), membership of parent's organizations (Correll et al., 2007), or sexual orientation (Baert, 2014), as proxies for their childbearing intentions and responsibilities. The evidence generally supports the existence of discrimination against those who (might) have family responsibilities. Several studies find no evidence of discrimination in job interviews based on gender, age, and marital status (e.g., Albert et al., 2011) or gender and parenthood status (e.g., Bygren et al., 2017). Becker et al. (2019) take a direct approach and explicitly signal fictitious applicants' marital status and children's ages. They find supportive evidence of discrimination in hiring based on potential and realized fertility.[1] One caveat of these studies pertains to their cross-sectional design in which employers receive multiple signals infrequently seen in job applications. It may raise concerns of potential confounds and external validity because employers may perceive such signals as unusual and react differently (Balfe et al., 2021).

In this paper, we conduct a two-wave correspondence study and take advantage of the heterogeneities in China's birth policy in 2014–2015 and the unexpected termination of the one-child policy in 2016 to investigate labor market discrimination based on expected family responsibilities. Our approach also necessitates the experimental intervention of disclosing an otherwise uncommon signal—job applicants' sibling status. However, we minimize the concerns that might arise and thus improve upon the previous literature by further leveraging the *temporal* variations provided by the unexpected 2016 policy change.

China's one-child policy, enacted in 1979, created a general restriction of one child per family. However, a major amendment implemented in 2014 and 2015 allowed *two children* for qualified families in which *at least one parent was an only child*. This amended policy, conditioning birth restrictions to one's sibling status (i.e., whether a person was an only child or had siblings), led to higher expected fertility for only-child adults relative to their siblinged counterparts (see a visual representation in Figure 1 and more details in Section 2). Such policy heterogeneities made us interested in researching family-related discrimination using the cross-sectional variations by individuals' *sibling status* and *gender*.

[Figure 1 about here]

On October 29, 2015, during our data collection, the government unexpectedly announced the shift to a universal two-child policy. It allowed all families to have up to two

[1] A related literature also shows that women are subject more vulnerable to the parenthood wage penalty relative to men (e.g., Anderson et al., 2002; Benard and Correll, 2015; Kleven et al., 2019).

children, delinking limitations on children per family from the parents' sibling status, effective January 1, 2016. This new policy increased the expected fertility for siblinged men and women (hence the treatment groups) but would not affect the only-child adults (hence the control groups). The field-leveling policy change offered us a unique opportunity to build a two-period design upon our original cross-sectional design. We thus combined a two-wave field experiment with a natural experiment (the policy change) to draw *causal* inferences on family-related labor market discrimination. This approach was first introduced in Agan and Starr (2017) to study discrimination against job applicants with criminal records.

We conducted this experiment in China's three most economically advanced megacities *before* and *after* the 2016 policy change. We sent 8,848 fictitious resumes with age between 22 and 29 in response to actual online job advertisements. We sent four resumes for each listing and systematically varied the fictitious applicants' gender and sibling status to signal expected fertility. We begin with a pre-policy cross-sectional difference-in-differences (DID) approach and find that when only-child adults were allowed to have one more child than siblinged adults before 2016, only-child women received differentially fewer callbacks than their siblinged female counterparts, relative to men. We then extend the analysis to temporal DIDs that leverage the policy change for more rigorous causal identification. We find that siblinged women, once allowed an additional child due to the 2016 policy change, received differentially fewer callbacks than the only-child women whose legal fertility remained unchanged. This impact is absent for men. We also construct a triple-differences estimator based on the temporal DIDs for women and men to difference out temporal variations that would similarly affect women and men. The results show that the *adverse change* resulting from the new policy in the callback difference between siblinged and only-child women is *worse* than the corresponding change between siblinged and only-child men. These findings provide strong evidence that women, but not men, are negatively affected by expected family responsibilities. Moreover, we find that age moderates such discrimination against women since our results reveal more pronounced discrimination against siblinged women in late 20s, the upper age limit in our sample. This finding coincides the deeply entrenched social norms and highly likely employers' beliefs that women of this age range are more likely to have "pent-up" demand for children than those under 25. The lack of this age pattern for men accentuates that family-based discrimination is a phenomenon directed at women.

To our knowledge, our study is one of the few, besides Agan and Starr (2017), that use pre- and post-policy field experiments to draw causal inferences regarding discrimination. It is also the first to apply such an approach to labor market discrimination based on family responsibilities. Our temporal DIDs and triple-differences strategies leveraging a policy and its exogenous change offer a more rigorous causal identification and improve previous literature built on cross-sectional variations in single-wave correspondence studies. We also demonstrate how to utilize exogenous, temporal sources of variations, such as a policy change, to alleviate methodological concerns related to signaling uncommonly seen individual traits in field experiments on discrimination (Balfe et al., 2021). In addition, we focus on the effect on labor market performance of the *possibility* of having one or two children (rather than the number of children that a person actually has), which is an important but understudied area noted by Bertrand and Duflo (2017, p. 17). Last but not the least, our study offers timely caution regarding China's most recent policy change in May 2021 that further relaxed the birth limits to allow *three* children per family. We will discuss policy implications in the conclusion.

2 Background

This section offers background information on China's birth policy from 2014 to 2016. We focus on the link between birth limits and an adult's sibling status in 2014 and 2015—which motivated our design—and the removal of this link after the 2016 policy change.

China's one-child policy went through multiple amendments since being enacted in 1979 to curb its population growth (see Appendix A). A major amendment, implemented in 2014 and 2015, *allowed two children per family if at least one parent was an only child*, but the one-child restriction remained for couples who were both siblinged. How did this seemingly preferential birth policy for only-child adults affect their job prospects compared to their siblinged counterparts? This was our initial question for our cross-sectional correspondence study which began in 2015, which turned out to be the first wave of our experiment.

On October 29, 2015, the government announced the shift to a universal two-child policy: all married couples could have two children starting on January 1, 2016. This radical change ended the decades-old, highly controversial one-child policy to address increasingly concerning socioeconomic issues such as a vast aging population and a severely skewed sex ratio. It made headlines instantly on all news networks and social media.

Despite the general expectation that the one-child policy would eventually end, the timing of its termination was completely unanticipated by the public, including us, the researchers.[2] Compared to the 2014–2015 policy, the 2016 policy change eliminated the dependence of allowed children on sibling status. It ended the discriminatory treatment in favor of only-child adults and against those with siblings. Note while this policy change

[2] See Appendix B for the evidence and discussions on the unexpectedness of this policy change. Even if anticipation was a concern, it would bias against us finding discrimination, and hence our results serve as *lower* bounds for the true degree of labor market discrimination based on family responsibilities.

rendered everyone equal in terms of reproductive rights, it directly affected those siblinged but not only-child adults. Figure 1 illustrates the position of each group in the pre- and post-2016 policy environments. The only-child men and women (hereafter MO and FO) were exempt from the one-child restrictions as early as 2014. However, men and women with siblings (MS and FS), until 2016, could have no more than one child unless their spouses were only children. Figure 1 highlights this change by contrasting the MSB/FSB to MSA/FSA groups (B and A designating before or after the policy change); the only-child groups saw no change (MOB/FOB vs. MOA/FOA). Hence, this policy change allowed us to expand our initial cross-sectional correspondence study to a two-wave design, as detailed in Section 3.

Evidence in the demographic literature indicates that the pre-2016 birth policy was binding for siblinged women of childbearing age or couples. A large-scale survey conducted by China's National Health and Family Planning Commission in 29 provinces in 2013 showed that the average preferred number of children was 1.79, 1.83, and 1.95 for couples where both were only children, one was an only child, or both were siblinged (Zhuang et al., 2014). China's Inter-Census Population Survey (2005), extrapolated to 2015, shows that 75.5% of men and 81.5% of women had a sibling(s) in the nation's 20–40 age group.[3] Therefore, the 2016 policy released a long-suppressed desire among many siblinged adults to have more children and may have potentially affected 90 million families (Reuters, 2015).

[3] This age group was mostly born after the inception of the one-child policy and were more likely to be influenced by the 2016 policy than other age groups. The sibling status composition was calculated based on the 10-30 age group in China's Inter-Census Population Survey in 2005. The existence of a large number of siblinged children despite the one-child policy may be due to the exemptions provided for ethnic minorities, policy relaxations that allowed a second child for qualified families, or enforcement challenges in rural areas.

The pent-up demand for a second child is evident in the national birth statistics and recent research. Data from the Chinese Bureau of Statistics (2015–2018) show that the birthrate of the *second* child in urban China among women aged between 15 and 49 increased dramatically from 8.13‰ in 2014 and 8.87‰ in 2015 to 12.32‰ in 2016 and 22.06‰ in 2017, while the first-child birth rate remained stable. Chen (2019) finds that the increase in the fertility rate of the second child was higher in urban than in rural areas and is particularly high for women with an associate's degree or higher.

The 2016 policy change bore importance for understanding how employers would update their beliefs toward their employees' expected family responsibilities. On the one hand, employers should infer no difference in the expected long-term family size of siblinged and only-child employees since both groups can have up to two children. On the other hand, employers may worry about an imminent increase in childbearing possibilities of siblinged women relative to those only-child women during employment.[4] If this were the case, we would expect a higher level of discrimination against siblinged women after the policy change, especially those who are somewhat older. Employers may try to avoid the costs associated with hiring them. We will discuss the repressed demand further in Section 5.

3 Experimental Design

Following the unanticipated shift to the two-child policy in January 2016, we extended our initial cross-sectional correspondence study to a two-wave experimental design before and

[4] Women receive better family-leave terms than men in China. The national maternity-leave policy entitles female employees to 98 days of maternity leave and no less than one hour of nursing time daily during workdays within their newborns' first year (China's State Council, 2012). No national policies stipulate such leave for men. All provincial-level administrations allow a more generous maternity leave—between 128 and 188 days (with a special term of one year in Tibet)—than the national standard. Some provinces allow women to take additional breastfeeding leave if employers agree. In contrast, paternity leave varies between 7 and 30 days at the provincial level, with 15 days in most provinces.

after the shift. We conducted the first wave from May to September 2015 and the second wave from January to July 2016, excluding the Chinese New Year (see Figure A1 of Appendix A) in China's three most economically advanced megacities, Shanghai, Guangzhou, and Shenzhen.[5] Appendix C details the experimental procedures.

In the experiment, we sent fictitious resumes to real, online job ads on 51job.com (mimicking "I-want-job.com" in the Chinese pronunciation; Nasdaq: JOBS), a nationally leading job board in China that includes primarily entry-level jobs. Many Chinese online job boards, including 51job.com, provide applicants with resume templates (Figure A3 in Appendix C). In addition to providing the standard information such as name, gender, date of birth, education background, and work experience, applicants must complete a brief *self-assessment*. This allows us to implement a two-wave, 2 (men/women) × 2 (high/low expected fertility) within-subject factorial design *before* and *after* the 2016 policy change. To manipulate the gender information, we complete the required gender information and use two Chinese-character, gender-salient names on the resumes. To signal expected fertility, we disclose sibling status on the fictitious resumes since, in China, it is widely known that under the 2014–2015 policy regime, the allowed number of children was tied to sibling status, which became irrelevant after 2016.

Specifically, we embed the sibling-status signal in the self-assessment section of the resume template (see Figure A3 of Appendix C) in which applicants introduce themselves as one might in a cover letter for a job application in the U.S. Our designed assessment starts with, "[a]s the only child (or one of the children) in my family …," followed by a personal statement on traits reflective of a good employee.[6] We reveal no information on the resume

[5] The policy change was announced in the middle of our first-wave study in Beijing (see more in Appendix C).
[6] We conducted a manipulation check to confirm the salience of the sibling-status signal (Appendix D).

of siblinged applicants of the number of siblings or birth order. Therefore, for fictitious applicants with siblings, such information is provided to signal that they are *not* only children, rather than whether they have a sibling(s).

We based our self-assessment design on a close examination of 3,000 real resumes drawn randomly from the job board's resume pool accessible by employers. We found it was *not* unusual for job applicants to self-disclose their sibling status—about 7% of the 3,000 resumes did so, 62% of which were only-child applicants. In addition, most resumes with such self-disclosure focused on the applicants' positive employee attributes (e.g., hardworking, confident, conscientious, motivated, reliable, team-player, and having good interpersonal skills). These keywords or phrases were collected to form a bank of 16 statements (Appendix Table A2). We categorized these statements into four 4-statement bins, randomly selected one statement from each bin with no replacement and combined them into four self-assessments, one for each of our four resume types (MO/FO/MS/FS).[7]

In correspondence studies on discrimination, researchers send employers signals on certain attributes of research interest by randomly assigning self-disclosed characteristics to the fictitious applicants. Some of these characteristics, such as sibling status in our study, do not often appear on real-life job applications and may lead to confounds.[8] Fortunately, our difference-in-differences and triple-differences approaches that leverage the policy change can substantially alleviate the concerns with this signal, as detailed in Section 4.[9]

[7] See a self-assessment example, the summary on our observations on 3,000 real resumes, and the bank of the statements and the detailed procedure to form our self-assessments in Appendix C.
[8] For example, Tilcsik (2011) discloses sexual orientation through a volunteer experience in a gay campus organization. Ameri et al. (2018) signal disability status on a cover letter by mentioning the applicants' volunteer work. Namingit et al. (2021) reveal a previous illness-related employment gap by mentioning involvement in a cancer-survivors support group on a cover letter.
[9] One question is whether merely signaling one's sibling status on the resumes would affect employers' decisions on callback, irrespective of the fertility issue. We investigated this question by conducting a companion experiment in Shanghai in 2016 using a separate but comparable sample on 51job.com. As detailed

The job board mainly comprises entry-level jobs for college graduates. Most require a three- or four-year college education and stipulate up to 5 years of work experience. Accordingly, our fictitious applicants held either a three-year associate's degree or a four-year bachelor's degree and had between 0 (exclusive) and five (inclusive) years of work experience based on specific requirements of the jobs. We adapted work experience and job skills from real resumes used for our targeted industries and occupations to make the fictitious applications appear authentic. For each job, we calculated age, a required field on the resume template, to match the stipulated educational background and work experience. We then generated each fictitious applicant's date of birth accordingly. Overall, our fictitious applicants were between 22 and 29, above the minimum legal marriage age of 20 and within the primary childbearing window. Our resumes did not reveal marital status and the number of children since such information was not required or commonly revealed on real resumes.[10]

We focused on the two industries in each city with the most job listings in the two months preceding our experiment. These industries were internet and finance in Shanghai, internet and fast-moving consumer goods in Guangzhou, and internet and electronics in Shenzhen.[11] We chose the three most popular job categories—sales, administrative assistance, and customer service—for each industry.[12] The college majors included

in Appendix D, we find no differences in callback rates between resumes with sibling-status signals and those without; we also find no differential gender gaps in callback rates between those with and without the signals.

[10] Unlike early studies, we chose not to reveal applicants' marital or parenthood status. Revealing parenthood status or number of children would be perceived as unprofessional and even odd for our targeted age group. We chose not to reveal marital status since our main contribution pertained to discrimination against the expected rather than current family responsibilities. In addition, revealing marital status may introduce other confounds (e.g., married men may be perceived as being more stable and responsible) and potentially bias our results.

[11] Fast-moving consumer goods are those with a short shelf life because of high consumer demand (e.g., soft drinks and toiletries) or perishable nature (e.g., meat and dairy products). They sell quickly at a relatively low cost.

[12] These occupations were also used in previous studies, e.g., Bertrand and Mullainathan (2004) and Kroft et al. (2013).

accounting, business administration, Chinese, economics, finance, history, international economics and trade, marketing, and psychology, which were employers' preferred college majors or those often listed on real applications for the targeted industries and occupations.

We sent 1,106 resumes for each resume type in each wave of our study (Table 1). In each city, we maintained the job compositions across waves so that the industries and occupations represented were identical between the two waves. This design helped avoid cross-sectional variations inadvertently brought into the study due to, for example, natural changes in the market dynamics (e.g., more or fewer jobs became available and thus were drawn into our sample in an industry or occupation). Although we did not design our second wave to exclude employers in the first wave, in practice, these employers did not overlap across waves because of the extensive labor demand in these occupations and industries. Overall, in the two-wave study, we sent 8,848 fictitious resumes in response to 2,212 job ads with four resumes for each ad, FO, MO, FS, and MS, representing four applicants with unique identities who differed in gender and sibling status but otherwise had comparable credentials.

For each randomly selected, newly published job ad that met our selection criteria, we sent the four resumes in a random order, one on each day within four consecutive days. We defined success as receiving a callback (i.e., a phone call, email, or text message from the employers) within two weeks of submission requesting an interview with the fictitious applicants.[13,14] Our research assistants answered the callbacks and informed the recruiters that the (fictitious) applicants were no longer available.

---

[13] More than 98% of callbacks were received within two weeks in our pilot, consistent with Kroft et al. (2013).
[14] Researchers have different views on the extent to which callbacks may measure hiring discrimination (Button and Walker (2020) summarize this debate). Much evidence shows that callbacks capture an important part of hiring discrimination (e.g., Riach and Rich, 2002; Neumark et al., 2019). Some recent research, however, shows that low interview costs may lessen the extent to which callbacks can do so (Cahuc et al., 2019), and focusing on callbacks may miss a significant amount of discrimination after callbacks (Quillian et al., 2020). In our study, the online recruitment markets comprise primarily private sector jobs in competitive industries for which

4 Empirical Strategies

When we designed our experiment's first wave, we did not know that the one-child policy was about to be terminated. We initially planned to take advantage of the cross-sectional source of variation—only-child adults could have one more child than siblinged adults under the pre-2016 policy—to identify potential family-related discrimination. However, this empirical strategy may be subject to some unobservable differences between the only-child and siblinged groups, irrespective of fertility. For example, regardless of any fertility concerns, only children may be disadvantaged in the labor market by negative stereotypes such as being spoiled, selfish, lonely, maladjusted, less trusting and trustworthy (e.g., Falbo, 1979; Cameron et al., 2013).[15] Since no studies find noticeable gender differences in these stereotypes, and the literature shows that men are not subject to discrimination for family responsibilities (e.g., Anderson et al., 2002; Benard and Correll, 2015), we could compare callback differences between the two sibling-status groups across gender, that is, using a cross-sectional DID to estimate discrimination against women for family responsibilities (see a visual illustration in the left part of Figure 1). We hypothesize that *the only-child women have a differentially lower callback rate than the siblinged women, relative to men, before the 2016 policy change, i.e., (FOB – FSB) < (MOB – MSB).*

However, not all the unobservable differences between the sibling-status types are gender neutral. For instance, sibling status may reveal one's birth order differentially for men and women. Due to a deep-rooted son preference in China (e.g., Qian, 2008), rural households or

Cahuc et al. (2019) argue interview costs are likely to be high. Hence, their critique may not apply to our study, and we believe that interview callback is an important recruiting stage where discrimination may occur. Nonetheless, we acknowledge that interview callback may not represent the entirety of hiring discrimination.

[15] Decades-long research shows no consensus, however. Feng (2010) suggests that only children are likely to be demonized. Other research finds no, or very small, differences in their personalities and behaviors (e.g., Falbo and Polit, 1986).

migrant families are more likely to have an additional child, with or without government approval, if the firstborn is a girl. Hence, having a sibling for a man suggests that he is likely to be a later child, while for a woman, it suggests she is likely to be the firstborn or an older child. These gendered inferences for birth order may, in turn, correlate with (employers' perceptions of) employee traits such as leadership, sense of responsibility, and work ethics.

Moreover, our signal of sibling status, especially only-child status, may carry other unintended fertility-irrelevant signals from employers' perspective since this information is uncommon (although not rare) in job applications. For example, among the 3,000 real resumes that we studied in preparation, some applicants mentioned that they wanted a local, stable job or had business-travel constraints due to caregiving for their aging parents. Although eldercare is not a pervasive concern for the age range in our study, employers may still perceive that only-child employees or women are more likely to have this possible caregiving role than siblinged employees or men. These fertility-irrelevant, unobservable differences that may vary by sibling status or gender could potentially affect applicants' desirability in employers' eyes but are unlikely to cancel out in our cross-sectional DID.

The unexpected policy change in 2016 lifted the one-child limit for all siblinged adults (the treated groups) and placed their childbearing potential on par with their only-child counterparts (the control groups; see Figure 1). The new policy and its serendipitous timing offered us a clean identification strategy to cancel out the gender-specific unobservable differences between sibling-status types (e.g., birth order and employers' perceptions of signals) since these differences were unlikely to be affected by a sudden change in birth policy. Specifically, we can leverage this policy change and apply a temporal-DID approach, within each gender group, to the only-child and siblinged applicants before and after the policy change to develop a causal estimate for family-related labor market discrimination against women. The identifying assumption for the *gender-contingent temporal DIDs* is that within each gender group, absent the policy, the callbacks for the only-child and siblinged applicants would have shared parallel trends before and after 2016. We hypothesize that women, but not men, are subject to discrimination based on family responsibilities. Specifically, *siblinged women receive differentially fewer callbacks than their only-child, same-gender counterparts after, relative to before, the policy change in 2016, that is, (FSA – FOA) < (FSB – FOB); the gaps in callbacks between the siblinged and only-child men are not different before and after the policy change, that is, (MSA – MOA) = (MSB – MOB).*

However, the gender-specific temporal DIDs may still be susceptible to confounds due to other temporal differences, such as seasonal variations in the labor market. To address this concern, we combine both the cross-sectional and temporal sources of variations discussed above and use a *triple-differences* approach to compare the change in the callback difference due to the new policy for women relative to the change over the same period for men. We hypothesize that *the differential, negative impact of the policy change on siblinged applicants relative to their only-child counterparts is <u>worse</u> for women than for men*, i.e., *(FSA – FOA) – (FSB – FOB) < (MSA – MOA) – (MSB – MOB)*. This analysis requires the *relative* callbacks of female siblinged and only-child applicants to trend similarly to the *relative* callbacks of male siblinged and only-child applicants, absent the policy change. In other words, it imposes no restrictions on the specific patterns of time trends for the four types of applicants other than requiring no contemporaneous shocks that differentially affected callbacks for siblinged women relative to the other three groups. Therefore, the parallel trend assumption for triple differences is weaker than that for the temporal DIDs. The triple differences can eliminate any potential impact of other temporal changes that are irrelevant to the 2016 birth-policy change, provided that they do not influence the relative callbacks between the sibling-status types differentially for men and women. It can also difference out any fertility-unrelated, unobservable cross-sectional differences between the only-child and siblinged applicants as long as they are time invariant across the two waves of our experiment. Thus, this approach adds more rigor to the causal identification of discrimination against family responsibilities than the temporal DIDs.

5 Results

We now turn to hypotheses testing using a cross-sectional DID followed by temporal DIDs and triple differences that leverage the policy change. We will discuss the validity of the underlying assumptions and possible heterogeneities in the results by age and occupation.

Our experiment's overall average callback rate is 32.3% for the 8,848 resumes sent to the 2,212 job ads, 31.8% before and 32.8% after the policy change (p = 0.340, test of proportions).[16] Table 1 presents callbacks for each category and policy environment. We find no significant callback differences in any pairwise comparisons across sibling types or policies for men (p > 0.10). The patterns of callback are very different for women, however. Before the policy change, the callback rate of 30.1% for only-child women is significantly lower than 35.4% for siblinged women (FOB vs. FSB, p < 0.001, McNemar's test). The callback rate for only-child women increases significantly from 30.1% to 36.6% after the

[16] Our callback rate is higher than that in some earlier studies on China's online job boards, e.g., Maurer-Fazio (2012) and Maurer-Fazio et al. (2015), maybe due to a tight labor market, suggested by the expected recruitment rate (about 9%) in our study. This rate was calculated using the total number of applications received and the intended number of recruits published for 1,562 job ads in our sample. We also observed a one percentage point increase in callback in wave two, showing a somewhat stronger economy. If discrimination is higher in economic downturns, as previous work finds (e.g., Baert et al., 2015; Dahl and Knepper, 2020), our temporal DID and triple-differences estimates are likely to be a lower bound for the true amount of family-based labor market discrimination.

policy change (FOB vs. FOA, $p = 0.001$, unpaired test of proportions). After the policy change, the callback rates for the only-child and siblinged women are reversed (36.6% for FOA vs. 34.4% for FSA, $p = 0.063$, McNemar's test).[17]

[Table 1 about here]

5.1 Regression Results

Table 2 reports the OLS analyses on the determinants of callback. The dependent variable is whether the fictitious applicant receives a callback. The independent variables in Column [1] include the dummy variables for the experimental treatments (with the female only-child group before the policy change, FOB, as the omitted reference group) and the city fixed effects. Column [2] adds other applicant characteristics, such as age, education, years of work experience, the number of previous jobs, college major, and university fixed effects. We further add job- and firm-related characteristics in Column [3], including occupations, the number of applications for a job ad (as a proxy for the competition for this position), the number of people who followed the firm on the job board (as a proxy for the firm's popularity), firm size (a categorical variable for the firm's specified employee numbers), firm ownership type, and industry type. Standard errors are clustered at the job ad level.[18] Since the results are remarkably similar in all columns, our discussion will focus on Column [3].

[Table 2 about here]

Cross-Sectional Difference-in-Differences Estimate

Our initial empirical strategy without anticipating the policy change was to apply a cross-sectional DID approach by sibling type and gender under the old policy. Table 2 shows that

[17] We refrain from comparing callbacks across the gender line and drawing inferences on gender discrimination other than discussing gender difference in discrimination based on family responsibilities. See the end of Section 5 for discussions of potential gender segregation in the occupations in our experiment.
[18] Table A5 in Appendix F shows similar results based on Probit.

this estimate is -5.32% ((FOB – FSB) – (MOB – MSB), $p = 0.005$, Column [3]), since the callback gap 5.22 percentage points ($p < 0.001$) between the only-child (FOB, the omitted category) and siblinged women (FSB) is significantly *lower* than the gap of 0.1 percentage points between the only-child and siblinged men. This DID based on the regression is similar to the simple DID of 5.5% ($p = 0.003$) calculated directly from Table 1. It indicates that when permitted two children under the pre-2016 policy, women who were only children had a differentially *lower* callback rate, due to their potential family responsibilities, than siblinged women compared to their male counterparts. This estimate exploits only the cross-sectional variations and is subject to the caveats discussed in Section 4. It thus serves as suggestive evidence of discrimination against women based on family responsibilities.

Temporal Difference-in-Differences Estimates by Gender Group

The 2016 policy delinked birth limits and sibling status and shifted to the universal two-child policy. The associated temporal variations allow us to improve the identification strategy via temporal DIDs. Table 2 shows a temporal DID estimate of -7.49% for women ((FSA – FOA) – (FSB – FOB), $p < 0.001$, Column [3]). Again, this regression-generated temporal DID is strikingly similar to the simple temporal DID of -7.5% for women ($p < 0.001$) in Table 1. It indicates that after the 2016 change, siblinged women, whose legally permitted fertility is increased by one, receive differentially *fewer* callbacks than their only-child female counterparts whose legally permitted fertility is unchanged. This improved estimate can difference out any gender-contingent, time invariant unobservable differences between the sibling-status types (e.g., birth order and employers' perceptions of an infrequent signal) as discussed in Section 4. It serves as more rigorous *casual* evidence of discrimination based on family responsibilities against women. Note the same temporal DID is 0.46% for men ((MSA

– MOA) – (MSB – MOB), $p = 0.790$), indicating no family-related discrimination against men, consistent with the literature (e.g., Anderson et al., 2002; Benard and Correll, 2015).

Triple-Differences Estimate of Expected-Family-Responsibility Effect

Combining the two temporal DIDs for women and men above, we can construct a triple-differences estimate of the effect of the policy change. Table 2 shows that this estimate, [(FSA – FOA) – (FSB – FOB)] – [(MSA – MOA) – (MSB – MOB)], is -7.95% ($p = 0.002$), almost the same as the triple differences of -8.0% ($p = 0.002$) in Table 1. It indicates that the differential, negative policy impact on callback rates for siblinged applicants relative to their only-child counterparts is *worse* for women than for men. As discussed in Section 4, since both cross-sibling-status and cross-gender time trends are differenced out, the triple-differences estimate is immune to both fertility-unrelated unobservable differences across sibling status (provided they are time invariant) and any 2016-policy-irrelevant temporal differences (provided they influence the relative callback rates between only-child and siblinged applicants similarly for women and men). This approach yields an even more rigorous, causal estimate of family-based discrimination relative to the temporal DIDs.

Recall our within-gender temporal DIDs require assumptions that, within each gender group, the callbacks across sibling status would have evolved similarly had the policy change not taken place. The assumption necessary for the consistency of our triple differences is that the *relative* callback rates of female siblinged and only-child applicants would have trended similarly to the *relative* rates of male siblinged and only-child applicants, absent the policy change. Like Agan and Starr (2017), who use the same set of identification strategies in a correspondence study combined with a policy change, our experimental data are also too short to allow pre-policy trend comparisons. Thus, we take an alternative approach to investigate whether any other contemporaneous policies could violate these assumptions. We conducted a thorough examination of all policies and regulations promulgated and implemented by national, provincial, and municipal governments during our experiments in all the cities. As detailed in Appendix G, we find that, except for the 2016 national birth-policy change, no other contemporaneous policies have affected callbacks to trend differently for the two gender groups or for sibling types within each gender group, or have differentially affected callbacks for siblinged women (FS) relative to the other three groups (FO/MS/MO). Therefore, we believe that the trend assumptions for our temporal DIDs and triple differences are very likely satisfied. It is noteworthy that neither the temporal DIDs nor the triple differences rely on similar time trends for women and men. We discuss this important point further in Subsection 5.2.

In sum, our DIDs and triple differences yield consistent evidence on family-related labor market discrimination against women but not men. Further analyses in Appendixes H and I show that our results are robust to the inclusion of the Beijing data, potential order effects of resume submissions, and the possibility that employers may detect our study.[19,20]

5.2 Heterogeneity Analyses

Age

Our fictitious resumes reveal no information on marital or parenthood status. One question is whether employers, when making callback decisions, may use an applicant's *age* to infer

[19] Our main analysis excludes Beijing to minimize the potential confounds due to the announcement of the policy change during our data collection there. Our results are robust with Beijing being included (see Appendix H). We find little evidence for employers' response to the policy announcement *per se* in Beijing.

[20] Appendix I examines the possible order effects of resume submissions and assesses the potential risk of detection by employers. We find no systematic patterns of order effects. Controlling for potential order effects and their interactions with resume types does not affect our DIDs and triple-differences. Moreover, our analysis on potential detection risks follows Balfe et al. (2021) by restricting data in different ways. The new estimates are largely consistent with those from the original sample but exhibit some increase in magnitude (albeit with lower statistical precision as expected), suggesting an even higher degree of discrimination against women.

their (expected) family responsibilities, as suggested in previous studies (e.g., Petit, 2007). This question bears particular importance for the age range we study for several reasons. First, the median age of first birth for women in China is 25 nationwide and 27 in urban areas; it is 27 for women with an associate's degree, 28 for those with bachelor's degrees, and 30 for those with postgraduate degrees (China's 2010 National Population Census, Issue 6-1). Second, statistics show substantially higher birth rates among women aged 25–29 than those aged 22–24. For example, in 2014 and 2015, the average birth rate was 71.8‰ for the first birth and 53.3‰ for the second birth among women aged 25 to 29, far exceeding 36.3‰ and 7.4‰ among those aged 22 to 24 in China's urban areas (Chinese Bureau of Statistics, 2015–2016).[21] Third, the traditional preference for younger brides in China's marriage market leads to a deep-seated social norm on the "appropriate" marriage and childbearing age for women, mirrored in the statistics mentioned above. Labeled with the derogatory term, leftover women (*shèng nǚ*), unmarried women past their mid-20s are subject to enormous disparagement and social pressure to marry, as featured in the documentary *Leftover Women* (2019) directed by Shlam and Medalia.

These facts together suggest the possibility that when making hiring decisions without knowing marital or parenthood status, employers may make age-based assumptions about the possible future family constraints on prospective employees. We investigate this possibility in Table 3 by splitting our sample by age and comparing the DIDs and triple differences across age categories.[22] We do not report the subsamples for those aged below 24 and aged 27 or above as they are too small to render accurate estimates.

---

[21] Note these statistics encompassed all women in urban areas regardless of education. Hence, birth rates for women aged 22 to 24 who held associate or college degrees may be dramatically lower than in these statistics.
[22] As in Table 1, we generate the DIDs and triple differences from OLS that include only the treatment dummies (i.e., with no other covariates), with the standard errors clustered at the job ad level.

[Table 3 about here]

Table 3 first shows that regardless of the age cutoffs, our estimates remain largely comparable across the somewhat older cohorts ($\geqslant$ 24, $\geqslant$ 25, or $\geqslant$ 26) and also comparable across the younger cohorts (< 25, < 26, or < 27). Furthermore, *before* the policy change, the cross-sectional DID is -5.57% (p = 0.024) for those aged 25 or above, comparable to -4.72% (p = 0.109) for those below age 25; it is -4.91% (p = 0.162) for those aged 26 or above, similar to -5.41% (p = 0.016) for those below age 26. Therefore, before the policy change, employers' preferences for only-child women (who were permitted one more child than siblinged women) are *lower* in a *similar* way for the somewhat older and younger cohorts. This observation indicates that age does not moderate the pre-policy cross-sectional effect.

In sharp contrast, the temporal DIDs for women and the triple differences exhibit noticeable heterogeneities across age cohorts. For example, women's temporal DID is -10.44% (p < 0.001) for age 25 or above, higher in size than the -6.62% (p = 0.011) for those below age 25. These estimates, combined with men's temporal DIDs in the corresponding age cohorts, yield the triple differences of -13.90% (p < 0.001) for age 25 or above, far exceeding -4.73% (p = 0.189) in size for those below age 25. In a similar vein, for the age cutoff of 26, the triple differences is -12.65% (p = 0.031) for the somewhat older cohort, in contrast to -7.38% (p = 0.010) for the younger cohort. Thus, age plays a crucial role in moderating the impact of the policy change. That is, labor market discrimination against women based on family responsibilities worsens as their probability of maternity increases in their late 20s, the upper age limit in our sample. The cohort aged 26 or above comprises merely 20% of our data, so the small sample size is likely to compromise the estimation precision. With this caveat in mind, we also observe that women's temporal DID is -5.77%

(p = 0.157) for age 26 or above, somewhat *smaller* in size than -8.28% (p < 0.001) for below age 26, exhibiting a reversed pattern in the ≥ 25 vs. < 25 comparison (-10.44% vs. -6.62%). The reversal in these two sets of temporal-DID comparisons is because those 25-year-old are included in the older cohort when we use 25, instead of 26, as the cutoff. This observation suggests the age of 25 may be pivotal in influencing employers' perceptions of employees' expected family responsibilities, in line with the widely-held social norm discussed above.

The findings in Table 3 show that age moderates the effect in temporal DIDs and triple differences that utilize the policy change, but age does not affect the pre-policy cross-sectional DID. These findings align with our conjecture on employers' possible beliefs on the pent-up demand for children (discussed in Section 2). Our interpretation is that because the new policy released this demand among siblinged women, employers may be worried about possible increases in their family-related work interruptions. These worries may be more severe and imminent concerning somewhat older siblinged women who have a higher chance of getting married, starting a family, or having a second child during their employment. Statistics show that such concerns are not ill-founded. Immediately following the two-child policy, the number of live births in 2015 was exceeded by 3.9 million (27.0%) and 3.0 million (20.9%) in 2016 and 2017, respectively (National Health Commission of P.R. China, 2020), reaching the highest levels since 2000. Second-born children accounted for more than 45% of births in 2016 and more than 50% in 2017. In these two years, 6.9% and 4.6% of the second-child births were by mothers aged 22 to 24, drastically outnumbered by the second-child births by mothers aged 25 to 29 of 37.2% and 29.8% for 2016 and 2017, respectively (Chinese Bureau of Statistics, 2017–2018). These nationwide statistics may overstate the childbearing rates of professional women with a college education in large metropolitan areas. However, they are very likely to reflect the scale of pent-up demand and its associated age distribution among women that, following the policy change, employers may have anticipated would have children. Overall, our investigation of age heterogeneities extends our main result and shows that women's higher probability of maternity, reflected by age, exacerbates labor market discrimination against them.[23]

Occupation

One remaining puzzle is that although the universal removal of one-child restriction in 2016 increased the average childbearing possibilities for women, the average callback rate barely changed for men (30.9% before vs. 30.1% after, $p = 0.873$, OLS with standard errors clustered at the job ad level) but increased moderately for women from 32.7% to 35.5% ($p = 0.125$). Job composition, which stayed identical by industry and occupation across cities in the two study waves, cannot explain it. Rather, it may suggest possible *gender segregation*—preferences for women in the jobs included in our study. Table A11 of Appendix J provides some support for this conjecture. It shows a much higher callback rate for women than for men for administrative assistance and customer service positions and a comparable callback rate for jobs in sales pre- and post-policy-change.[24] The gender patterns in callback rates may reflect gendered perceptions about the characteristics of these jobs (decent pay or high

[23] Our study relies on China's policy change that released pent-up demand for a second child. Hence, our estimated discriminatory effect may be larger than an estimate without such a policy change. On the other hand, this pent-up demand (and the perceived need for its release) is positively related to women's reproductive age and is more dramatic for those older than 30. This is particularly true for educated women working in the largest metropolitan areas. Our study focuses on age 22–29 because of the entry-level jobs listed on 51job.com. Women in this age range may be less likely to have a second child than those aged 30 and above and may thus become more desirable substitutes in the labor market. Hence, our estimate of discrimination is likely to be smaller than if our sample had encompassed the age range older than 30.

[24] Annual surveys conducted in China by a competitor of 51job.com, Zhilian Zhaopin (2018, 2020), another nationally leading job board, show that the proportion of women was 45% in sales, 75% in administrative assistance, and 77% in customer service in 2018 (45%, 74%, and 68% in 2020). These statistics are substantive to the callback patterns by gender and occupation in our study.

stability rather than high status or salary); employers may perceive women as valuing job stability more than remuneration relative to men. Hence, women's relative dominance in these positions may account for the lack of substitution between men and women following the policy change.[25]

Gender segregation may prevent callbacks for the jobs in our study from evolving similarly across gender had the policy change not taken place. However, it should not nullify the assumptions necessary for the consistency of our temporal DIDs and triple-differences estimates. We apply the temporal-DID approach *within* each gender group, and thus these estimates should be immune from potential violations of similar gender trends. Furthermore, these potential violations should not invalidate the assumption for our triple-differences approach which requires the *relative* callbacks of female siblinged and only-child applicants to trend similarly to the *relative* callbacks of male siblinged and only-child applicants, absent the policy change. It does not require as a prerequisite that women and men share a similar time trend. In other words, the assumption for our triple differences should still hold, despite possible violations of similar gender trends, so long as *no* contemporaneous shocks during our study differentially affected callbacks of the siblinged women, relative to the other three groups. We validate this in Appendix G and discuss it in Subsection 5.1.

[Table 4 about here]

Table 4 shows that the temporal DIDs and triple differences by occupation are similar to those in Tables 1 and 2. Specifically, the temporal DIDs are all small and statistically

[25] Other studies also find higher callback rates for women in certain occupations which prefer women, e.g., retail sales and restaurant service (Neumark et al., 2019; Button and Walker, 2020). Some research shows that gender discrimination and its direction depend on the proportion of women and the degree of applicants' gender-(mis)match with these occupations. The callback rate is higher for women in gender-integrated and women-dominant occupations, e.g., accountant, sales representative, and administrative assistant, but lower in men-dominant occupations, e.g., software engineering (Albert et al., 2011; Maurer-Fazio and Lei, 2015).

insignificant for men ($p > 0.10$ in these occupations) but are economically sizable for women and statistically significant for sales (-8.22%, $p < 0.001$) and customer service (-7.87%, p = 0.041), albeit insignificant for administrative assistance (-3.01%, p = 0.539) presumably due to the small subsample. The associated triple differences are -8.64% (p = 0.009), -9.36% (p = 0.078), and -1.50% (p = 0.774) for sales, customer service, and administrative assistance, respectively.[26] These observations further corroborate our main result that women, not men, are negatively affected by perceived family responsibilities.

6 Conclusion

We conducted a two-wave correspondence study before and after China shifted its one-child policy to a universal two-child policy in January 2016. We find strong evidence of labor market discrimination based on expected family responsibilities that applies only to women, particularly those in their late 20s, approaching the peak childbearing years.

In May 2021, after the national census revealed low birth rates, the Chinese government further relaxed the birth limits to allow three children per family. Many are skeptical about the efficacy of this change, given the skyrocketing costs of raising a family and concerns over careers, especially for women. Our study bolsters this skepticism by showing that further relaxing birth limits is likely to exacerbate labor market discrimination against women of childbearing age and, as a result, suppress their desire to have more children.

A direct implication of our study is that the new policy or future ones that may ultimately remove birth limits may not successfully reverse the declining birth rate unless policymakers

[26] Substitution of the only-child women for the siblinged women in sales and customer service after the policy change can be attributed to these occupations being highly customers oriented and responsible for front-line communication. Their presence and performance may directly influence the customers' perceptions of their company's dedications to meet the customers' needs. These occupational attributes may make employers particularly sensitive to their employees' expected family responsibilities.

effectively address discrimination in the labor market against women's (expected) family responsibilities. Policymakers may consider reevaluating their complementary policies (e.g., increasing parental leave) intended to help working women that, unfortunately, end up hurting them (e.g., Lalive and Zweimüller, 2009). They may consider offering financial incentives to employers to reduce the costs of hiring or retaining working women. For example, in 2004, the French government granted private firms tax incentives to cover employees' childcare or improve their work-family balance (Pailhé and Solaz, 2019). The government may also consider providing or ramping up incentives to working women and families to offset their childcare expenses and the impact on their careers of childbearing and childrearing. Milligan (2005) finds that Quebec's tax incentives to families of up to C$8,000 for a child birth between 1988 and 1997 boosted fertility. Kalwij (2010) finds that increases in national expenditure on family policy programs (e.g., family allowances, maternal and parental leave benefits, and childcare subsidies) led to positive fertility responses in 16 Western European countries.

Our evidence of discrimination against women based on perceived family responsibilities pertains to the interview callback stage. The fact that we did not cover the subsequent hiring processes does not veil the indisputable insight of our study: any implicit cues of intended maternity, once available, may play an essential role in employers' hiring decisions as early as pre-interview screening. One direction for fruitful future research is to combine correspondence studies with empirical data that match employers with their employees (e.g., Hellerstein et al., 1999). This approach will allow researchers to investigate potential labor market inequalities during the succeeding stages of recruitment and employment.

References

Ameri, Mason, Lisa Schur, Meera Adya, F. Scott Bentley, Patrick McKay, and Douglas Kruse. “The disability employment puzzle: A field experiment on employer hiring behavior.” *ILR Review* 71, no. 2 (2018): 329-364.

Agan, Amanda, and Sonja Starr. “Ban the box, criminal records, and racial discrimination: A field experiment.” *The Quarterly Journal of Economics* 133, no. 1 (2018): 191-235.

Albert, Rocío, Lorenzo Escot, and José Andrés Fernández-Cornejo. “A field experiment to study sex and age discrimination in the Madrid labour market.” *The International Journal of Human Resource Management* 22, no. 02 (2011): 351-375.

Anderson, Deborah J., Melissa Binder, and Kate Krause. “The motherhood wage penalty: Which mothers pay it and why?.” *American Economic Review* 92, no. 2 (2002): 354-358.

Baert, Stijn. “Career lesbians. Getting hired for not having kids?.” *Industrial Relations Journal* 45, no. 6 (2014): 543-561.

Baert, Stijn, Bart Cockx, Niels Gheyle, and Cora Vandamme. “Is there less discrimination in occupations where recruitment is difficult?.” *ILR Review* 68, no. 3 (2015): 467-500.

Balfe, Catherine, Patrick Button, Mary Penn, and David Schwegman. *Infrequent Identity Signals and Detection Risks in Audit Correspondence Studies*. No. w28718. National Bureau of Economic Research, 2021.

Becker, Sascha O., Ana Fernandes, and Doris Weichselbaumer. “Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment.” *Labour Economics* 59 (2019): 139-152.

Bertrand, Marianne, and Sendhil Mullainathan. “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination.” *American Economic Review* 94, no. 4 (2004): 991-1013.

Bertrand, Marianne, and Esther Duflo. “Field experiments on discrimination.” *Handbook of economic field experiments* 1 (2017): 309-393.

Benard, Stephen, and Shelley J. Correll. “The motherhood wage penalty and status discrimination.” *Working in America: Continuity, conflict, and change in a new economic era* (2015): 273-294.

Button, Patrick, and Brigham Walker. “Employment discrimination against Indigenous peoples in the United States: Evidence from a field experiment.” *Labour Economics* 65 (2020): 101851.

Bygren, Magnus, Anni Erlandsson, and Michael Gähler. “Do employers prefer fathers? Evidence from a field experiment testing the gender by parenthood interaction effect on callbacks to job applications.” *European Sociological Review* 33, no. 3 (2017): 337-348.

Cahuc, Pierre, Stephane L. Carcillo, Andreea Minea, and Marie-Anne Valfort. “When correspondence studies fail to detect hiring discrimination.” (2019).

Cameron, Lisa, Nisvan Erkal, Lata Gangadharan, and Xin Meng. “Little emperors: behavioral impacts of China's One-Child Policy.” *Science* 339, no. 6122 (2013): 953-957.

Chen, Wei. “The Two-Child Policy and Fertility Rate in China.” *Journal of Peking University (Philosophy and Social Sciences)* (in Chinese) 56, no. 5 (2019): 57-68.

China’s State Council, 2012. Articles 7 and 9, *The Special Rules on the Labor Protection for Female Employees,* Decree No. 619. (URL: http://www.gov.cn/gongbao/content/2012/content_2136749.htm, accessed on December 17, 2020)

Chinese Bureau of Statistics, 2015-2020. *China Population and Employment Statistics Yearbook*. China statistics publishing house, Beijing.
Correll, Shelley J., Stephen Benard, and In Paik. "Getting a job: Is there a motherhood penalty?" *American Journal of Sociology* 112, no. 5 (2007): 1297-1339.
National Health Commission of P. R. China, 2020. *China Health Statistics Yearbook*. Peking Union Medical College Press, Beijing.
Dahl, Gordon B., and Matthew M. Knepper. *Age discrimination across the business cycle*. No. w27581. National Bureau of Economic Research, 2020.
Falbo, Toni. "Only children, stereotypes, and research." In *The child and its family*, pp. 127-142. Springer, Boston, MA, 1979.
Falbo, Toni, and Denise F. Polit. "Quantitative review of the only child literature: research evidence and theory development." *Psychological Bulletin* 100, no. 2 (1986): 176.
Feng, Xiaotian. "The construction and demonstration of the only child's negative image in media." *Sociological Studies*, no. 3 (2010): 177-198.
Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. "Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations." *Journal of labor economics* 17, no. 3 (1999): 409-446.
Helleseter, Miguel Delgado, Peter Kuhn, and Kailing Shen. "The age twist in employers' gender requests evidence from four job boards." *Journal of Human Resources* 55, no. 2 (2020): 428-469.
Kalwij, Adriaan. "The impact of family policy expenditure on fertility in western Europe." *Demography* 47, no. 2 (2010): 503-519.
Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11, no. 4 (2019): 181-209.
Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. "Duration dependence and labor market conditions: Evidence from a field experiment." *The Quarterly Journal of Economics* 128, no. 3 (2013): 1123-1167.
Lalive, Rafael, and Josef Zweimüller. "How does parental leave affect fertility and return to work? Evidence from two natural experiments." *The Quarterly Journal of Economics* 124, no. 3 (2009): 1363-1402.
Maurer-Fazio, Margaret. "Ethnic discrimination in China's internet job board labor market." *IZA Journal of Migration* 1, no. 1 (2012): 1-24.
Maurer-Fazio, Margaret, and Lei Lei. ""As rare as a panda": How facial attractiveness, gender, and occupation affect interview callbacks at Chinese firms." *International Journal of Manpower* (2015).
Maurer-Fazio, Margaret, and Sili Wang. "Does marital status affect how firms interpret job applicants' un/employment histories?." *International Journal of Manpower* (2018).
Milligan, Kevin. "Subsidizing the stork: New evidence on tax incentives and fertility." *Review of Economics and Statistics* 87, no. 3 (2005): 539-555.
Namingit, Sheryll, William Blankenau, and Benjamin Schwab. "Sick and tell: A field experiment analyzing the effects of an illness-related employment gap on the callback rate." *Journal of Economic Behavior & Organization* 185 (2021): 865-882.
Neumark, David. "Experimental research on labor market discrimination." *Journal of Economic Literature* 56, no. 3 (2018): 799-866.

Neumark, David, Ian Burn, and Patrick Button. “Is it harder for older workers to find jobs? New and improved evidence from a field experiment.” *Journal of Political Economy* 127, no. 2 (2019): 922-970.
Pailhé, Ariane, and Anne Solaz. “Is there a wage cost for employees in family-friendly workplaces? The effect of different employer policies.” *Gender, Work & Organization* 26, no. 5 (2019): 688-721.
Petit, Pascale. “The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector.” *Labour Economics* 14, no. 3 (2007): 371-391.
Qian, Nancy. “Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance.” *The Quarterly Journal of Economics* 123, no. 3 (2008): 1251-1285.
Quillian, Lincoln, John J. Lee, and Mariana Oliver. “Evidence from field experiments in hiring shows substantial additional racial discrimination after the callback.” *Social Forces* 99, no. 2 (2020): 732-759.
Reuters, 2015. “China says one-child policy stays in effect for now.” 2015-11-1. URL: https://www.reuters.com/article/us-china-population/china-says-one-child-policy-stays-in-effect-for-now-idUSKCN0SR0E020151102, accessed on May 23, 2018.
Riach, Peter A., and Judith Rich. “Field experiments of discrimination in the market place.” *The Economic Journal* 112, no. 483 (2002): F480-F518.
Rich, Judith. “What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000.” (2014).
Tilcsik, András. “Pride and prejudice: Employment discrimination against openly gay men in the United States.” *American Journal of Sociology* 117, no. 2 (2011): 586-626.
Zhilian, Zhaopin. 2018&2020. The 2018&2020 Investigation Report on the Present Situation of Chinese Women’s Workplace. http://news.fengone.com/d/20170307/592939.html.
Zhuang, Yaer, Yu Jiang, Zhili Wang, Chengfu Li, Jianan Qi, Hui Wang, Hongyan Liu, Bohua Li, Min Qin. “Fertility intention of rural and urban residents in China: Results from the 2013 national fertility intention survey.” *Population research* 38, no. 3 (2014): 3-13.

Table 1. Callback Rate (%)

	Men			Women		
	Treated Siblinged	Control Only child	Difference	Treated Siblinged	Control Only child	Difference
Before policy change	MSB	MOB	MSB – MOB	FSB	FOB	FSB – FOB
	30.8	31.0	-0.2	35.4	30.1	5.3
	(1,106)	(1,106)	[0.884]	(1,106)	(1,106)	[<0.001]
After policy change	MSA	MOA	MSA – MOA	FSA	FOA	FSA – FOA
	30.2	29.9	0.3	34.4	36.6	-2.2
	(1,106)	(1,106)	[0.823]	(1,106)	(1,106)	[0.063]
After – Before	MSA – MSB	MOA – MOB	(MSA – MOA) – (MSB – MOB)	FSA – FSB	FOA – FOB	(FSA – FOA) – (FSB – FOB)
	-0.6	-1.1	0.5	-1.0	6.5	-7.5
	[0.747]	[0.579]	[0.794]	[0.624]	[0.001]	[<0.001]
	Triple differences = -8.0 [0.002]					

Notes: This table reports the callback rates (the number of resumes in parentheses) for the four categories of fictitious applicants before and after the policy change. For the cross-resume-categories comparisons conditional on policy, *p* values in brackets are based on McNemar's Chi-square test for the paired binomial data. For the before-and-after-policy differences in callbacks, *p* values in brackets are based on the unpaired test of proportions. The triple differences are defined as $\{[(FSA - FOA) - (FSB - FOB)] - [(MSA - MOA) - (MSB - MOB)]\}$. The temporal DIDs and triple-differences estimates are based on OLS with no covariates being included and standard errors clustered at the job ad level.

Table 2. Determinants of the Callbacks (OLS)

	[1]	[2]	[3]
Regression estimates			
Male only-child before	0.009	0.004	0.004
(MOB)	(0.014)	(0.015)	(0.015)
Female with siblings before	0.052	0.052	0.052
(FSB)	(0.013)	(0.013)	(0.013)
Male with siblings before	0.007	0.003	0.003
(MSB)	(0.014)	(0.014)	(0.014)
Female only-child after	0.065	0.061	0.070
(FOA)	(0.020)	(0.024)	(0.023)
Male only-child after	-0.002	-0.011	-0.001
(MOA)	(0.019)	(0.024)	(0.023)
Female with siblings after	0.043	0.038	0.048
(FSA)	(0.020)	(0.023)	(0.023)
Male with siblings after	0.001	-0.007	0.002
(MSA)	(0.019)	(0.023)	(0.023)
Cross-sectional DID (%)			
(FOB − FSB) − (MOB − MSB)	-5.42	-5.25	-5.32
	[0.003]	[0.006]	[0.005]
Temporal DIDs (%)			
(FSA − FOA) − (FSB − FOB)	-7.50	-7.50	-7.49
	[<0.001]	[<0.001]	[<0.001]
(MSA − MOA) − (MSB − MOB)	0.45	0.45	0.46
	[0.794]	[0.796]	[0.790]
Triple differences (%)			
[(FSA − FOA) − (FSB − FOB)] −	-7.96	-7.95	-7.95
[(MSA − MOA) − (MSB − MOB)]	[0.002]	[0.002]	[0.002]
Applicant characteristics	No	Yes	Yes
Employer and job characteristics	No	No	Yes
City fixed effects	Yes	Yes	Yes
Observations	8,848	8,848	8,848

Notes: The dependent variable is whether a fictitious applicant receives a callback. The reference group (omitted) in the regressions is FOB, i.e., female only-child applicants before the policy change. The callback for FOB is 30.1%. The applicants' characteristics include age, education, years of work experience, the number of previous jobs, college major, and university fixed effects. Job and firm characteristics include occupations, number of applications submitted to the ad, the number of people who followed the firm on the job board, firm size, ownership, and industry type. Standard errors in parentheses are clustered at the job ad level. The lower panel provides the Wald tests with the *p* values reported in brackets.

Table 3. DIDs and Triple Differences by Age Category (%)

	≥ 24	**< 25**	**≥ 25**	**< 26**	**≥ 26**	**< 27**
Cross-sectional DID						
(FOB − FSB) − (MOB − MSB)	-5.94	-4.72	-5.57	-5.41	-4.91	-3.91
	[0.001]	[0.109]	[0.024]	[0.016]	[0.162]	[0.051]
Temporal DIDs						
(FSA − FOA) − (FSB − FOB)	-9.08	-6.62	-10.44	-8.28	-5.77	-7.36
	[<0.001]	[0.011]	[<0.001]	[<0.001]	[0.157]	[<0.001]
(MSA − MOA) − (MSB − MOB)	2.68	-1.89	3.46	-0.90	6.88	-1.61
	[0.171]	[0.441]	[0.208]	[0.644]	[0.093]	[0.377]
Triple differences						
[(FSA − FOA) − (FSB − FOB)] −	-11.76	-4.73	-13.90	-7.38	-12.65	-5.75
[(MSA − MOA) − (MSB − MOB)]	[<0.001]	[0.189]	[<0.001]	[0.010]	[0.031]	[0.030]
Observations	7,196	4,632	4,216	7,052	1,796	8,044

Notes: Similarly to Table 1, the DIDs and triple differences are based on OLS that includes only the treatment dummy variables, with the standard errors clustered at the job ad level. *P* values are in brackets. Results are similar if we include all covariates in the OLS. The two subsamples below age 24 and age 27 or above are too small to render accurate estimates and are not reported here.

Table 4. DIDs and Triple Differences by Occupation (%)

	Sales	Administrative Assistance	Customer Service
Cross-sectional DID			
(FOB − FSB) − (MOB − MSB)	-6.52	-3.01	-3.75
	[0.007]	[0.416]	[0.308]
Temporal DIDs			
(FSA − FOA) − (FSB − FOB)	-8.22	-3.01	-7.87
	[<0.001]	[0.539]	[0.041]
(MSA − MOA) − (MSB − MOB)	0.42	-1.50	1.50
	[0.857]	[0.595]	[0.651]
Triple differences			
[(FSA − FOA) − (FSB − FOB)] −	-8.64	-1.50	-9.36
[(MSA − MOA) − (MSB − MOB)]	[0.009]	[0.774]	[0.078]
Observations	5,648	1,064	2,136

Notes: Similarly to Table 1, the DIDs and triple differences are based on OLS that includes only the treatment dummy variables, with the standard errors clustered at the job ad level. *P* values are in brackets. Results are similar if we include all covariates in the OLS.

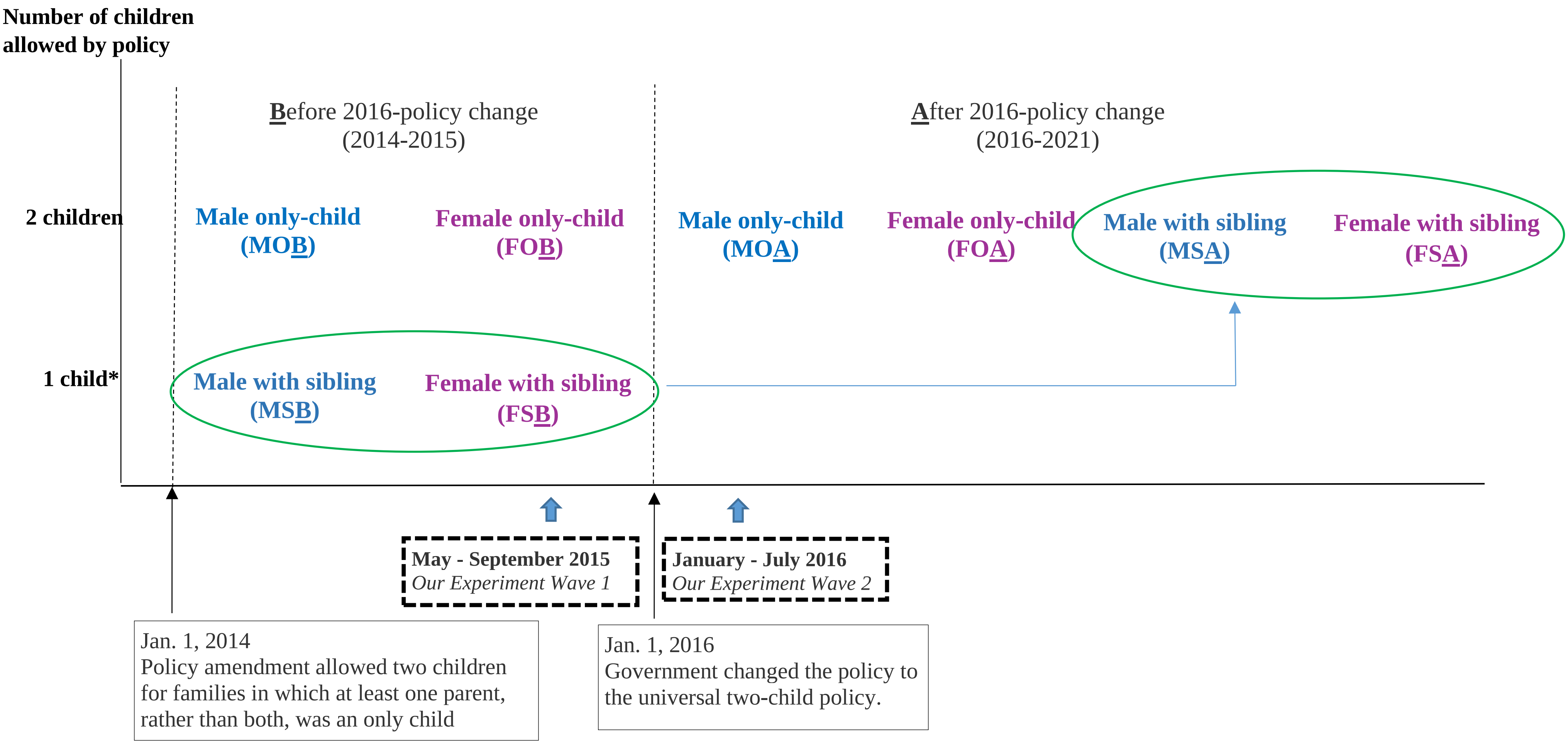

Figure 1. Policy Impact on Each Category

Notes: This diagram illustrates where each group stands regarding the policy-allowed fertility under the pre- and post-2016 policy environments. "B" and "A" in the acronyms stand for before or after the 2016 policy change, respectively. The oval highlights the groups treated by the 2016 policy change.
*During 2014 and 2015, those with siblings could have only one child unless their spouses were an only child.

Online Appendixes

Appendix A. Brief History of China's One-Child Policy and the Timeline of Two Waves of Our Experiments

As shown in Figure A1, China introduced the one-child policy in 1979 to curb its population growth. The policy limited the Han ethnic majority (over 90% of the population) to only one child per family. The enforcement of this policy mostly relied on propaganda, incentives, and punishments for violations (e.g., steep fines, demotion or loss of public-sector employment). Non-Han ethnic minorities were exempted from the one-child policy. See Wang et al. (2016) for more details on this history. Despite its efficacy in slowing down population growth, the restrictive birth policy was widely criticized for the brutality of enforcement and skewing the sex ratio of men to women. In 1984, only five years later, the government amended the policy to allow a second child for two types of families: those in rural areas where the firstborn was a girl and, regardless of geographic regions, where both parents were only children themselves. The amendment regarding rural areas was partly intended to accommodate a traditional preference for sons and partly due to enforcement difficulties in these areas. The exemption for parents who were both only children was to battle the so-called "4-2-1 problem"; an adult who was an only child would need to care for two aging parents and two pairs of even older grandparents at some point in life. The 1984 amendment marked the start of birth limits as tied to an adult's sibling status.

Thirty years later, on January 1, 2014, the government, concerned for the stagnating birth rate, further expanded the sibling-status-based exemption to families with at least one spouse, rather than both, being an only child. Starting in 2014, couples with either parent being an only child could have two children; the one-child restriction remained unchanged for couples who were both siblinged. The differential treatment in the birth policy increased the anticipated family responsibilities for men and women who were only children.

On October 29, 2015, the Communist Party leadership announced the shift to the two-child policy: all married couples could have two children starting on January 1, 2016.

As noted in Figure A1, the first wave of our study was conducted in 2015 before the policy change and the second wave in 2016 after the policy change. Table A1 in Appendix C reports the detailed timeline.

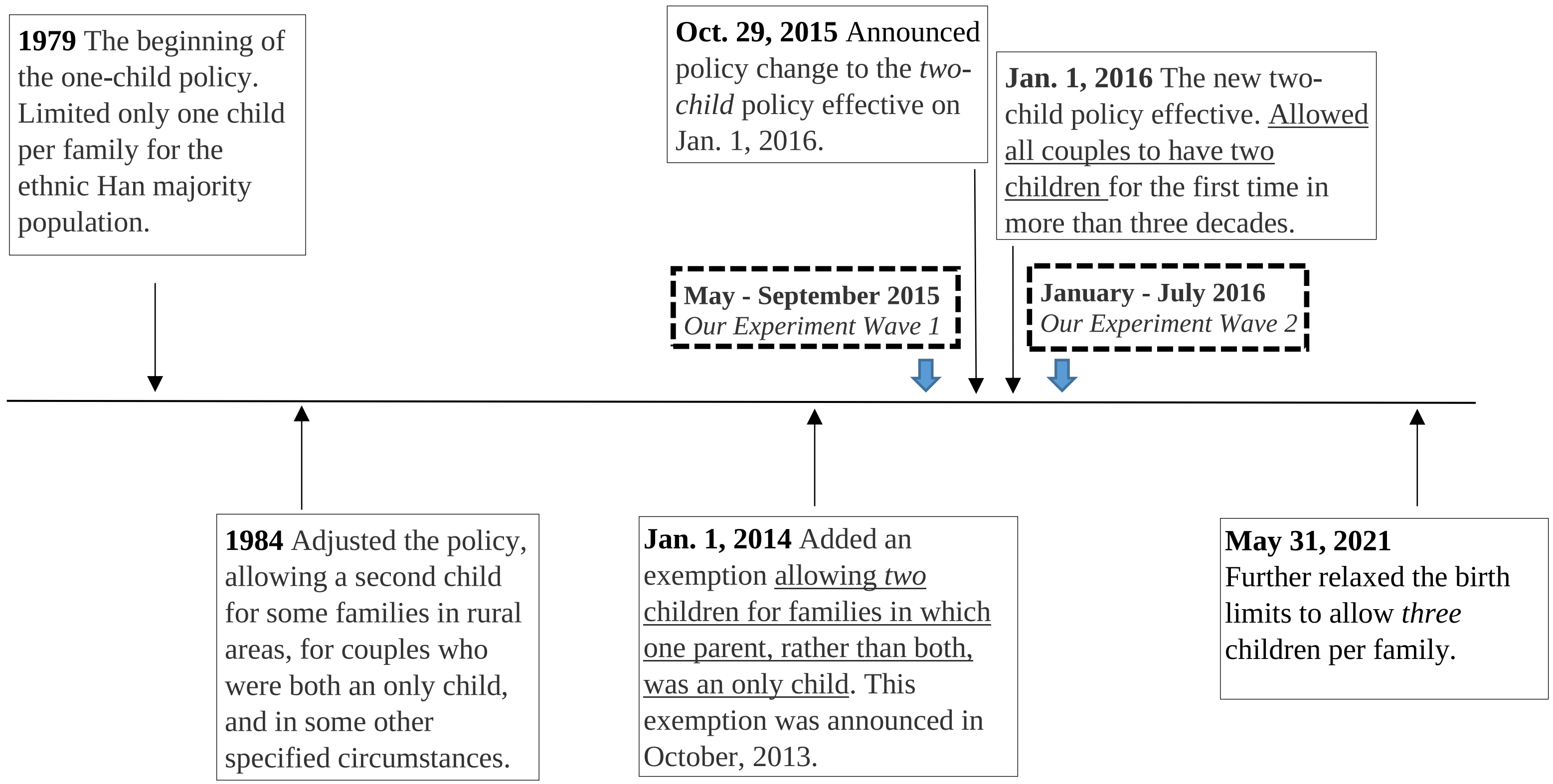

Figure A1. Timeline of China's One-Child Policy and Two Waves of Our Experiments

Source: The Guardian
(URL: https://www.theguardian.com/world/2013/nov/15/china-one-child-family-policy-timeline)

Appendix B. Evidence of Unexpected Termination of the One-Child Policy

The timing of the termination of the one-child policy was completely unanticipated, especially because the policy had undergone a major amendment only two years earlier in 2014. The unexpectedness of this change is evident in the web-search trends set out in Figure A2 below. This shows the search volume index on baidu.com, China's Google, for the phrase "second child for all" (*quán miàn èr tāi*), a term commonly used by policymakers and public media to describe the 2016 policy relaxation of the one-child policy. After being near zero over many years, this search index surged and reached its peak immediately after the government's announcement of the reform on October 29, 2015, the last day of the summit of the Communist Party's policy-making Central Committee.

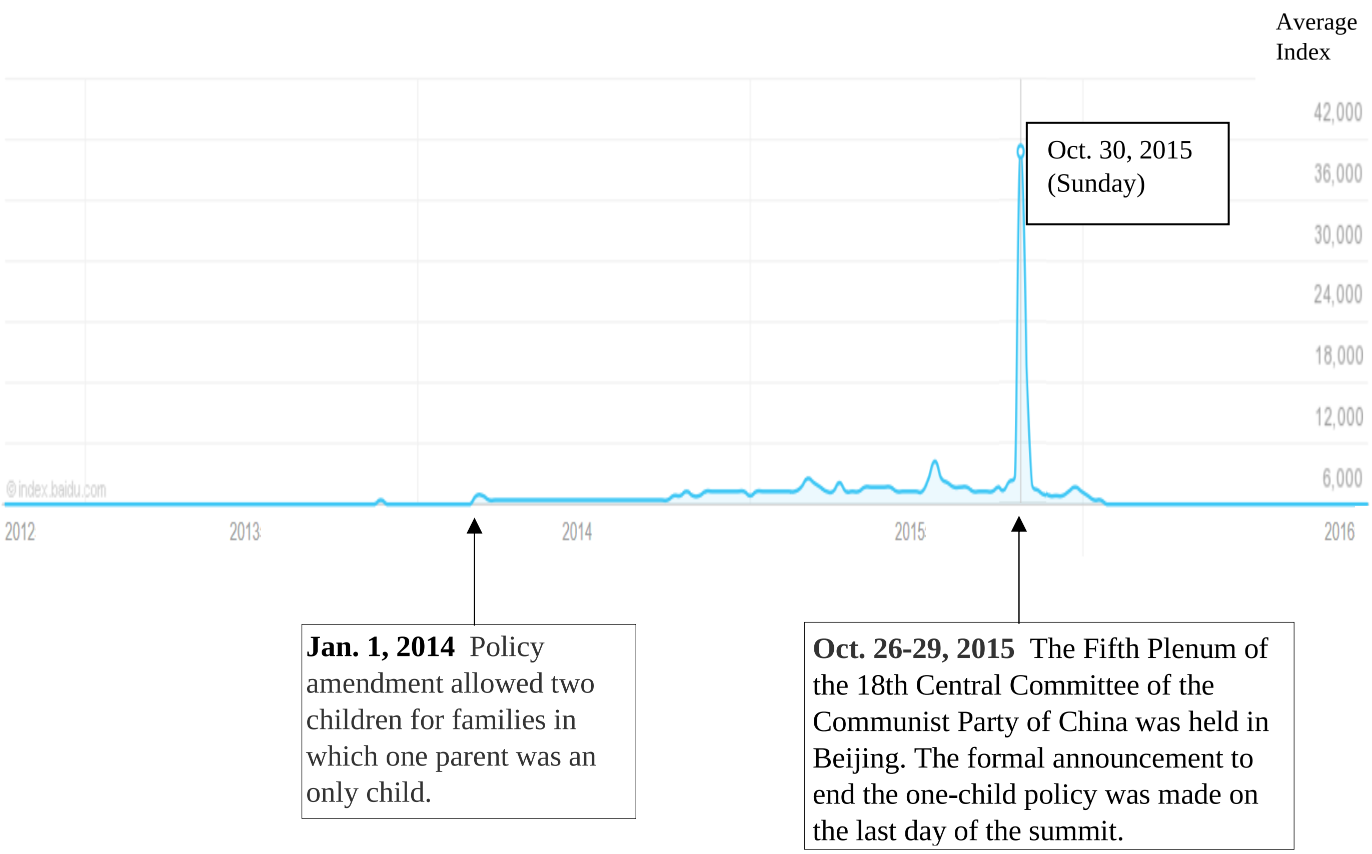

Figure A2. Baidu Search Volume Index for "Second Child for All" (*quán miàn èr tāi*)

Appendix C. Details of Experimental Design and Procedures

Job Board

We conducted our study on 51job.com (Nasdaq JOBS), China's leading human resource services provider headquartered in Shanghai. The company, founded in 1998, completed its Nasdaq IPO in September 2004, becoming China's first publicly listed HR-services firm. The company leverages its technology (e.g., online recruitment platforms and mobile applications) and staff expertise to connect thousands of domestic and multinational corporate clients with millions of job seekers throughout the talent management process from recruitment to career development and employee retention. It offers a broad range of services (e.g., recruiting, training, evaluations, and HR tools) through its online platforms and 26 subsidiaries in Mainland China and Hong Kong.

According to their website, in 2015, 51job.com had over 100 million registered users and a database of 96 million resumes; over 3.2 million job ads were posted, and about 38 million applications were delivered to prospective employers weekly. In 2020, they had 100 million resumes in their database, peak traffic of over 300 million average daily page views, over 5 million job postings, and about 42 million applications delivered to potential employers every week. The site is considered "China's Most Influential Recruitment Website" by domestic and international industry watchers such as The China Internet Network Information Center—the administrative agency responsible for Internet affairs under China's Ministry of Industry and Information Technology—and Euromonitor International—the world's leading independent provider of strategic market research (51job website).

Cities

Shanghai, Guangzhou, Shenzhen, and Beijing are the four most economically advanced megacities in China, ranked as the top four cities by GDP in 2014. Shanghai, a global business and financial center, is China's most populous city, with over 24 million people. Guangzhou (Canton to westerners) is the capital of Guangdong Province in the southeast and the third-largest city behind Beijing and Shanghai. With a population of more than 13 million, it is a manufacturing hub of clothing, electronics, plastic goods, and toys. Shenzhen, another major city in Guangdong Province immediately adjacent to Hong Kong, is a manufacturing center for high technology products and has a population of more than 18 million. Beijing is the country's capital, political and cultural center and the largest city by area with approximately 19 million people. Science, technology, research, and finance are important components of its local economy. These cities provide job opportunities for the locals and attract hundreds of thousands of migrant workers nationally.

In the main two-wave study, we sent a total of 10,528 resumes in response to 2,632 job ads from May 2015 to July 2016. Table A1 reports the number of resumes by city and the corresponding time frame. In the companion experiment to test the potential impact on callbacks of signaling sibling status, 948 resumes were sent to 237 job ads in Shanghai (See Appendix E).

Table A1. Number of Resumes Sent and Time Frame by City and Wave

	Shanghai	Guangzhou	Shenzhen	Beijing
Wave 1	1,878	1,302	1,244	1,680
	(5/10/2015–6/23/2015)	(7/4/2015–8/7/2015)	(8/15/2015–9/22/2015)	(10/10/2015–12/24/2015)*
Wave 2	1,878	1,302	1,244	No data collected
	(1/17/2016–4/14/2016)	(5/10/2016–6/2/2016)	(5/31/2016–7/3/2016)	

Notes: *The government announced the unexpected policy change on October 29, 2015.
The time taken to collect data in each city was primarily determined by the number of research assistants (RAs) available during the time window. For example, it took longer to complete the data collection in Beijing as we ran into a shortage of RAs toward the end of 2015 due to final exams. We hired new RAs in early 2016 after the winter break. The time for their training contributed to the duration of Wave 2 in Shanghai.

As noted in Table A1, the government announced the unexpected policy change on October 29, 2015, during our first-wave data collection in Beijing. We, nonetheless, completed the planned data collection on December 24, 2015. The mere announcement of the termination of the one-child policy would potentially affect employers' perceptions of applicants' expected family responsibilities. Therefore, the first-wave data from Beijing was likely to be "contaminated", and we thus did not collect the second-wave data in Beijing. We, therefore, focus on the data collected in the three other cities not subject to this “contamination.” We show, however, that our results are robust by including the Beijing data partially or entirely in Appendix H.

Industries, Occupations, and Job Ads

The 51job.com site and other leading job boards in China mainly cover entry-level jobs. To ensure sufficient observations and reduce logistical difficulties in each city, we selected two industries with the most job listings during the two months before the first wave of the experiment. These industries were internet and finance in Shanghai and Beijing, internet and fast-moving consumer goods in Guangzhou, and internet and electronics in Shenzhen. The two criteria applied in our choice of occupations were, a) a sufficiently large number of job listings, and b) they hired enough women. We selected three job categories—sales, administrative assistance, and customer service. These occupation categories are also used in previous audit studies such as Bertrand and Mullainathan (2004) and Kroft et al. (2013).

To ensure a sizable number of job ads and limit the large degree of heterogeneity in their requirements, we applied several criteria to our selection. We excluded jobs that were to be filled immediately; our pilot study indicated that to fill such positions quickly, employers often offered interviews to all applicants. We excluded postings asking for photographs; very few employers required photos, and most of the site’s resumes did not include these. We further excluded job postings requiring lower than an associate degree or higher than a bachelor’s degree or targeted fresh college graduates or applicants with five or more years of work experience. The reason was we found that 68% of the publicly listed 51job.com postings required a college education, and 82% required some work experience, of which 76% required fewer than five years. Finally, we excluded postings stipulating the desired gender.

In the second wave, we used the same two industries and three occupations in the corresponding cities. The selected job postings still satisfied our selection criteria discussed above. In this wave, we applied for the same number of jobs (1,106) as in the first wave. We did not intend to preclude employers present in our first wave. However, there was no overlap across waves in practice because of the vast labor demanded by each city’s many employers in these occupations and industries. Thus, we were able to maintain the job compositions from the first to the second wave.[27]

简 历 (Resume)

个人信息 (Personal Information)

姓名 (Name)	性别 (Gender)
出生日期 (Date of birth)	居住地 (Address)
工作年限 (Years of work experience)	电子邮件 (Email)
手机 (Cell-phone number)	

自我评价（**Self-Assessment**）

求职意向（**Career Objectives**）

到岗时间 (Starting time)
工作性质 (Employment type)
希望行业 (Intended industry)
目标地点 (Intended city)
目标职能 (Intended occupation)
求职状态 (Employment status)

工作经验 **(Work Experiences**）

教育经历 **(Educational Background**）

Figure A3. 51job.com Resume Template

Appendix Figure A3 shows the 51job.com required resume template. Our 2 × 2 factorial experimental design was implemented by a) explicitly entering the gender information in the

[27] See Neumark et al. (2019) on how an unbalanced sample due to changes in job compositions may affect analysis, and how this can be empirically resolved.

required field, and b) signaling potential fertility through self-disclosing sibling status in the first sentence in the Self-Assessment field.

Name and Gender The fictitious name consisted of two Chinese characters–a common last name (Wu and Yang, 2014) and a popular male- or female-sounding first name. We avoided gender-neutral first names to make the applicants' names gender salient and specified the applicants' gender in the required field.

Signaling Expected Fertility through Sibling Status The second part of the resume template is the "Self-Assessment" field, allowing applicants to freely introduce themselves as in the U.S. job application's short cover letter. We use this field to include our sibling-status signal. Specifically, we start the self-assessment with either "[a]s the only child in my family…" or "[a]s one of the children in my family…", followed by fictitious applicants' self-evaluations on quality traits relevant to being a good employee. An example follows:

> "As the only child [or one of the children] in my family, I consider myself mature, kindhearted, trustworthy, modest, and sensible with a pleasant personality. I am a well-organized, fast learner. I am meticulous and resourceful. I am personable, have good self-management skills, and treat others sincerely. I have good interpersonal skills and like communicating with others."

Signaling sibling status on one's resume was not unusual on this job board. To design the self-assessment above, we closely studied 3,000 resumes randomly drawn from the resume pool assessable by employers. These resumes were posted or updated by their owners up to six months prior to our experiment. Among these resumes, 211 (7%) explicitly revealed sibling status, including 61% who self-disclosed as an only child and 39% who disclosed their siblinged status. Among the 211 resumes that self-disclosed, 71% focused on their positive employee attributes regardless of their sibling status, 24% tried to leverage their sibling status to cater to specific jobs (e.g., as an only child, they wanted a local job that could offer good financial stability to care for elderly parents; as a siblinged child, they were less geographically limited and had no issues with business traveling), 3%—all only children—openly admitted their single-child stereotypical shortcomings (e.g., they needed to improve teamwork) while trying to highlight their positive employee traits, and 2%—half only children and half siblinged children—specifically disclosed the number of their children. Therefore, when we designed the self-assessment, we followed the majority case above and focused on positive employee attributes (see statements adopted in our fictitious resumes in Appendix Table A2). This ensured that while we varied the sibling status, the rest of the self-assessment appeared largely comparable across the four resume types (MO/FO/MS/FS).[28]

Table A2. Bank of Self-Assessment Statements
(Translated from Chinese)

Bin	Number	Statement
1	1	I have good coordination skills and enjoy working with others. I can acclimate to a new workplace with little difficulty.
	2	I take my work seriously. I am thoughtful, conscientious, and motivated.
	3	I have good interpersonal skills and like communicating with others.
	4	I complete assignments on schedule and get along well with my coworkers.
2	5	I am hardworking, sincere, confident, and self-driven. I highly value teamwork and collectivism.
	6	I am personable, have good self-management skills, and treat others sincerely.
	7	I am perseverant, persistent, reliable, and always hew to employers' rules.
	8	I am competent and dedicated to my work. I am voluble and willing to share my expertise with others.
3	9	I am warmhearted, easy-going, and enterprising with great group spirit.
	10	I can adapt to changes easily. I am bright, quick-witted, creative, and eager to take initiative.
	11	I always bear in mind the importance of integrating theory and practice and can work effectively when facing difficulties and challenges.
	12	I consider myself mature, kindhearted, trustworthy, modest, and sensible with a pleasant personality.
4	13	I am a well-organized, fast learner. I am meticulous and resourceful.
	14	I am self-effacing, prudent, and efficient. I have the courage to take responsibility.
	15	I am a self-starter and aspire to excellence. I work diligently and have a strong sense of duty.
	16	I undertake my tasks with rigor and care. I am adept at examining my work and learning from my experiences.

[28] The cases in which applicants tried to leverage their sibling status to cater to specific jobs (24% of the 211 resumes) were too difficult to implement in a consistent manner across the four resume types without generating additional unintended confounds. Other cases in which applicants specified the number of children or admitted upfront their only-child-related stereotypical shortcomings (5% of the 211 resumes) were too rare.

To construct the positive attributes, we first formed a bank with 16 statements, classified into four 4-statement bins, on skills and characteristics of a good employee (see Appendix Table A2). To ensure authenticity, we adopted key words and phrases of positive employee traits from the 3,000 real resumes obtained from 51job.com and closely studied them in preparation for our experiment. We randomly selected one statement from each of the four bins in the bank, with no replacement, and combined them to form four short paragraphs (i.e., the self-assessments), one for each of the four resume types. No statements appeared on more than one resume for any job ad. The sample self-assessment provided earlier was the combination of Statements 3, 6, 12, and 13 from Table A2.

Educational Background All our fictitious applicants had a three- or four-year college education since about 68% of jobs on 51job.com had such requirements at the time of our experiment. We used two local and two non-local mid-ranked, comprehensive public universities as the graduating schools and randomly assigned them across the four resume types. The two non-local universities were in one of the three other megacities with a GDP and population closely resembling that of the target city. The fictitious applicants held either a three-year associate's degree or a four-year bachelor's degree, depending on the stipulations in the job ads. Major fields of study included accounting, business administration, Chinese, economics, finance, history, international economics and trade, marketing, and psychology. These were employers' preferred college majors or those often listed on real applications for the industries and occupations that we targeted.

Work Experience For our targeted industries and occupations on 51job.com, at the time of our experiment, employers stipulated work experience as a requirement for 82% of positions, among which 76% stipulated up to 5 years of work experience. Therefore, we included 0 (exclusive) – 5 (inclusive) years of work experience on our resumes. Specifically, we used the number of years of experience stipulated plus up to an extra year. For example, for an ad stipulating one year of work experience, our four resumes listed experience of between one and two years. If there was no such stipulation, we first generated a random integer between zero and four and added up to one year for those resumes.

In the corresponding field on the resume, we followed the common practice of listing work experience in bullet points and ensured they looked different across the four applications for the same position. To make the fictitious resumes look authentic, we built a database using work experience and job skills adopted from real resumes. For each job ad in our experiment, we found four sets of work experience and skills in our database that would match the job in the corresponding industry and occupation in the targeted city. We then randomly assigned these four sets of experience and skills to each of the four resume types. To prevent our fictitious resumes from coinciding with these real applications to the same job ads, when building our database, we only included the information from real resumes last updated before January 1, 2015, and hence inactive at the time of our experiment.

Age To determine the age distribution in our study, we collected information on employers' required educational backgrounds and years of work experience in the targeted industries and occupations. We then mapped this information to the ages of applicants. Specifically, their age should be equal to the sum of the years of formal schooling and work experience, assuming they started the first grade at age seven.[29] We then generated a date of birth that would match the stipulated education and work experience.

Other Items on Resume We used the People's Republic of China Resident Identity Card for the required identification field. The identification numbers were generated in Matlab to be consistent with the fictitious applicants' age and gender.[30] We used local addresses on all the resumes. We used the lab's designated mobile phones and purchased local mobile phone numbers in the corresponding city for each part of our experiment. We made sure that these phone numbers, particularly the first three digits, looked substantially different. For each fictitious resume, we registered an email address free of charge at 163.com, one of China's largest email service providers. Our RAs were assigned to duty at the lab during regular work hours. They were able to answer the phones or check phone or email messages promptly.

In the Career Objectives category, we specified "within a week" for starting time, "full time" for employment type, "actively searching for a job" for employment status, and the

[29] The first grade starts at the age of six or seven in China, depending on geographic region. We computed the years of schooling by summing up six years in the elementary school, three years in the middle school, three years in the high school, and three or four years in college to match the fictitious applicant's educational background and degree in the Educational Background field on the resume.

[30] In China, the ID number on one's resident identity card consists of 18 digits. The 7th to the 14th digits indicate the birth year, month, and date in the YYYYMMDD format. The 17th digit indicates sex, with odd numbers for male and even numbers for female.

industry, occupation, and city listed on the job ad for the intended industry, occupation, and city, respectively.

In sum, we generated individual resumes by compiling each part based on the above procedures, rather than recycling similar resumes from previous applications.

Resume Submission

Based on our capacity (mainly determined by the number of RAs available), we randomly selected a certain number of job ads from 51job.com in our targeted industries and occupations that were newly published within 24 hours and met our selection criteria. We then assigned the ads and the corresponding resumes to our RAs who would register accounts on 51job.com, submit resumes, and record information about the resumes, the matched employers and positions, and the submission time. Our RAs submitted the four different types of resumes to targeted job ads, one each day within four consecutive days, in random order. This order was determined by the sizes of four decimals randomly generated by a computer. We gave the resumes and corresponding positions unique numbers to facilitate cross-checking; upon completion, our RAs saved the job ads and submission pages permanently as PDF files.

Employer Response

We followed the literature to define the success of a job application as a positive response (e.g., an interview invitation) from employers via phone, text message, or email. Our RAs checked the messages regularly on the designated cellular phones and email accounts and answered callbacks promptly. They matched employer responses to specific resumes with little difficulty and recorded the information based on the employer names, the fictitious applicant names, and phone numbers or emails. The RAs responded to interview invitations by informing the recruiters that the applicant had accepted another offer and was no longer available. Our pilot showed that we received 97% of callbacks within ten days and more than 98% within two weeks. Therefore, consistent with Kroft et al. (2013), we tracked callbacks from employers for up to two weeks from submission.

Research Assistants

We recruited our RAs among the undergraduate students at the Beijing Normal University. Based on what we had learned from the pilot study (conducted in March–April 2015), we designed step-by-step, detailed procedures for each part of the experimental process (e.g., how to screen job ads, register 51job.com accounts, submit resumes, collect employers' publicly available information, track callback, and record data) and provided thorough training. The RAs conducted the assigned tasks by closely following these pre-designed protocols on the designated computers and cellular phones at Beijing Normal University's Lab for Experimental Economics. The two local coauthors took turns supervising the RAs at the lab and answered questions promptly. They also audited the completed work daily to ensure that the RAs properly followed protocols and accurately recorded data.

Appendix D. Manipulation Check on the Salience of Sibling Status

We conducted a manipulation check on the salience of signaling job applicants' sibling status by following Kroft et al. (2013). In the manipulation check, we recruited 99 MBA students at the Beijing Normal University with a specialization or previous work experience in human resource management. As our human resource evaluators, they read a given job ad and reviewed two fictitious resumes that we used in the experiment. They then recommended one applicant for an interview.

As shown in Table A3, we found that 68% of our evaluators could recall both applicants' sibling status correctly without referring to the resumes. This percentage is similar to those for correctly recalling the number of jobs previously held by the applicants (74%, $p = 0.273$), significantly higher than recalling their respective work experience (9%, $p < 0.001$) and universities attended (44%, $p < 0.001$), and marginally lower than recalling their respective educational degrees (80%, $p = 0.058$). When asked to comment on the resumes, no evaluators mentioned anything strange or unusual about the resumes, e.g., revealing the only-child or siblinged status. When asked which applicant would better suit the job, 20 out of the 90 evaluators ($p < 0.001$, test of proportions) referred to the fictitious applicant's sibling status in their recommendations.

Table A3. Accuracy Rate of Recalling Resume Information in Manipulation Check

Variables	Fraction of MBA evaluators who got correct answers for both resumes	Comparison with the only-child variable in Row 1 (*p* value, McNemar test)
1. Is the applicant an only child?	0.68	1.000
	(0.47)	
2. Number of previous jobs	0.74	0.273
	(0.44)	
3. Work experience	0.09	<0.001
(years and months)	(0.29)	
4. Educational degree	0.80	0.058
(3-year associate's or 4-year bachelor's)	(0.40)	
5. Name of university attended	0.44	<0.001
	(0.50)	

Notes: We code each of the variables above as one if an evaluator answered it correctly for *both* resumes and as zero otherwise. Standard deviations are listed in the parentheses. The last column reports the *p* values of McNemar tests comparing each of Variables 2–5 with the first variable on recalling sibling status.

Appendix E. Did Signaling the Sibling Status Affect Callbacks? A Companion Experiment in Shanghai

One question in our study was whether signaling a job applicant's sibling status *per se* could affect the callback, irrespective of the fertility issue. We investigated this question by conducting a companion experiment immediately after the second wave of the main study in Shanghai in April and May 2016. In this companion experiment, we sent 948 resumes in response to 237 job ads in Shanghai. For 120 jobs, we sent four resumes for each ad, including a male or female with *sibling-child* status (just as in the main study) and a male or female *without* such information revealed. For the other 117 jobs, we sent four resumes for each ad, including male or female with the *only-child* status (just as in the main study) and male or female *without* revealing such information.[31] The selection criteria for job ads, fictitious resume creation, and the submission procedure were otherwise identical to that in the main study. Since, in the post-policy environment, everyone, regardless of sibling status, was allowed two children, this experiment allowed us to isolate any impact of the sibling-status signaling, irrespective of expected fertility.

Table A4 reports the results from the OLS regression analysis. The resumes without the sibling-status signal are in the omitted reference group. Coefficient estimates of the main independent variable *Sibling-status info* show *no* significant differences in callbacks between resumes with the sibling-status signal and those without ($p > 0.10$, Columns [1] and [4]); the interaction term between *Sibling-status info* and *Male* shows *no* differential gender gap in

[31] Dividing the job ads into two bins was to avoid an alternative design that was to send to each job ad six resumes including male (or female) with sibling-child status, male (or female) with only-child status, and male (or female) without such information revealed. We did not choose this alternative design since it would involve submitting too many resumes for each ad.

callbacks between those with and without signals included ($p > 0.10$, Columns [2], [3], [5], and [6]). The results from the Probit analysis are very similar and thus omitted.

Table A4. Determinants of Callbacks in the Companion Experiment in Shanghai (OLS)

Variables	No Info. vs. Siblinged Status [1]	[2]	[3]	No Info. vs. Only-Child Status [4]	[5]	[6]
Sibling-status info.	-0.025	-0.025	-0.021	0.021	0.017	0.006
	(0.020)	(0.030)	(0.031)	(0.028)	(0.038)	(0.041)
Male		-0.092	-0.085		-0.034	-0.084
		(0.041)	(0.043)		(0.042)	(0.052)
Sibling-status info × male		0.000	-0.001		0.009	0.049
		(0.053)	(0.059)		(0.051)	(0.055)
Constant	0.488	0.533	1.840	0.410	0.427	-1.765
	(0.041)	(0.046)	(1.749)	(0.041)	(0.046)	(1.606)
Control for other variables	No	No	Yes	No	No	Yes
Observations	480	480	480	468	468	468

Notes: The dependent variable is an indicator variable for whether a fictitious applicant receives a callback. Standard errors clustered at the job-ad level are in parentheses. The control variables (i.e., applicant, employer, and job characteristics) in Columns [3] and [6] are the same as in Columns [3] and [6] in Table 2.

Appendix F. Robustness Check of Main Results (Table 2) with Probit

Table A5. Determinants of the Callbacks (Probit)

	[1]	[2]	[3]
Regression estimates			
Male only-child before	0.008	0.002	0.001
(MOB)	(0.014)	(0.015)	(0.015)
Female with siblings before	0.051	0.050	0.051
(FSB)	(0.013)	(0.013)	(0.013)
Male with siblings before	0.006	0.002	-0.000
(MSB)	(0.014)	(0.014)	(0.014)
Female only-child after	0.063	0.059	0.067
(FOA)	(0.020)	(0.024)	(0.024)
Male only-child after	-0.004	-0.014	-0.006
(MOA)	(0.020)	(0.025)	(0.024)
Female with siblings after	0.041	0.036	0.045
(FSA)	(0.020)	(0.024)	(0.024)
Male with siblings after	-0.000	-0.008	-0.001
(MSA)	(0.020)	(0.024)	(0.024)
Cross-sectional DID (%)			
(FOB − FSB) − (MOB − MSB)	-5.30	-5.09	-5.18
Temporal DIDs (%)	[0.004]	[0.007]	[0.006]
(FSA − FOA) − (FSB − FOB)	-7.31	-7.32	-7.28
	[<0.001]	[<0.001]	[<0.001]
(MSA − MOA) − (MSB − MOB)	0.55	0.58	0.51
	[0.756]	[0.744]	[0.774]
Triple differences (%)			
[(FSA − FOA) − (FSB − FOB)] −	-7.86	-7.90	-7.79
[(MSA − MOA) − (MSB − MOB)]	[0.002]	[0.002]	[0.002]
Applicant characteristics	No	Yes	Yes
Employer and job characteristics	No	No	Yes
City fixed effects	Yes	Yes	Yes
Observations	8,848	8,848	8,848

Notes: The empirical specifications in this table are the same as those in Table 2. The only difference is that we use Probit and report marginal effects here. Results are very similar to the OLS results in Table 2.

Appendix G. Investigation on Other Potential Contemporaneous Policies

To verify the trend assumptions required by our temporal DIDs and triple differences, we thoroughly examined government policies and regulations that applied to the four cities during our two-wave experiment. We find that *no* other contemporaneous policies or regulations affected callbacks to trend differently for the two sibling-status types for either gender or differentially affected callback of the siblinged women, relative to the other three groups (FO/MS/MO). Therefore, the required assumptions are very likely to hold for both our temporal DIDs and triple differences. In this appendix, we detail our methodology and procedures.

In China, the Ministry of Human Resources and Social Security (MOHRSS) of the central government is a part of the State Council "which is responsible for national labor policy standards, regulations - and managing the national social security system" as well as "labor force management, labor relations, social insurance management and legal construction of labor" (China's State Council website). Its subordinate departments and bureaus are responsible for these functions on the provincial and local government levels. The two directly administered municipalities, Beijing and Shanghai, have the same rank as provinces, whereas Guangzhou and Shenzhen fall under the Guangdong Province. Our examination thus encompassed policies and regulations promulgated and implemented during our two-wave study by all the levels of relevant government agencies: the Ministry of Human Resources and Social Security on the national level, the Human Resources and Social Security Department of the Guangdong Province, and the Municipal Human Resources and Social Security Bureau in each of the four cities.

To identify potentially relevant policies and regulations, we started with a comprehensive search based on the following categories with keywords included in parentheses: gender

(gender *xìng bié*, male *ná*, and female *nǚ*), marriage (marry/marriage *hūn/hūn yīn*, married *yǐ hūn*, and unmarried *wèi hūn*), family (child(ren) *zǐ nǚ*, and family *jiā tíng*), fertility (fertility *shēng yù*, pregnancy *yùn*, maternity/paternity leave *chǎn jià*, and nursing *pǔ rǔ*), and sibling status (only-child *dú shēng*, and siblinged *fēi dú/fēi dú shēng*). We screened out the policies that covered unrelated topics (e.g., medical insurance, prescription drugs, medical supplies, management of medical and health professionals), industries (e.g., constructions, steel, and coal, and startup businesses), or populations (e.g., overseas talents, K-9 teachers, and rural migrant works). We found only one policy that might bear some relevance to our study: MOHRSS and the Ministry of Finance lowered the cap on employer-paid premiums for maternity insurance from 1% to 0.5% of total wage payments, effective on October 1, 2015.[32]

Maternity insurance is a part of China's social insurance system in addition to pensions and medical, unemployment, and work-injury insurance. It has several key features. For example, employers, *not employees*, are responsible for paying premiums according to a certain percentage of total wage payments for *all* employees, regardless of gender. Female and male employees are all eligible for insurance benefits. Since neither employee gender nor sibling status affects employer costs and employee benefits, reducing the cap on employer-paid premiums should not affect their decisions on hiring based on prospective employees' gender and sibling status.

To verify this point more objectively, we conducted a survey among 56 MBA students at the Beijing Normal University who had previous work experience in human resources management. The survey asked the students how familiar they were with the features of the

[32] The specific contribution rate is determined by the municipal government based on the local economic conditions, upon the approval of the provincial government.

maternity insurance mentioned above, whether the premium reduction would affect their decisions on hiring male versus female and only-child versus siblinged applicants and whether they thought the premium reduction would affect such decisions for *other recruiters*. Table A6 reports the summary results. We find only a negligible 5% (or 4%) of these MBA students reported that the premium decrease would affect their (or other recruiters') decisions on hiring men or women, but none of those who responded yes to these questions had previous *recruiting* experience. In addition, no one reported that the premium reduction would affect their or other recruiters' decisions on hiring only-child or siblinged applicants.

In sum, we find no policies or regulations contemporaneous with the 2016 universal two-child policy that would affect the parallel trend assumptions required by our temporal DIDs and triple differences. In addition, no contemporaneous policies or regulations would result in different time trends for men and women.

Table A6. MBA Survey on Premium Reduction of Employers Paid Maternity Insurance

Variables	Mean (Standard deviation)
1. Familiarity with key features of maternity insurance	4.17
(1 very unfamiliar ~ 5 very familiar)	(1.62)
2. Would affect my decision on hiring men/women	0.05
(1 Yes; 0 No)	(0.23)
3. Would affect my decision on hiring only-child/siblinged applicant	0
(1 Yes; 0 No)	(0)
4. Would affect other recruiters' hiring decisions on men/women	0.04
(1 Yes; 0 No)	(0.19)
5. Would affect other recruiters' hiring decision on only-child/siblinged applicant	0
(1 Yes; 0 No)	(0)

Appendix H. Beijing

Our data collection in Beijing lasted two months, from October 10 to December 24, 2015 (Table A1, Appendix C). The government announced the policy change on October 29, 2015, resulting in some complications in the Beijing data. In this appendix, we report why we exclude Beijing from the main analysis, but our results (Table 2 in the main text) are robust if we include Beijing. We also investigate if the employers responded to the policy announcement, using the sample including Beijing.

Figure A4 illustrated how the policy announcement affected the data we gathered in Beijing. Since the policy change was announced on October 29, our two-week callback window meant the last day to submit a resume to receive a possible callback *before* the policy change was October 15, 2015. Thus, the 39 jobs for which all four resume submissions were completed by October 15 were free of any potential interference from the policy announcement.

The data associated with the 93 jobs we applied to between October 16 and 29 (both inclusive and highlighted by the gray bar in Figure A4) were subject to potential confounds; it was unclear whether a callback decision was made before or after the policy announcement. For example, among the four resumes submitted for a specific job during this time, it was possible for some decisions to occur before the policy announcement and some afterward. Even cases in which we received callbacks after the policy announcement could have been decided beforehand.

For the 265 jobs for which we sent resumes after October 29, callback decisions were undoubtedly made after the policy announcement. However, it was unclear whether and to what extent the employers responded to the *announcement* itself before the new policy took effect on January 1, 2016. This ambiguity introduced an unknown degree of disparity between Beijing's *post-announcement* data and the *post-policy-change* data of the other three cities.

In sum, since 90% of the Beijing data were subject to these potential confounds resulting from the policy announcement, we excluded Beijing from the main analysis. We next conduct robustness checks by adding the data back to the analysis.

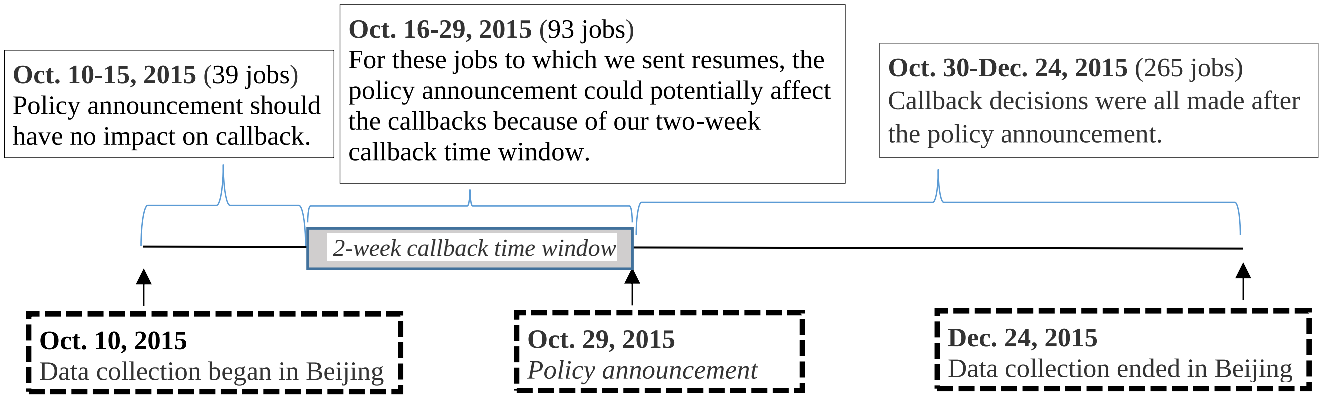

Figure A4. Policy Announcement and Data Collection in Beijing

Table A7 reports the robustness checks. To tackle the data confounds mentioned above, we gradually incorporate the Beijing data into the main, three-city analysis (OLS, Column [3] of Table 2) in the following ways: adding only the "clean" portion of the pre-announcement data (i.e., the 39 jobs applied for by October 15) in Column [1], adding all the pre-announcement data (i.e., the 132 jobs applied for by October 29) in Column [2], adding the "clean" portion of the pre-announcement data and all the post-announcement data (i.e., the 39 jobs applied for by October 15 and the 265 jobs applied for after October 29) in Column [3], and adding all the Beijing data (i.e., all 397 jobs) in Column [4]. In other words, Columns [1] and [2] only add Beijing's pre-announcement data to the three-city main dataset, while Columns [3] and [4] further add post-announcement data. Columns [1] and [3] differ from Columns [2] and [4] by excluding the resumes associated with the 93 jobs for which we made a pre-announcement application between October 16 and 29, 2015. We also include a Beijing dummy as we do for other cities in the main analysis to control for the city fixed effects. The Probit results are very similar and hence omitted. The comparisons of Table A7 with Column [3] of Table 2 show that all our main results—DIDs and triple differences—are robust no matter whether we add the Beijing data partially or entirely.

Next, we investigate whether employers in Beijing responded to the *announcement* of the policy change, using the policy announcement on October 29 as a source of identification. We provide some caveats to this analysis. The unexpected policy announcement arbitrarily split our planned first wave of data collection in Beijing. This caused an imbalance in the industries and occupations that employers represented before and after the announcement, unlike in the other three cities where the industries and occupations were, by design, balanced before and after the policy change. Moreover, the estimates may lack statistical precision due to the limited numbers of pre- and post-announcement observations. As shown in Table A8, we augment the empirical models in Columns [3] and [4] of Table A7 by interacting the Beijing dummy and the other-three-cities dummy with the four post-announcement (or post-policy-change) resume types FOA, MOA, FSA, and MSA, respectively. These interactions allow us to separately estimate the impact of policy *announcement* on employers' decisions in Beijing and the impact of *policy change* in the other three cities. For the reader's convenience, Table A8 reports the regression estimates in the top panel, test results for Beijing in the middle panel, and test results for the other three cities in the bottom panel. We find that the DIDs and triple differences for the other three cities pooled are similar to the main results in Column [3] of Table 2 as well as the robustness check results in Columns [3] and [4] of Table A7. In addition, the test results for Beijing are qualitatively consistent with the main results in Column [3] of Table 2 but statistically insignificant ((FSA − FOA) − (FSB − FOB) = -2.66%, $p = 0.332$; triple differences = -3.94%, $p = 0.343$ in Column [2]), showing little evidence for employers' response to the policy *announcement per se* in Beijing.

Table A7. Robustness Check on Main Results by Incorporating Beijing Data (OLS)

	Adding the "clean" pre-announcement data by Oct. 15	Adding all pre-announcement data	Adding all Beijing data except pre-announcement data between Oct. 16 and 29, 2015	Adding all Beijing data
	[1]	[2]	[3]	[4]
Regression estimates				
Male only-child before	0.003	0.001	0.002	-0.001
(MOB)	(0.014)	(0.014)	(0.014)	(0.014)
Female with siblings before	0.050***	0.044***	0.049***	0.044***
(FSB)	(0.013)	(0.013)	(0.013)	(0.013)
Male with siblings before	0.000	-0.002	0.000	-0.003
(MSB)	(0.014)	(0.013)	(0.014)	(0.013)
Female only-child after	0.069***	0.067***	0.046**	0.039*
(FOA)	(0.023)	(0.023)	(0.022)	(0.021)
Male only-child after	-0.002	-0.003	-0.014	-0.020
(MOA)	(0.023)	(0.023)	(0.022)	(0.021)
Female with siblings after	0.046**	0.044*	0.031	0.024
(FSA)	(0.023)	(0.023)	(0.022)	(0.021)
Male with siblings after	0.001	-0.001	-0.008	-0.015
(MSA)	(0.023)	(0.023)	(0.022)	(0.021)
Cross-sectional DID (%)				
(FOB − FSB) − (MOB − MSB)	-5.21	-4.74	-5.05	-4.55
	[0.005]	[0.007]	[0.006]	[0.010]
Temporal DIDs (%)				
(FSA − FOA) − (FSB − FOB)	-7.22	-6.68	-6.45	-5.90
	[<0.001]	[<0.001]	[<0.001]	[<0.001]
(MSA − MOA) − (MSB − MOB)	0.55	0.53	0.69	0.68
	[0.749]	[0.754]	[0.677]	[0.675]
Triple Differences (%)				
[(FSA − FOA) − (FSB − FOB)] - [(MSA − MOA) − (MSB − MOB)]	-7.77	-7.21	-7.14	-6.58
	[0.002]	[0.003]	[0.003]	[0.005]
Applicant characteristics	Yes	Yes	Yes	Yes
Employer and job characteristics	Yes	Yes	Yes	Yes
City fixed effects	Yes	Yes	Yes	Yes
Observations	9,004	9,376	10,084	10,456

Notes: The analysis in this table uses the OLS empirical specification with the complete set of covariates in Column [3] of our main analysis in Table 2. Columns [1] and [2] only add Beijing's pre-announcement data to the three-city main dataset, while Columns [3] and [4] further add post-announcement data. Columns [1] and [3] differ from Columns [2] and [4] by excluding pre-announcement data collected between October 16 and 29, 2015.

Table A8. Impact of Policy *Announcement* and Policy Change (OLS)

	Three cities plus all Beijing data except pre-announcement data between Oct. 16 and 29, 2015 [1]	Three cities plus all Beijing data [2]
Regression estimates		
Male only-child before	0.002	-0.0003
(MOB)	(0.014)	(0.014)
Female with siblings before	0.049***	0.044***
(FSB)	(0.013)	(0.013)
Male with siblings before	0.0003	-0.002
(MSB)	(0.014)	(0.013)
Female only-child after, Beijing	-0.112*	-0.066
(FOA)	(0.064)	(0.042)
Male only-child after, Beijing	-0.122*	-0.075*
(MOA)	(0.064)	(0.042)
Female with siblings after, Beijing	-0.096	-0.049
(FSA)	(0.064)	(0.043)
Male with siblings after, Beijing	-0.111*	-0.064
(MSA)	(0.063)	(0.042)
Female only-child after, other 3 cities	0.068***	0.066***
(FOA)	(0.023)	(0.023)
Male only-child after, other 3 cities	-0.004	-0.005
(MOA)	(0.023)	(0.023)
Female with siblings after, other 3 cities (FSA)	0.045*	0.043*
	(0.023)	(0.023)
Male with siblings after, other 3 cities (MSA)	-0.0003	-0.002
	(0.023)	(0.023)
Beijing cross-sectional DID (%)		
(FOB − FSB) − (MOB − MSB)	-5.08	-4.60
	[0.006]	[0.009]
Beijing temporal DIDs (%)		
(FSA − FOA) − (FSB − FOB)	-3.25	-2.66
	[0.242]	[0.332]
(MSA − MOA) − (MSB − MOB)	1.26	1.28
	[0.684]	[0.678]
Beijing triple differences (%)		
[(FSA − FOA) − (FSB − FOB)] − [(MSA − MOA) − (MSB − MOB)]	-4.51	-3.94
	[0.282]	[0.343]
3 cities cross-sectional DID (%)		
(FOB − FSB) − (MOB − MSB)	-5.08	-4.60
	[0.006]	[0.009]
3 cities temporal DIDs (%)		
(FSA − FOA) − (FSB − FOB)	-7.23	-6.69
	[<0.001]	[<0.001]
(MSA − MOA) − (MSB − MOB)	0.55	0.53
	[0.749]	[0.752]
3 cities triple differences (%)		
[(FSA − FOA) − (FSB − FOB)] −	-7.77	-7.22

[(MSA − MOA) − (MSB − MOB)]	[0.002]	[0.003]
Applicant characteristics	Yes	Yes
Employer and job characteristics	Yes	Yes
City fixed effects	Yes	Yes
Observations	10,084	10,456

Notes: The analysis in this table uses data of all four cities and extends the OLS empirical specification in Columns [3] and [4] of Table A7 by interacting the Beijing dummy and the other-three-cities dummy with the four post-policy resume types FOA, MOA, FSA, and MSA. The Beijing temporal DIDs and triple differences in the middle panel provide estimates on the impact of policy *announcement*; the three cities' DIDs and triple differences in the lower panel provide estimates of the impact of policy *change*.

Appendix I. Robustness Checks on Potential Detection Risks

One concern in sending multiple correspondences with uncommon signals in a correspondence study is that employers may sense they are being studied and hence behave differently (Balfe et al., 2021). We use two sets of robustness checks to investigate this question.

We first examine whether the order of resume submission affects callbacks. We extend our benchmark empirical specification in Column [3] of Table 2 by adding resume-submission-order dummies. We present the results in Table A9. Column [1] simply incorporates these order dummies (i.e., Order2, Order3, and Order4) with the first submission as the omitted reference category. We find that, on average, the second or the fourth resumes are less likely to receive callbacks relative to those submitted first or third ($p = 0.062$ between the second and the third, $p < 0.05$ in other pairwise comparisons). However, there is no callback difference between the first and third resumes ($p = 0.472$) or between those submitted second and fourth ($p = 0.664$). Therefore, our analysis does not indicate a consistent pattern of order effect, although the second and the fourth resumes submitted are associated with lower callback rates. Importantly, controlling for the submission orders does not affect any estimates of the resume types and other covariates or the DIDs and triple differences, compared to the main results reported in Table 2. Column [2] further includes interactions of the order dummies with the resume types (i.e., MO, FS, and MS) to allow for the potential, differential impact of submission order on different types of resumes.[33] We find no differential impact of submission orders on resume types ($p > 0.10$). In addition, the DIDs and triple differences that are evaluated for the average submission order are similar to the main results in Table 2. Therefore, we find that our results are robust to the order of resume submission. This should be unsurprising since we randomized the submission orders across the four resume types in our experiment.

[33] Note we assume that the differential impact, if any, does not depend on the policy environments. Therefore, we do not interact the submission orders with the policy-dependent resume types (e.g., MOB, MOA).

Table A9. Resume-Submission Order (OLS)

	[1]	[2]
Regression estimates		
Male only-child before	0.004	-0.019
(MOB)	(0.015)	(0.030)
Female with siblings before	0.052***	0.068**
(FSB)	(0.013)	(0.030)
Male with siblings before	0.003	0.003
(MSB)	(0.014)	(0.030)
Female only-child after	0.070***	0.070***
(FOA)	(0.023)	(0.023)
Male only-child after	-0.002	-0.024
(MOA)	(0.023)	(0.034)
Female with siblings after	0.047**	0.064*
(FSA)	(0.023)	(0.035)
Male with siblings after	0.002	0.002
(MSA)	(0.023)	(0.035)
Order2	-0.024**	-0.030
	(0.009)	(0.028)
Order3	-0.007	-0.011
	(0.010)	(0.028)
Order4	-0.028***	-0.023
	(0.010)	(0.028)
MO × Order2		0.018
		(0.041)
MO × Order3		0.048
		(0.042)
MO × Order4		0.022
		(0.041)
FS × Order2		0.010
		(0.044)
FS × Order3		-0.026
		(0.043)
FS × Order4		-0.045
		(0.042)
MS × Order2		0.000
		(0.041)
MS × Order3		-0.004
		(0.042)
MS × Order4		0.004
		(0.042)
Cross-sectional DID (%)		
(FOB − FSB) − (MOB − MSB)	-5.24	-6.92
	[0.006]	[0.007]
Temporal DIDs (%)		
(FSA − FOA) − (FSB − FOB)	-7.47	-7.47
	[<0.001]	[<0.001]
(MSA − MOA) − (MSB − MOB)	0.46	0.46
	[0.793]	[0.791]

Triple differences (%)		
[(FSA − FOA) − (FSB − FOB)] −	-7.93	-7.93
[(MSA − MOA) − (MSB − MOB)]	[0.002]	[0.002]
Applicant characteristics	Yes	Yes
Employer and job characteristics	Yes	Yes
City fixed effects	Yes	Yes
Observations	8,848	8,848

Notes: The analysis in this table extends the OLS empirical specification in Column [3] of Table 2 by adding the resume submission orders (Column [1]) and further interacting the orders with resume types (Column [2]). * DIDs and triple differences in Column [2] are all evaluated for the average submission order.

Balfe et al. (2021) propose another approach to analyze whether employers may detect a correspondence study that sends multiple uncommon signals, or whether spillover issues may occur when fictitious resumes interact with one another or potentially affect the quality of an employer's entire applicant pool (Phillips, 2019). Our second approach, in line with the spirit of Balfe et al. (2021), is to conduct robustness checks by restricting our sample in two ways (Table A10). The only-child signal may be more likely to be perceived as odd by employers compared to the sibling-child signal before the policy change. Therefore, Column [1] restricts the sample to cases where employers in our study have only been exposed to *one* only-child resume, i.e., we drop all second only-child resumes and any resumes sent afterward. This restricted sample, based on the idea underlying the restricted Sample #1 in Balfe et al. (2021), may still be detectable since many employers are still exposed to multiple resumes in our study unless the first resumes they receive include the only-child signal. To address this concern, Column [2] of Table A10 further restricts the sample to include only the *first* resumes submitted for all the job openings, so each employer has been exposed to only *one* fictitious resume with one signal. Note while the first restricted sample in Column [1] compromises some benefits of our original balanced within-subject design, the second restricted sample in Column [2] relies entirely on the between-employer variations and should provide the cleanest, detection-free estimates on discrimination against women on the basis of family responsibilities. However, both restricted samples come with the cost of reduced statistical power compared to the full sample, with the second restricted sample sacrificing more statistical power since it comprises only one-quarter of the full dataset.

Results show that temporal DIDs are -9.39% (p = 0.002) for women and 0.85% (p = 0.767) for men, and the triple differences are -8.54% (p = 0.061) in Column [1] using the first restricted sample. Temporal DIDs are -10.99% (p = 0.052) for women and 0.08% (p = 0.988) for men, and the triple differences are -11.07% (p = 0.152) in Column [2] using only the *first* resumes submitted for all the job openings. Although these estimates are largely consistent with those from the full sample, they show somewhat increases in magnitude (albeit with lower statistical precision), suggesting even greater discrimination against women.

Table A10. Tests on Potential Detection and Spillover Risks Using Restricted Samples

	Employers were exposed to only *one* only-child resume [1]	Including only the *first* resumes submitted to jobs [2]
Regression estimates		
Male only-child before	-0.016	-0.053
(MOB)	(0.028)	(0.041)
Female with siblings before	0.070***	0.074*
(FSB)	(0.022)	(0.041)
Male with siblings before	-0.003	-0.016
(MSB)	(0.022)	(0.040)
Female only-child after	0.063**	0.063
(FOA)	(0.031)	(0.045)
Male only-child after	-0.023	-0.046
(MOA)	(0.032)	(0.046)
Female with siblings after	0.039	0.027
(FSA)	(0.030)	(0.045)
Male with siblings after	-0.019	-0.009
(MSA)	(0.029)	(0.045)
Cross-sectional DID (%)		
(FOB − FSB) − (MOB − MSB)	-5.69***	-3.69
	[0.090]	[0.510]
Temporal DIDs (%)		
(FSA − FOA) − (FSB − FOB)	-9.39***	-10.99**
	[0.002]	[0.052]
(MSA − MOA) − (MSB − MOB)	-0.85	0.08
	[0.767]	[0.988]
Triple differences (%)		
[(FSA − FOA) − (FSB − FOB)] −	-8.53*	-11.07
[(MSA − MOA) − (MSB − MOB)]	[0.061]	[0.152]
Applicant characteristics	Yes	Yes
Employer and job characteristics	Yes	Yes
City fixed effects	Yes	Yes
Observations	5,085	2,212

Notes: The analysis in this table uses our benchmark OLS empirical specification in Column [3] of Table 2 but restricts the sample to reduce the concerns of potential detection and spillover risks. Column [1] restricts the sample to when employers in our study have only been exposed to *one* only-child resume, i.e., we drop all second only-child resumes and any resumes sent afterward. This sample restriction is in the spirit of the restricted Sample #1 in Balfe et al. (2021). Column [2] further restricts the sample to include only the *first* resumes submitted for all the job openings.

Appendix J. Possible Gender Segregation

Table A11. Callback Rates by Occupation (%)

		Men		Women	
		Treated Siblinged	Control Only child	Treated Siblinged	Control Only child
Sales	Before policy	MSB 37.4 (706)	MOB 38.2 (706)	FSB 39.7 (706)	FOB 34.0 (706)
	After Policy	MSA 36.0 (706)	MOA 36.4 (706)	FSA 37.8 (706)	FOA 40.4 (706)
Administrative Assistance	Before policy	MSB 6.0 (133)	MOB 5.3 (133)	FSB 16.5 (133)	FOB 12.8 (133)
	After Policy	MSA 6.8 (133)	MOA 7.5 (133)	FSA 16.5 (133)	FOA 15.8 (133)
Customer Service	Before policy	MSB 25.8 (267)	MOB 24.7 (267)	FSB 33.3 (267)	FOB 28.5 (267)
	After Policy	MSA 26.6 (267)	MOA 24.0 (267)	FSA 34.1 (267)	FOA 37.1 (267)

Notes: This table reports the callback rates by gender, sibling status, and occupation before and after the policy change, with the number of resumes in parentheses. The DIDs and triple differences derived based on this table are presented in Table 4.

Additional References in Appendixes

China's State Council, Ministry of Human Resources and Social Security, URL: http://english.www.gov.cn/state_council/2014/09/09/content_281474986284102.htm (accessed on June 10, 2021).

51job website, URL: www.51job.com/bo/AboutUs_e.php, accessed on December 16, 2020.

Neumark, David, Ian Burn, Patrick Button, and Nanneh Chehras. "Do state laws protecting older workers from discrimination reduce age discrimination in hiring? Evidence from a field experiment." *The Journal of Law and Economics* 62, no. 2 (2019): 373-402.

Phillips, David C. "Do comparisons of fictional applicants measure discrimination when search externalities are present? Evidence from existing experiments." *The Economic Journal* 129, no. 621 (2019): 2240-2264.

Wang, Fei, Liqiu Zhao, and Zhong Zhao. "China's family planning policies and their labor market consequences." *Journal of Population Economics* 30, no. 1 (2017): 31-68.

Wu, J. and Yang, J. "Zhang, Wang, Li, Zhao, who is the most common: The structure and distribution of surnames in China 2010 Population Census." *China Statistics* (in Chinese), 6 (2014): 21-22.