ESSAYS ON GENDER INEQUALITY IN CHINA

A Dissertation

Presented to

The Faculty of the Department

of Economics

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Xuejing Zuo

May, 2018

ESSAYS ON GENDER INEQUALITY IN CHINA

Xuejing Zuo

APPROVED:

Elaine M. Liu, Ph.D.
Committee Chair

Chinhui Juhn, Ph.D.
Committee Co-Chair

C. Andrew Zuppann, Ph.D.

Jessica Leight, Ph.D.
American University

Antonio D. Tillis, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of Hispanic Studies

Abstract

The dissertation consists of two applied microeconomics studies on gender inequality in China. In the first study, my coauthors and I explore the multiple switching behavior (MSB) in Multiple Price List (MPL) instrument. This instrument has been widely used to measure gender differences in risk preferences in experimental economics. MSB is believed to indicate low quality decision making. We develop a "nudge" protocol for the MPL that reduced multiple switching behavior (MSB) from 31% to 10% (p-value <0.001) without limiting the choice set. We further develop a conceptual framework to formally test three leading explanations for the nature of low quality decision-making in the MPL using the covariance of responses in the MPL with a second, simple risk instrument. Using a counter-balanced within and between-group experimental design, we find that low quality decision-making in the MPL is best explained by task-specific miscomprehension.

In the second study, I examine the effect of the State-owned enterprise (SOE) reform on gender inequality in labor market outcomes. Between 1996 and 2001, the national SOE reform resulted in a massive layoff of over 35 million workers in urban China. I employ both difference-in-differences and instrumental variable strategies to identify the causal impact of this SOE on gender gaps in employment and earnings. I find that that the reform negatively affected women's labor market outcomes substantially more than men's outcomes. To explore whether the response to the SOE reform in related with traditional gender norms, I calculate male-to-female sex ratio using under age 10 cohorts and I find that the widening gender employment gap is entirely driven by the areas with high sex ratio. But the impact on gender rank gap is small and not significant in both areas.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor Prof. Elaine M. Liu for the continuous support of my Ph.D study and related research, for her patience, motivation, and immense knowledge. Without her guidance and persistent help this dissertation would not have been possible.

Besides my advisor, I would like to thank the rest of my dissertation committee: Prof. Chinhui Juhn and Dr. Andrew Zuppann, for their insightful comments and encouragement. My sincere appreciation also goes to Prof. Aimee Chin, Dr. Willa Friedman, Dr. Jessica Leight, and Dr. Fan Wang, who provided me thoughtful feedback on this dissertation.

Last but not the least, I am indebted to the unconditional love from my family in China: for my parents and aunts who support me whenever I need, for my younger sister who took all the family responsibility which I should have taken, for my husband who never doubts the importance of what I am pursuing.

Contents

List of Figures

List of Tables

Chapter 1

Multiple Switching and Data Quality in the Multiple Price List (with Y. Jane Zhang and Chi Wai Yu)

1.1 Introduction

Reliable and meaningful measurement of individual risk preferences is critical for understanding a wide range of economic decision-making. Experimental economics has contributed many tools to measure individual risk preferences using incentivized choice situations (see Harrison and Rutström (2008) for a survey), although different risk elicitation methods are not always correlated with one another (see, for example, Charness and Viceisza (2016) and the references in Niederle (2016)). The Multiple Price List (MPL) instrument, often called the Holt and Laury (2002) instrument, is one of the most widely used methods to elicit risk individual preferences. The MPL is also used in settings other than the measurement of risk preferences, such as pricing commodities (Kahneman et al., 1990; Cassar et al., 2016) and measuring discount rates (Harrison et al., 2002; Andersen et al., 2008). An attractive feature of the MPL is that it can be used to elicit arbitrarily precise intervals of risk aversion estimates (Charness et al., 2013; Tanaka et al., 2010; Jacobson and Petrie, 2009).

Despite its popularity, an empirical difficulty that researchers encounter when using the

MPL is that a substantial proportion of subjects switch back and forth between the safe and the risky choice columns in the instrument (i.e., engage in multiple switching), which is behavior incompatible with standard assumptions on preferences (Charness et al., 2013). Such multiple switching behavior (MSB) is generally considered low quality decision-making, and the observed responses are treated as noise, although some studies argue that MSB may indicate indifference between a range of options (Andersen et al., 2006). MSB is especially pronounced in developing countries. Whereas typical studies in developed countries find multiple switching to affect approximately 10% of the subjects (e.g., Holt and Laury (2002) reports 13% multiple switchers; Dave et al. (2010) report 8.5% multiple switchers), in developing countries the multiple switching rate can be over 50% (Jacobson and Petrie, 2009; Charness and Viceisza, 2016).

A common experimental practice used to reduce MSB is to ask subjects to indicate the row in which they would like to switch from the risky option to the safe option (e.g., Andersen et al. (2006); Tanaka et al. (2010)). This eliminates MSB, but also reduces the choice set, so we do not know whether the subject would have engaged in multiple switching if they had been free to do so. In this study we develop a "nudge" protocol to increase cognitive effort without limiting the choice set.[1] After subjects complete the MPL task, we ask them if they are sure of their responses and give them the option of hearing the instructions one more time. We found a reduction in MSB from 31% using the standard protocol to 10% in the nudge protocol (p-value of difference < 0.001). This suggests that at least 2/3 of MSB can be categorized as mistakes which are corrected upon further reflection, which sets it apart from the deliberate randomization behavior described in Agranov and Ortoleva (2017), for example.

Although the literature generally views MSB as equivalent to low decision-making quality, an individual can make low quality decisions that do not result in MSB. The potential for non-multiple switchers to make low quality decisions is not well understood and may be an important source of noise in the data. We develop a conceptual framework, which formally defines decision-making quality independently of MSB, and, using the covariance between responses on the MPL task and responses on a simpler lottery selection task, allows us to test between three explanations for low-quality decision-making in the MPL suggested by the findings in the

[1] A nudge, as defined by Thaler and Sunstein (2008) is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."

literature.

Of the three explanations, perhaps the most pessimistic is that bad decision-making in the MPL task is a stable attribute of the decision-maker. This explanation is supported by the evidence in Jacobson and Petrie (2009), which shows that multiple switchers in MPL instruments make sub-optimal decisions in other areas of their lives. Similarly, Choi et al. (2014) finds that people make bad decisions in many areas of their lives and individuals who make low-quality decisions in an experimental setting have less wealth, controlling for their current income and a slew of demographic and socioeconomic status variables.[2] Under this view, decision-making quality is not readily improvable, and should not respond to an unintrusive stimulus such as the nudge treatment.

At first pass, this explanation seems immediately incompatible with our finding that MSB can be reduced by the nudge protocol. However, since we do not make the assumption that MSB captures the full extent of low quality decision-making, the finding that MSB can be reduced does not necessarily imply that decision-making quality can be improved. For example, individuals could have inferred that experimenters wanted less MSB and in their desire to be helpful make fewer multiple switches but continue to give noisy responses that do not reflect their true risk preferences. Charness et al. (2013) raises a similar concern in the context of treatments that eliminate MSB but do not induce higher quality decision-making, which would mask data quality issues.

A second explanation is that low quality decision-making in the MPL is incidental to the complexity of the MPL instrument. For example, Charness and Viceisza (2016) argue that a lack of comprehension is a serious concern with using the MPL in developing countries. Charness et al. (2013) uses the term "failure to understand". Andreoni and Sprenger (2011) directly tells subjects that "Most people begin by preferring Option A and then switch to Option B, so one way to view this task is to determine the best row to switch from Option A to Option B," in an effort to improve comprehension. Furthermore, Dave et al. (2010) demonstrates that the MPL produced noisier estimates of risk aversion than did a simple lottery selection instrument developed in Eckel and Grossman (2002), especially for low math ability individuals,

[2] Low quality decisions in Choi et al. (2014) are defined as choices inconsistent with the generalized axiom of revealed preferences in their experiment which asks subjects to choose bundles of goods under varying budget slopes.

which suggests that cognitive ability plays a role in comprehension of the MPL. Under this interpretation, short-term treatments such as the nudge protocol can improve decision-making quality in the MPL by improving comprehension.

A third explanation also views low quality decisions in the MPL as improvable in the short-term, but assumes that individuals who make low-quality decisions in the MPL also make low-quality decisions using other instruments because they are careless. This explanation was mentioned in Brick et al. (2012) but has not been given much attention in the literature. The main difference between carelessness and miscomprehension is that low quality decision-making due to carelessness is not unique to the MPL instrument. The implication is that the MPL, by virtue of allowing for MSB, which may be an indicator of low-quality decision making, may in fact be preferred to simpler instruments that do not allow for MSB and which obscures low quality decision-making.

To test the different explanations, we compare responses given in the MPL to responses given in a simple lottery selection task (henceforth, LS) in which MSB is not possible, in the control and nudge treatment groups.[3]

In the conceptual framework that we develop below, we show that the covariance of the responses given in the MPL and the LS tasks will be the same for the control and nudge groups under the stable attribute explanation, it will be larger for the nudged group than the control group under the miscomprehension explanation, and will be weakly larger for the control group than the nudge group under the carelessness explanation. To make inferences about the difference between two population covariances, we derived the variance of the difference between two sample covariances and propose an estimator which is consistent and asymptotically normal (see Appendix B.1.2).

Our test finds consistent evidence in support of the miscomprehension explanation. Furthermore, this finding in conjunction with our conceptual framework motivates a novel metric to quantify data quality using the correlation of the MPL task and the simpler LS task, which is independent of the multiple switch rate. Using this metric we find that the nudge treat-

[3]The lottery selection task was independently developed in Binswanger (1980) and Eckel and Grossman (2002). To give a brief explanation of the LS task, it requires subjects to select one out of six different coin-flip lotteries. Lotteries with higher expected value also have higher variance. Selecting a lottery with higher expected value (and higher variance) is indicative of higher risk tolerance. In this task, all choices are compatible with standard assumptions on preferences.

ment conservatively improves data quality by 143%. Data from both multiple switchers and non-multiple switchers under the standard protocol are characterized by low decision-making quality, implying that MSB does not capture the full extent of low quality decision-making, and that discarding multiple switchers does not ensure high data quality, as is often believed. Although cognitive ability predicts MSB, which is consistent with the previous literature, cognitive ability is not a significant determinant of data quality. To the extent that the nudge treatment was able to elicit more cognitive effort from respondents, the findings suggest that cognitive ability is not a limiting factor in achieving high data quality in the MPL, but cognitive effort is. Increased cognitive effort has also been proposed as an explanation for the reduction in MPL errors in choice situations where subjects can incur a loss (Von Gaudecker et al., 2011) and when stakes are high (Holt and Laury, 2002). This relates to the literature on bounded rationality, which finds cognitive effort plays an important role in decision-making independent of intelligence, and that humans are prone to cognitive laziness (Kahneman, 2003, 2011).

This paper is organized as follows. In Section 2 we develop our conceptual framework. In Section 3 we describe our experimental design and our subjects. Section 4 presents the empirical analysis. Section 5 presents additional results and Section 6 concludes.

1.2 Conceptual Framework

In this section, we develop a simple conceptual framework (see Figure 1.1) to motivate our experimental design and analysis, presented in the next section. Let the signal in risk tolerance, or "true" risk preferences, be denoted S, with $\mu_s = E(S)$, and $\sigma_s^2 = Var(S) > 0$. η_{ij} and ν_{ij} are i.i.d. noise terms in measured risk aversion for the MPL and the LS tasks, respectively. $i \in \{1, 2\}$ denotes type, $j \in \{s, n\}$ denotes treatment status, where s represents standard MPL and n represents nudged MPL. Suppose there are two types of individuals - those who are confused by the MPL and those who are not. Type 1, occurring with probability p, are not confused by the MPL and give"high quality" responses which reflect both signal and noise. Without additional intervention, Type 2 individuals, occurring with probability $1-p$, are confused by the MPL and give "low quality" responses that reflect only noise. Note that $1-p$ can be larger than the proportion of multiple switchers because Type 2 individuals could by chance avoid multiple switching. The maintained assumption is that $0 < p < 1$. The presence of the two types leads to a mixture distribution in measured risk aversion.

Scenario I: Low quality decision-making is a stable attribute of the decision-maker

For the control group using the MPL, the response from Type 1 is $X_{1s} = S + \eta_{1s}$, and the response from Type 2 is $X_{2s} = \eta_{2s}$. For the LS task, the response from Type 1 is $Y_{1s} = S + \nu_{1s}$. Because under this scenario confused individuals consistently make low quality decisions, we expect the response on the LS task for Type 2 to be $Y_{2s} = \nu_{2s}$.

Since the nudge treatment cannot induce confused individuals to make "high quality" decisions, the responses of the treatment group will be equal in distribution to those in the control group. That is to say, $X_{1n} \overset{d}{=} X_{1s}$; $X_{2n} \overset{d}{=} X_{2s}$; $Y_{1n} \overset{d}{=} Y_{1s}$, and $Y_{2n} \overset{d}{=} Y_{2s}$.

The response in the MPL is denoted by MPL_j, $j \in \{s, n\}$, whose density function is the mixture of the density functions of X_{1j} and X_{2j}, where p is the weight placed on the density function of X_{1j}. Similarly, the response in the LS task is denoted by LS_j, whose density function is the mixture of the density functions of Y_{1j} and Y_{2j}, where p is the weight placed on the density function of Y_{1j}.

Because of the equality in distribution of the component distribution functions, $MPL_s \overset{d}{=} MPL_n$, and $LS_s \overset{d}{=} LS_n$. This implies that $Cov(MPL_s, LS_s) = Cov(MPL_n, LS_n)$,. Similarly, $Var(MPL_s) = Var(MPL_n)$ and $Var(LS_s) = Var(LS_n)$, implying $Corr(MPL_s, LS_s) = Corr(MPL_n, LS_n)$.

Scenario II: Task specific miscomprehension

Under this scenario, the confusion is specific to the MPL task. For the control group using the MPL, the response is identical to Scenario I: The response from Type 1 is $X_{1s} = S + \eta_{1s}$, and the response from Type 2 is $X_{2s} = \eta_{2s}$.

For the control group using the LS task, because the confusion is specific to the MPL, the responses from Type 1 and Type 2 are $Y_{1s} = S + \nu_{1s}$ and $Y_{2s} = S + \nu_{2s}$, respectively, where $\nu_{1s} \overset{d}{=} \nu_{2s}$.

For the treatment group using the nudged MPL, the response from Type 1 is $X_{1n} = S + \eta_{1n}$, where $\eta_{1n} \overset{d}{=} \eta_{1s}$. If the treatment fully "unconfuses" Type 2s, then the response from Type 2 is $X_{2n} = S + \eta_{2n}$, where $\eta_{2n} \overset{d}{=} \eta_{1n}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status, for both types of individuals. That is to say, $Y_{1n} = S + \nu_{1n}$ and $Y_{2n} = S + \nu_{2n}$, where $\nu_{1n} \overset{d}{=} \nu_{1s}$ and $\nu_{2n} \overset{d}{=} \nu_{2s}$. Therefore, $LS_s \overset{d}{=} LS_n$.

By the properties of mixture distributions,

$$Cov(MPL_j, LS_j) = \sum_i [p_i Cov(X_{ij}, Y_{ij}) + p_i(\mu_{Xij} - \mu_{Xj})(\mu_{Yij} - \mu_{Yj})], \tag{1.1}$$

where $\mu_{Xij} = E(X_{ij})$, $\mu_{Yij} = E(Y_{ij})$, $\mu_{Xj} = E(X_j)$, and $\mu_{Yj} = E(Y_j)$.

For the control group, $Cov(MPL_s, LS_s) = pCov(X_{1s}, Y_{1s}) + (1-p)Cov(X_{2s}, Y_{2s}) = p\sigma_s^2$, and for treatment group, $Cov(MPL_n, LS_n) = \sigma_s^2$. This yields the result that $Cov(MPL_s, LS_s) < Cov(MPL_n, LS_n)$.

More generally, we can assume that the nudge treatment "unconfuses" a proper subset of Type 2s. Appendix B.1.1 shows that in that case $Cov(MPL_s, LS_s) = p_1\sigma_s^2$ and $Cov(MPL_n, LS_n) = (p_1 + p_2)\sigma_s^2$ where p_1 is the proportion who are not confused in the control group and p_2 is the additional proportion who have become unconfused by the nudge treatment in the treatment group. We will have the same result that $Cov(MPL_s, LS_s) < Cov(MPL_n, LS_n)$ in this general case, if $p_2 > 0$. Scenario II also implies that for both the control and treatment groups, the lower the proportion of confused individuals, the higher will be the covariance between the responses on the MPL and the LS tasks.

Because this scenario does not produce clear predictions on the relative sizes of $Var(MPL_s)$ and $Var(MPL_n)$, it does not produce clear predictions on the relative sizes of $Corr(MPL_n, LS_n)$ and $Corr(MPL_s, LS_s)$.

Scenario III: Carelessness

This scenario assumes that the confusion of Type 2 individuals is due to non-task specific carelessness, affecting both the MPL and the LS tasks for the control group. The nudge treatment, which is only applied to the MPL, removes confusion only in the MPL task.

Identical to Scenario I and II, for the control group using the MPL, the response from Type 1 is $X_{1s} = S + \eta_{1s}$, and the response from Type 2 is $X_{2s} = \eta_{2s}$.

For the control group using the LS task, the response from Type 1 is $Y_{1s} = S + \nu_{1s}$, and the response from Type 2 is $Y_{2s} = \nu_{2s}$. Because carelessness is not task specific, the responses on the LS task for type 2 individuals also only capture noise.

For the treatment group using the nudged MPL, the response from Type 1 is $X_{1n} = S + \eta_{1n}$, where $\eta_{1n} \overset{d}{=} \eta_{1s}$. If the treatment fully "unconfuses" Type 2s, then the response from Type 2 is $X_{2n} = S + \eta_{2n}$, where $\eta_{1n} \overset{d}{=} \eta_{2n}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status. For the treatment group using the LS task, the response from Type 1 is $Y_{1n} = S + \nu_{1n}$ and the response from Type 2 is $Y_{2n} = \nu_{2n}$, where $\nu_{1s} \overset{d}{=} \nu_{1n}$ and $\nu_{2s} \overset{d}{=} \nu_{2n}$, so that $LS_s \overset{d}{=} LS_n$.

Using the fact that $\mu = p\mu_1 + (1-p)\mu_2$, Equation 1.1 simplifies to

$$\begin{aligned} Cov(MPL_j, LS_j) = {} & pCov(X_{1j}, Y_{1j}) + (1-p)Cov(X_{2j,} Y_{2j}) \\ & + p(1-p)(\mu_{X1j} - \mu_{X2j})(\mu_{Y1j} - \mu_{Y2j}). \end{aligned} \tag{1.2}$$

For the control group, $Cov(MPL_s, LS_s) = p\sigma_s^2 + p(1-p)(\mu_{X1s} - \mu_{X2s})(\mu_{Y1s} - \mu_{Y2s})$. For the treatment group, $Cov(MPL_n, LS_n) = p\sigma_s^2$. As long as the difference in the expected value of measured risk tolerance in the control group for Type 1 and Type 2 is not in the opposite direction for the two tasks, then $Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$. We would violate this assumption if, for example, confused individuals without intervention are more risk averse in the MPL task, but less risk averse in the LS task. We do not know of any theory or evidence that predicts this pattern.

The intuition for this result is that unlike in scenario II, there are no gains in the covariances between the two tasks for Type 2 individuals under the nudge treatment because the carelessness of Type 2 individuals is not improved for the LS task. On the other the hand, the relative magnitudes of the expected values of risk tolerance for Type 1 and Type 2 individuals are allowed to have the same pattern for MPL and LS tasks in the control group, which adds to the overall covariance of the control group, but they do not have this "similarity" in the treatment group.

It can be demonstrated that under the more general assumption where the nudge treatment "unconfuses" a proper subset of Type 2s, we have the same result that
$Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$. See Appendix B.1.1 for the proof.

1.3 Experimental Design

1.3.1 Experimental Setting

Subjects were recruited from a rural middle school (7th to 9th grade) in an ethnically diverse region of southwest China. The county in which this middle school is located has been on the register of nationally recognized "poor" counties since the criteria for the designation were established in 1986. According to the provincinal statistical yearbook, in 2014, the county annual average GDP per capita was 11,345 RMB (1650 USD).

From the complete school rosters, we randomly drew students from all regular classes in the middle school.[4] Class size ranges from 51 to 72. Our final sample consists of 193 out of 212 students selected by us, for a response rate of 91%. Non-response is largely due to student absenteeism and the roster not being updated for students who had dropped out of school.

The experiments were conducted in the spring semester of 2015, mainly during the 4pm to 7pm break time on campus.[5] Students completed the experiments one-on-one with our experimenters. At the beginning of the experiments, students were told that they would play two games and only one of them would be chosen randomly to realize their final payment. Students were also asked to fill out a short survey after all experiments are completed to capture basic demographic and socioeconomic status information. Subject payments were handed out after the surveys were completed. Average payout was 6.19 RMB, not including a pencil and eraser as a show up gift. Student test scores were separately obtained from the school administrators.

1.3.2 Balance Tests

Subjects were randomly assigned to either the control or the treatment group.[6] Table 2.1 reports the balance of demographic and socioeconomic status variables between the treatment and the control groups. We found no statistically significant differences between the control and treatment groups in age, size of household, distance to school, mother's educational attainment,

[4]We omitted the honors class students at the request of school administrators.

[5]All students are in school through the evening self-study period (wanzixi) which ends at 10pm. This middle school is a boarding school. Over 85% of students live on campus and are only allowed to return home on weekends. The rest live within a 10 min walk to school.

[6]Because treatment status was assigned prior to the date of the experiment, the final number of subjects in each group was not identical.

mother's occupation, monthly household income, monthly allowance, or students' test scores.

1.3.3 Experimental Design

Both the treatment group and control group are administered two tasks: the MPL task and the LS task. The control group is administered MPL using the standard protocol while the treatment group is administered the MPL with a nudge protocol, explained below. There is no difference in the administration of the LS task for the control and treatment groups. The order of the two tasks was randomized within each group. In Appendix Table A.1 we report balance by treatment status and the order in which the MPL and LS tasks were administered. The results show that there are no statistically significant differences between the four groups defined by treatment status and task order.

1.3.4 MPL instrument and the Nudge Treatment

The MPL task follows the design in Dohmen et al. (2011).[7] Subjects are required to make six choices between a lottery and a certain payout (see Appendix B.2 for the instrument). Option A is a coin flip lottery (risky choice) with 50% chance of paying 10 RMB and 50% chance of paying 0 RMB. Option A does not change across the six choices. Option B is the certain cash payout (safe choice), in increasing increments of 1 RMB, from 1 RMB to 6 RMB. One of the 6 pairs of options is randomly selected from the instrument for each subject after she makes her choices, and the option she chose from the selected pair will be implemented. The instrument is incentive compatible. For example, a participant who values the coin flip lottery (option A) at 3.5 RMB certain payout should choose the lottery for all values of the certain payout below 3.5 RMB and should choose the certain payout when it is above 3.5 RMB. For this subject we should observe three choices of the risky choice before switching to make three safe subjects. Under standard assumptions on preferences, subjects should make at most one switch from the risky choice to the safe choice, nevertheless, in practice we find that many subjects switch back to the risky choice after having made a safe choice.

In the nudge treatment, the subjects are first given the exact same instructions used in the

[7] While Dohmen et al. (2011) asks subjects after making the first switch from the risky choice to the safe choice whether they would also like higher amounts of the safe choice, we do not overtly discourage subjects from multiple switching in either the control or treatment groups.

control group to administer the MPL task. As each subject hands in his or her responses, we say the following: "Have you decided? You can think about your choices again carefully and can change your choices. If you would like, we can explain this game one more time." See Appendix B.2 for the protocol. Those who indicate the need are given the instructions again. The treatment is designed to encourage the subjects to put more cognitive effort into the task, without taking away the ability of subjects to engage in MSB. Indeed, 10% of the treatment group exhibited MSB.

The choice results using the MPL instrument are shown in Table A.4. We report the distribution of subjects choosing each possible number of lottery (risky) options (from 0 to 6) before making their first switch to the safe choice. We also report the range of the implied CRRA coefficient corresponding to each number of lottery options chosen, assuming no multiple switching. In the full sample, about 20% of subjects were multiple switchers, which falls in the range of many previous findings (Charness et al., 2013).

1.3.5 LS Instrument

The format of the LS instrument follows Barr and Genicot (2008). The appeal of this design is its simplicity. Subjects are only allowed to choose one coin flip lottery out of six, with the first lottery offering a certain amount and all other alternatives offering higher expected payoff along with higher variance (see Appendix B.2 for the instrument). A more risk tolerant individual is more likely to choose lotteries with higher expected payoffs and higher variance. All choices are consistent with standard assumptions on preferences.

Table 2.3 reports the simple LS game results for the control and treatment group separately. We report the low and high payoffs for each lottery, the implied range of the CRRA coefficient corresponding to each choice and the percentage of subjects choosing each lottery in each group. The distribution of choices are similar in the two groups. In each group, the lottery chosen with the highest frequency is the third safest lottery and a Mann-Whitney test finds no significant distributional differences between the treatment and control groups in the lottery chosen (p-value = 0.33).

1.4 Empirical Analysis

The last row of Table 2.1 reports the share of subjects who are multiple switchers in the control and treatment groups. The nudge treatment reduces the share of multiple switchers from 31% in the control group to 10% in the treatment group. The p-value of the difference in the multiple switching rate is less than 0.001. The fact that the nudge treatment, which does not limit choice sets or overtly discourage MSB is able to eliminate 66% of MSB suggests that the majority of MSB are mistakes that are corrected upon further reflection rather than the result of deliberate choice.

As demonstrated in Section 2, the relative size of the covariances between the MPL and the LS task in the control and treatment groups will allow us to pin down the explanation for MSB most consistent with the data. Table 1.4 reports the covariance between choices made in the MPL task (number of risky choices) and the LS task (riskiness of lottery chosen), which correspond to the risk tolerance ranking of the implied CRRA coefficients of the choices made in each instrument. Because of the presence of multiple switching, we used three different methods to code the number of risky choices in the MPL task. Method A uses the total number of risky choices (this method was suggested by Holt and Laury (2002). Method B uses the number of risky choices made before the first point at which individuals switch from the risky choice to the safe choice, or the "first switch point" (this method was used in, for example, Harrison and Rutström (2008) and Meier and Sprenger (2013)). Method C uses the average of the decision number preceding the first switch from a risky choice to a safe choice and the decision number preceding the last switch from a risky choice to a safe choice, or the last decision number if the last decision was a risky choice (this method is inspired by the argument in the literature that multiple switching is due to indifference (Andersen et al., 2006)). The last column in Table 1.4 reports the p-value of the difference between the covariance in the control and treatment groups. Because we were unable to find an estimator in the literature for the difference between two population covariances, to do inference, we derived the variance of the difference between two sample covariances in Appendix B.1.2, and propose an estimator which is consistent and asymptotically normal. This method is also used to find the significance levels of the estimated covariances.

To simulate the magnitude of the covariance between the two tasks if subjects did not understand the MPL at all and made their choices randomly, we randomly generated 10,000 bootstrap samples of MPL choices from the empirical distribution of the number of risky choices made in the MPL, separately for the control and treatment groups.[8] We find the covariance of MPL responses in each bootstrap sample with the actual responses in the LS task and report the mean of sample covariance and the standard error of sample covariance in Table 1.4. The last column reports the average p-value of the difference between the covariance in the control and in the treatment groups over the 10,000 observations.

In the treatment group, methods A, B and C produce covariances between the MPL and LS tasks of 1.51, 1.67, and 1.58, respectively, all significantly different from 0 at the 1% level. However, in the control group, the covariance between the responses on these two tasks are small and only method B results in a marginally significant covariance of 0.71. The p-value of the difference in the covariances are 0.01, 0.08, and 0.03, respectively. The randomly generated MPL choices produce a mean covariance close to zero for both the control and treatment groups, which are not statistically different from each other. For the control group, the 95% confidence interval constructed from the bootstrap samples, using normal approximation, is $[-0.753, 0.747]$, which contains 0.300, 0.713, and 0.489, indicating that the covariance of control group responses (using any of the three coding methods) is not statistically significantly different from the mean covariance obtained from randomly generated MPL choices.[9]

Table 1.5 shows the Pearson correlation coefficients between responses in the MPL and LS tasks. Although the conceptual framework does not produce clear predictions on the relative sizes of the variances of the MPL and LS responses in the treatment and control groups, empirically we find that the variances of the MPL responses and the variances of the LS responses are not statistically different from each other in the treatment versus the control groups (see Table 1.6). Therefore, the correlation results should provide a similar pattern as the covariance results. In the treatment group, methods A, B and C produce correlation coefficients between the MPL and LS tasks of 0.424, 0.449, and 0.440, respectively, all significantly different from 0 at the 1% level. However, in the control group, the correlation coefficients between these two

[8]The number of risky choices is coded using method B.

[9]See Appendix Figure A.1 for the empirical distribution of the bootstrap sample covariances.

tasks are small and only method B results in a marginally significant correlation coefficient of 0.185. The p-value of the difference in the correlation coefficients are statistically significant at the 5% level for all three coding methods. Results from the 10,000 randomly generated bootstrap samples also show a similar pattern as the covariance results.

This set of results consistently support Scenario II - task specific miscomprehension of the MPL, which is the only explanation for low decision-making quality from the literature that predicts greater covariance in the treatment group. These results also rule out the interpretation that the lower rate of MSB in the treatment group is due to experimenter demand effects. If that were the case, choices in the nudged MPL would be just as noisy as the choices in the standard MPL, and there should be no difference in the covariance or correlation coefficient between the control and treatment groups.

Results using method C also speaks to the potential for indifference to account for MSB. If we assume MSB is a result of indifference between the lottery and a range of certain payout values defined by the first switch point and last switch point from the risky to the safe choice, then of the three methods, method C, which uses the midpoint of these certain payouts to value the lottery, should give the best approximation to true risk aversion. However, as Table 1.4 and 1.5 show, the covariance and the correlation coefficient between the two tasks are no larger using method C than the other two methods, for either the control or treatment group. Indifference in the true sense also should not respond to the nudge treatment, whereas we find a substantial reduction in MSB.

1.5 Additional Results

1.5.1 Data Quality Metric

Under the interpretation of Scenario II, a higher covariance between the two tasks is indicative of higher decision-making quality (that is to say, a higher proportion of individuals making high-quality decisions). The percent of maximum decision-making quality achieved is the ratio of the actual covariance to the maximum possible covariance: σ_s^2. To approximate σ_s^2 we have two candidates, $Var(MPL_j)$ and $Var(LS_j)$, which are equal to $\sigma_s^2 + \sigma_{\eta j}^2$ and $\sigma_s^2 + \sigma_{\nu j}^2$, respectively. The geometric mean of $Var(MPL_j)$ and $Var(LS_j)$ gives an upperbound estimate on σ_s^2, or maximum decision-making quality. Therefore, the correlation coefficient between the two tasks, $Cov(MPL_j, LS_j)/\sigma_{MPLj}\sigma_{LSj}$, gives a lowerbound estimate of percent of maximum decision-making quality achieved.[10] Interpreting the correlation coefficient as a data quality metric, we can further conclude from Table 1.5 that the nudge treatment increased data quality, more precisely, the proportion of high quality responses, by 143% to 371%. A lowerbound estimate of 42% to 45% of maximum decision-making quality is achieved using the nudge protocol.

1.5.2 Is MSB a Good Proxy for Data Quality?

Although the literature generally views MSB as indicative of low decision-making quality, potential low decision-making quality among non-multiple switchers is not well understood. Here we can separately identify decision-making quality and MSB. We explicitly test decision-making quality for multiple switchers and non-multiple switchers in Table 1.7. The results show that data quality is insignificantly different from 0 for both multiple switchers and non-multiple switchers in the control group, and is only significantly different from 0 for the non-multiple switchers in the treatment group. Data quality under the nudge treatment is significantly higher than the data quality of both multiple and non-multiple switchers under the standard protocol (p-values = 0.016 and 0.019, respectively). This implies that lack of MSB does not necessarily indicate high quality decision-making and the common practice of restricting the data to non-multiple switchers does not necessarily resolve data quality issues (Charness et al., 2013).

1.5.3 Cognitive Ability, MSB, and Data Quality

The previous literature shows that MSB is related to cognitive ability, in particular math ability (Dave et al., 2010; Meier and Sprenger, 2013). In the following analysis we first check whether subjects' multiple switching behaviors are correlated with their school test scores and then check if a relationship exists between test scores and data quality. The tests are uniform across the middle schools in the county, and the test results are provided to us by the school

[10] An alternative method is to use the smaller of $Var(MPL_j)$ and $Var(LS_j)$ as an upperbound estimate of σ_s^2, which, according to Table 1.6, is $Var(LS_j)$. In this case, our data quality metric would be $Cov(MPL_j, LS_j)/Var(LS_j)$, or the slope coefficient in a regression of MPL_j on LS_j.

administrators.

Table 1.8 reports the results from the linear regression of multiple switching on test scores for the control and treatment groups separately. Test scores are the average of standardized math and standardized verbal test scores, standardized within each grade. Column 1 shows that, in line with our expectations, individuals' cognitive ability is strongly correlated with MSB in the standard MPL protocol. Column 2 adds a set of control variables: gender, monthly household income, mother's educational attainment, mother's occupation, the number of children in the household, and grade fixed effects. The results are essentially unchanged. Column 2 shows that an increase of one standard deviation in test scores is associated with a 14.9 percentage point decrease in the likelihood of multiple switching. This corresponds to a 49% (0.112/0.307) reduction in the probability of multiple switching. Appendix Table A.2 shows that both math and low verbal scores are significant predictors of MSB. These findings suggest that one reason that multiple switchers using the standard MPL protocol make worse financial decisions (Jacobson and Petrie, 2009) could be their lower cognitive ability, which leads to both MSB and poor financial decision-making.

Columns 3 and 4 in Table 1.8 show that cognitive ability no longer predicts MSB when using the nudged MPL protocol. Appendix Table A.3 Table shows that this is also the case for both verbal and math test scores.

Because MSB is not a good proxy for data quality, to examine the relationship between decision-making quality and cognitive ability, Table 1.9 reports the correlation between the responses on the LS and MPL tasks by overall cognitive ability, for the control and treatment groups. The results show that data quality is low (insignificantly different from 0) for both high and low cognitive ability individuals using the standard MPL. For the treatment group, data quality is significantly different from 0 for both high and low cognitive ability individuals. The point estimates indicate that low cognitive ability individuals exhibited higher quality decision-making with the nudge protocol than high cognitive ability individuals did using the standard protocol. Table 1.9 also shows that cognitive ability is not a significant determinant of data quality, for either the control or treatment group.[11]

[11] Using the regression coefficient in a regression of MPL responses on LS responses as an alternative data quality metric, we found that in a regression of MPL responses on LS responses interacted with test scores, the interaction term was insignificant for both the control and treatment groups (control group p-value = 0.789;

1.6 Conclusion

In this study we developed a conceptual framework defining decision-making quality in the MPL to test several prominent explanations of low decision-making quality suggested by the findings in the literature using a novel experimental design and treatment protocol. In a departure from previous literature, our study provides a direct test of and finds evidence in support of task specific miscomprehension as the explanation for low quality decision-making in the MPL.

Our framework further lead us to propose a novel metric to quantify data quality separately from MSB and we showed that MSB is not a good proxy for data quality, as is often believed. Using this metric we show that the nudge treatment conservatively increased high quality responses by 143%. Data quality improvement is not limited to the multiple switchers. Data quality under the nudge protocol is significantly higher than data quality of both multiple and non-multiple switchers under the standard protocol.

We find that cognitive ability explains MSB in the standard protocol, but it is not a significant determinant of data quality. To the extent that the nudge treatment was able to elicit more cognitive effort from respondents, the findings suggest that cognitive ability is not a limiting factor in achieving high data quality, but cognitive effort is. This does not preclude cognitive ability from influencing decision-making quality in other areas of life, or in other instruments. Our findings speak narrowly to the MPL instrument, and imply that protocol design innovations in the MPL can reveal risk preferences that would otherwise be obscured due to poor decision-making quality.

These findings may be particularly relevant for researchers who would like to employ the MPL but are concerned that the subject pools they study will be prone to MSB. More broadly, our findings point to the importance of investing in efforts to increase subject comprehension and data quality using the MPL, which resonate with the conclusions in Dave et al. (2010) and Charness and Viceisza (2016). One strategy can be to use a nudge protocol similar to ours in conjunction with any existing MPL protocol. Other strategies include using the framing device in Andreoni and Sprenger (2011), discussed previously, reading instructions out loud in

treatment group p-value = 0.836), confirming the finding that cognitive ability is not a significant determinant of data quality.

addition to providing written instructions (Bruner, 2011), and using a visual representation of the MPL (Bauermeister and Musshoff, 2016).

Figure 1.1: Conceptual Framework

Mixture Distribution Weight	I - Stable attribute		II - Task specific miscomprehension		III - Carelessness	
	MPL - standard	LS	MPL - standard	LS	MPL - standard	LS
p	$X_{1s} = S + \eta_{1s}$	$Y_{1s} = S + \nu_{1s}$	$X_{1s} = S + \eta_{1s}$	$Y_{1s} = S + \nu_{1s}$	$X_{1s} = S + \eta_{1s}$	$Y_{1s} = S + \nu_{1s}$
1-p	$X_{2s} = \eta_{2s}$	$Y_{2s} = \nu_{2s}$	$X_{2s} = \eta_{2s}$	$Y_{2s} = S + \nu_{2s}$	$X_{2s} = \eta_{2s}$	$Y_{2s} = \nu_{2s}$
	MPL - nudged	LS	MPL - nudged	LS	MPL - nudged	LS
p	$X_{1n} = S + \eta_{1n}$	$Y_{1n} = S + \nu_{1n}$	$X_{1n} = S + \eta_{1n}$	$Y_{1n} = S + \nu_{1n}$	$X_{1n} = S + \eta_{1n}$	$Y_{1n} = S + \nu_{1n}$
1-p	$X_{2n} = \eta_{2n}$	$Y_{2n} = \nu_{2n}$	$X_{2n} = S + \eta_{2n}$	$Y_{2n} = S + \nu_{2n}$	$X_{2n} = S + \eta_{2n}$	$Y_{2n} = \nu_{2n}$
Result	$Cov(MPL_s, LS_s) = Cov(MPL_n, LS_n)$; $Corr(MPL_s, LS_s) = Corr(MPL_n, LS_n)$		$Cov(MPL_s, LS_s) < Cov(MPL_n, LS_n)$		$Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$	

Table 1.1: Balance Check and MSB

	Control (1)	Treatment (2)	p-value for H0 (1)=(2)
Female	.51	.61	0.192
	(.5)	(.49)	
Age	14.42	14.34	0.671
	(1.2)	(1)	
Number of children in the household	2.09	2.26	0.197
	(.87)	(.92)	
Number of family members in the household	5.34	5.72	0.290
	(2.05)	(2.8)	
Distance from home to school (=1 if less than or equal to 30min walk)	.38	.43	0.410
	(.49)	(.5)	
Mother's educational attainment (=1 if less than or equal to primary)	.68	.68	0.981
	(.47)	(.47)	
Mother's occupation (=1 if agricultural)	.71	.77	0.354
	(.45)	(.42)	
Monthly hh income (=1 if less than or equal to 750RMB)	.45	.45	0.999
	(.5)	(.5)	
Monthly allowance (=1 if less than or equal to 300RMB)	.82	.78	0.497
	(.38)	(.41)	
Test score	.08	-.09	0.177
	(.82)	(.92)	
Multiple switcher	.31	.1	0.000
	(.46)	(.3)	
Observations	101	92	

Notes: Means and standard deviations are presented. Standard deviations in parentheses. Exchange rate: 1RMB = 0.16 US dollars.

Table 1.2: Distribution of Responses in the Multiple Price List

Number of risky choices	Implied CRRA coefficient range	% of subjects
0	$r > 0.69$	17.10%
1	$0.57 < r < 0.69$	13.99%
2	$0.42 < r < 0.57$	18.13%
3	$0.24 < r < 0.42$	7.77%
4	$0 < r < 0.24$	8.81%
5	$-0.36 < r < 0$	7.25%
6	$r < -0.36$	26.94%

Notes: $N = 193$. The number of risky choices is coded using the number of risky choices made before the first point at which individuals switch from the risky choice to the safe choice. The implied CRRA coefficient range is calculated as the range of r in the function $u = x^{1-r}/(1-r)$ for which the subject makes the corresponding number of risky choices, assuming no multiple switching.

Table 1.3: Distribution of Responses in the Lottery Selection Task by Treatment

Lottery	Low payoff	High payoff	Implied CRRA coefficient range	% of subjects	
				Control	Treatment
1	3	3	$r > 4.17$	15.84%	13.04%
2	2.5	5	$0.99 < r < 4.17$	11.88%	9.78%
3	2	6	$0.81 < r < 0.99$	35.64%	29.35%
4	1.5	7.5	$0.32 < r < 0.81$	6.93%	17.39%
5	0.5	9	$0 < r < 0.32$	7.92%	8.70%
6	0	10	$r < 0$	21.74%	21.78%

Notes: $N = 101$ for the control group; $N = 92$ for the treatment group. The implied CRRA coefficient range is calculated as the range of r in the function $u = x^{1-r}/(1-r)$ for which the subject chooses each lottery.

Table 1.4: Covariance between Responses on the MPL and LS Tasks

	Control (1)	Treatment (2)	P-values of H0: (1) = (2)
Method A: Number of risky choices			
Covariance with LS task	0.300	1.514***	0.014
Method B: First switch point			
Covariance with LS task	0.713*	1.670***	0.076
Method C: Average switch point			
Covariance with LS task	0.489	1.580***	0.028
Randomly generated MPL choices			
Mean of Covariance with LS task	-0.0032	-0.0006	0.4941
Standard Error of Covariance with LS task	0.3825	0.3928	
N	101	92	

Notes: Method A defines MPL response as the total number of risky choices; Method B defines MPL response as the number of risky choices made before the "first switch point." Method C defines MPL response as the average switch point when the subject exhibits MSB. Randomly generated MPL choices uses 10,000 bootstrap samples of MPL choices from the empirical distribution of the number of risky choices made in the MPL (coded using method B), separately for the control and treatment groups. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.5: Correlation between Responses on the MPL and LS Tasks

	Control (1)	Treatment (2)	P-values of H0: (1) = (2)
Method A: Number of risky choices			
Correlation with LS task	0.090	0.424***	0.014
Method B: First switch point			
Correlation with LS task	0.185*	0.449***	0.043
Method C: Average switch point			
Correlation with LS task	0.135	0.440***	0.021
Randomly generated MPL choices			
Mean of Correlation with LS task	-0.0009	-0.0001	0.5000
Standard Error of Correlation with LS task	0.0995	0.1061	
N	101	92	

Notes: Method A defines MPL response as the total number of risky choices; Method B defines MPL response as the number of risky choices made before the "first switch point." Method C defines MPL response as the average switch point when the subject exhibits MSB. Randomly generated MPL choices uses 10,000 bootstrap samples of MPL choices from the empirical distribution of the number of risky choices made in the MPL (coded using method B), separately for the control and treatment groups. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.6: Variance of Responses on the MPL and LS Tasks

	Control (1)	Treatment (2)	P-values of H0:(1) = (2)
Method A: Number of risky choices			
Variance of MPL	3.809	4.692	0.308
Method B: First switch point			
Variance of MPL	5.086	5.101	0.986
Method C: Average switch point			
Variance of MPL	3.974	4.753	0.382
Variance of LS	2.929	2.716	0.715
N	101	92	

Notes: Method A defines MPL response as the total number of risky choices; Method B defines MPL response as the number of risky choices made before the "first switch point." Method C defines MPL response as the average switch point when the subject exhibits MSB.

Table 1.7: MSB and Data Quality - Correlation between MPL and LS Responses

	Multiple Switchers (1)	Non-multiple Switchers (2)	P-values of H0: (1)=(2)
Control	-0.040	0.104	0.520
N	31	70	
Treatment	0.000	0.446***	
N	9	93	

Notes: MPL is coded using Method B, the number of risky choices made before the "first switch point." The p-value for the difference in correlations for the treatment group could not be calculated because the N for multiple switchers is too small. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.8: Cognitive Ability and MSB in the MPL

Dep.Var.= 1 if multiple switcher in MPL task

	Control		Treatment	
	Mean = 0.307		Mean = 0.098	
	(1)	(2)	(3)	(4)
Test scores	-0.136**	-0.149***	-0.037	-0.050
	(0.053)	(0.054)	(0.027)	(0.032)
Other controls	No	Yes	No	Yes
N	99	95	92	87

Notes: Students' test scores are the average of standardized math and standardized verbal tets scores within each grade. Robust standard errors are in parentheses. Other controls include: gender, monthly household income, mother's educational attainment, mother's occupation, number of children in the household, and grade fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.9: Cognitive Ability and Data Quality - Correlation between MPL and LS Responses

	Low (1)	High (2)	P-values of H0:(1)=(2)
Control	0.106	0.230	0.533
N	52	49	
Treatment	0.419***	0.477***	0.734
N	46	46	

Notes: MPL is coded using Method B, the number of risky choices made before the "first switch point." High cognitive ability is defined as above having median test scores (average of standardized verbal and math scores) in the control and treatment groups separately. * significant at 10%, ** significant at 5%, *** significant at 1%.

Chapter 2

Holding up Half the Sky? Affirmative Action, Labor Market Restructuring and Gender Inequality in Urban China, 1988–2007

2.1 Introduction

When discussing how public polices can promote gender equality, most existing studies consider the existence of gender gaps as given. Many researchers have suggested a fundamental role for institutional reform in promoting gender equality (King and Mason, 2001; Wong, 2012; Niederle et al., 2013; UN et al., 2015); others, however, find little evidence that the affirmative-action and family friendly policies substantially reduce gender inequality (Goldin, 2014; Bertrand et al., 2014; Bagues et al., 2015; Blau and Kahn, 2016).

This paper examines a unique historical context in which gender equality was enforced through a central labor planning arrangement in urban China before the 1990s. Specifically, strict and extreme gender equality regulations were implemented during this period, which resulted in a high female labor participation rate (around 90%) and a relatively small gender wage gap - female to male earnings ratio was 88%.[1] I am particularly interested in understanding whether women were differentially affected when such strong government intervention was suddenly lifted through reform of state-owned enterprises (SOE) in the late 1990s, and I utilize detailed household survey data, and industry-specific employment data, employing both difference-in-differences (DID) and instrumental variable (IV) strategies to identify the causal effect of this reform.

The unique features of the period between 1950 and 1990, along with the subsequent SOE reform, make China a compelling natural laboratory in which to study the pervasiveness of gender inequality. The principle of gender equality was not only enforced ideologically but also legally to create a socialist society before the 1990s (Booth et al., 2016). For example, Mao's central government cultivated an ideology of "Women Can Hold Up Half the Sky" through media and school education since the 1950s.[2] Moreover, the central planning system forcefully ensured every urban resident full-time and lifetime employment in the SOEs.[3] Upon graduation, every urban adult was assigned a job with a fixed income under the principle of "equal job with equal pay"; with very few exceptions, quitting or moving between firms was disallowed. Most SOEs were responsible for individuals' job arrangements, as well as the welfare of the whole family, including housing, health care, child care, and sometimes even education. The fact that most firms failed to make a profit pushed the government to reform most SOEs in the late 1990s.

This late-1990s SOE reform marked the end of an era in which government intervened in almost every aspects of people's lives in the urban areas.[4] In 1997, the central government announced a policy to privatize, merge or close most SOEs in the urban areas.[5] As a result, in

[1]Author's calculation by using China Household Income Survey 1988 and 1995. Current literature finds gender earnings gap is between 7% and 14% in 1980s, which is smaller than most OECD countries.(Kidd and Meng, 2001)

[2]Figure A.2 is an example of the propaganda of gender equality ideology in the most important newspaper "China Daily".

[3]Generally speaking, there are two types of firms in urban areas during that time: state-owned firms and collective-owned firms. The ownership of state-owned firms was the central government, while that of the collective-owned firms was the local government or local community. In this paper, I do not distinguish between these two types of firms; I define both state- and collective- owned firms as one group and refer to them as SOE.

[4]China's economic reform started in 1979 in rural areas. Four Special Economic Zones were set up in the urban areas in 1980 (all are located in the coastal province, Guangdong). The central planning system was still working in most parts of urban China until the mid-1990s.

[5]The policy is "grasping the big, letting go of the small" (*zhua da fang xiao*).

a short period following the policy's introduction, over 35 million workers were laid off and by 2007, about 80% of enterprises were privatized (Meng, 2000; Smyth et al., 2001; Solinger, 2002; Wu and Xie, 2003; Hsieh and Song, 2015).[6] The massive laying off period lasted about six years, from 1996 to 2001, with the intensity varying across industries and regions. After 1997, market power played a substantially more important role in allocating labor, and government intervention geared toward gender equality was ended essentially. Figure 2.1 compares the change of gender gaps in employment and earnings between China and U.S..China started with an extremely low gender inequality in the labor market outcomes because of the specifical "affirmative action policies" and then it has been dramatically increasing across years, which shows a reversed pattern comparing with most other developed countries, such as U.S.

Although a large number of studies have documented the increased gender gaps in employment and earnings in the process of labor market restructuring, very few have established a causal link between SOE reform and gender gaps.[7] Moreover, none has investigated whether women are affected differently, and if so, why. In this paper, I first extend existing literature to document gender gaps in employment, monthly earnings, and choices of occupations/industries to form a more complete picture of women's labor market performance in pre- and post- reform period.[8]

One important contribution of my descriptive analysis is that I not only provides the traditional gender gap in earnings level but also shows the gender gap in earnings rank. For any given quantile, the level difference measures the difference in earnings between a female and a male at the same quantile of their respective earnings distributions, while the rank gap asks how far below the quantile in her gender's distribution a female's earnings would rank in the male distribution. Taken together, these two measures give a more complete picture of female relative earnings than does either alone.

My main analysis exploits the variation in regional intensity of the SOE reform to identify

[6] The term *xiagang* (step down from the post) was used instead of "laying off" in China to describe someone being forced to leave his working unit, because it was politically sensitive to say someone was laid off in a socialist society. Sometimes, firms forced workers to retire early, so early retirement is another form of laying off during that period.

[7] Literature finds that the gender earnings gap is about 9% before the reform and 14%-20% after the reform. These papers are summarized in the literature review section.

[8] Most previous studies investigate labor force participation and wages as outcomes (Gustafsson and Li, 2000; Shu and Bian, 2003; Whalley and Xing, 2014; Meng, 2012).

the causal effect of this reform on gender inequality in labor market outcomes. In order to proxy the intensity of the SOE reform, I take advantage of the fact that it is characterized by the regional variation of a massive laying off of workers, and then I calculate the local change of employment share in the state-owned sectors as a measurement of reform intensity. I first employ DID to compare individuals in relatively high reform intensity areas to those individuals exposed to low laying off intensity, both before and after the reform.

Although the change of employment share in the state-owned sectors captures the key features of SOE reform, it may suffer from endogeneity. For instance, the outcome of employment and the local change of SOE employment share are simultaneously determined. Moreover, if private firms are more likely to enter the markets where the male workers were more productive, the firm entry could correlate with the change of SOE employment share and the gender gap in employment.

To address these endogeneity concerns, I implement an IV strategy, predicting the laying off intensity with a measure of pre-reform industry composition, augmented by the national industry-specific change of SOE employment share. This Bartik shift-share instrument has been widely used to study labor market issues in developed countries' context (Bartik, 1991; David et al., 2013; Basso and Peri, 2015) and to study regional growth questions in China (Luo and Xing, 2015; Dong, 2016; Ha et al., 2016); my study, however, is the first to compile a detailed pre-reform industrial employment data set to examine gender inequality in China. The intuition behind this instrument is straightforward. The SOE reform was targeting all state-owned sectors in the urban areas, but the pre-determined differential importance of each industry in the economy generated the regional variation of laying off intensity. For example, if some areas specialized in mining in the pre-reform period and the state-owned mining sector experienced a large decline in the employment nationally, I would expect those areas to have a high laying off intensity.

Overall, I find that the SOE reform causes an increase in gender inequality in the labor market. My DID estimate results suggest that the SOE reform can explain the 13.3%- 24.5% increase in the employment gender gap, and the 33% - 36.4% increase in the monthly earnings gender gap. My IV estimation produces similar results. A one standard deviation (20%) increase in the reform intensity causes the gender gap in employment to increase by 6.9 percentage points, and the gender monthly earnings gap to increase by 8.4%. A simple back of the envelope calculation suggests that over 50% increased gender gaps can be explained by the SOE reform. Another interesting finding is that the employment result is almost entirely driven by the relatively old groups (between the ages of 40 and 54); however, the increased gender monthly earnings gap is detected among the young groups (between 30 to 40). Younger and older groups are not affected by the SOE reform in monthly earnings. Moreover, men are not affected by the SOE reform in all outcomes: employment and monthly earnings.

I also find increased household income cannot explain the increased gender employment gap, which suggests women are not voluntarily leaving the labor market because of the specialization in the household work. Furthermore, the effect is almost driven by those low educated women, which may indicate the decreasing demand for low skilled workers in the emerging labor market. Similarly, I find industries, occupations, household income, and the existence of children under age 6 can hardly explain the increased gender earnings gap. Therefore, it is interesting to see if the response to the SOE reform is related with traditional gender norms. Current literature suggests that regional variation in sex ratio is associated with the differential intensity of gender norms in China (Qian, 2008; Alesina et al., 2013). So I use the 1990 census to calculate the sex ratio of the birth cohorts who were under age 10 and I find my employment gap and earnings level gap results are almost entirely driven by the high sex ratio areas. More details about the variation of sex ratio across regions will be discussed in section 7.3.

Although the fact that increased gender employment gap is only detected in those high sex ratio areas is consistent with the theory that women are somehow forced to leave the labor market, I cannot conclude that women are discriminated in pay simply by observing the fact that increased earnings level gap is driven by the high sex ratio areas. The most important reason is that the increased earnings level gap is fundamentally driven by two different forces: the change of wage structure and the change of gender-specific factors (i.e, discrimination).[9] In order to disentangle these two effects, I further study the impact of the SOE reform on gender rank gap. The idea that change of gender rank gap is only driven by gender specific factors

[9]Gender-specific factors describe "male-female differences in skills and on the relative treatment of women by employers" (Blau and Kahn, 1997). Wage structure is defined as "the array of prices set for various labor market skills (measured and unmeasured) and rents received for employment in particular sectors of the economy" (Blau and Kahn, 1996)

originates from the extensive work of wage decomposition (Juhn et al., 1991, 1993; Blau and Kahn, 1996; Bayer and Charles, 2016; Blau and Kahn, 1997; Lemieux, 2006; Bayer and Charles, 2016). The way I construct gender rank gap is discussed in details in section 6.1.2.

Specifically, I compare the impact of the SOE reform on the gains or losses of females in terms of relative positions in the male earnings distribution. Although the point estimate is bigger in areas with stronger gender discrimination culture (i.e. high sex-ratios), the effects are small and not significant in both high and low sex ratio areas. In another paper, I conducted an aggregate level wage decomposition exercise and the results suggest that the dominant factor that drives the increased gender earnings gap is the change in wage structure. So, the importance of traditional gender norms in the contribution of increased gender earnings gap is not obvious.

This paper is organized as follows. Section 2 discusses the related literature and contribution. Section 3 introduces the institutional background on gender equality-related cultures and government policies in the pre-reform period and the SOE reform in the 1990s. In Section 4, I describe a simple conceptual framework and discuss the possible forces that would lead to increased or decreased gender gaps in the labor market outcomes. Section 5 describes the data source. Section 6 presents the empirical analysis, including main methodology and regression results. Section 7 discusses possible mechanisms and shows the suggestive evidence of persistence of traditional gender norms on increased gender employment and the inconclusiveness of the increased gender earnings gap. Section 8 concludes.

2.2 Related Literature

My paper contributes to a large number of literature which studies the issue of gender gaps in urban China. My study has several advantages. First, most of the existing papers do not discuss the causal relationship between SOE reform and the increased gender gaps. For instance, some studies simply compare the gender earnings gap between 1980s and 1990s without discussing any mechanisms (Gustafsson and Li, 2000; Shu and Bian, 2003; Millimet and Wang, 2006); others present some descriptive evidence to suggest there might be a link between SOE reform and increased gender earnings gap without identification strategy (Whalley and

Xing, 2014; Meng, 2012). In addition, several studies use the Oaxaca decomposition method to explore which factors are contributing to the sizable gender earnings gap and conclude that after controlling for a series of individual characteristics and occupations/ownership/industry, there is still a large unexplained part of the gender earnings gap (Bauer et al., 1992; Liu, 2011; Ni et al., 2005; Su and Heshmati, 2011; Shi et al., 2011). Generally speaking, these papers suggest that education plays a more important role in determining workers' wages in the post- than pre-reform period, and ownership of the industry can explain part of the increased gender earnings gap. But less than 50% of the gender earnings gap can be explained by the observed individual characteristics, and most studies agree with the existence of discrimination and/or unobserved productivity differences without providing further evidence (Cai et al., 2008).

The most relevant study is by Jenq (2015). In this paper, the author uses 1990, 2000, and 2005 census data, employing seemingly-unrelated regression (SUR) and OLS regression to study the effect of SOE reform on the aggregate level change of gender gap in employment. The author calculates prefectural change of SOE employment share and change of female employment share and argue that these are exogenous in China's setting. She finds that female industry-biased privatization can explain almost 50% of the increase in the employment gender gap. My study distinguishes from this paper in several aspects. First and foremost, I rely on time and regional intensity variation, employing both DID and IV strategies to establish a reliable causal link between the privatization movement and gender gaps. Jenq's paper does not include any further identification strategies beyond assuming that the change of employment share is exogenous. As I have discussed, the endogeneity concern of this measurement, her assumption without identification strategy, would be biased. Secondly, I am particularly interested in understanding whether females are affected differently due to this SOE reform, not the aggregate level change of gender gap. The increased gender gap could be the result of an increase in male employment or a decrease in female employment; it could also be the result of a decrease in both that negatively affected women to a greater extent. Only looking at the aggregate level change will not convey women's relative economic performance in the labor market. Lastly, I use four waves detailed household survey data to investigate more individual outcomes, including employment, retirement, earnings, and so on; The paper by Jenq mainly uses census data to look at the aggregate employment as outcome.

Bartik instrument has been widely used to study labor market issues in many developed countries' settings (Altonji and Card, 1991; Card, 2001; Basso and Peri, 2015; David et al., 2013), but due to data availability, my study is the first one which compiles a detailed pre-reform industry-specific employment composition data set from a large number of sources to investigate gender related labor market outcomes in China.

My paper is also related to recent studies on the impacts of government intervention on gender inequality and the historical origins of gender roles. Starting from Norway in December 2003, Spain, Iceland, Italy, Finland, France, and the Netherlands have all passed similar reforms requiring some specific percentage (eg, 40% in Norway) of representation of each gender on the board of directors of publicly limited companies. Bertrand et al. (2014) did not find evidence that this policy reduces the gender wage gap or increases female representation in top positions. Furthermore, they find little evidence that the reform affected the major, marital or fertility decisions of young women. On the other hand, Alesina et al. (2013) and Hansen et al. (2015) emphasize the importance of pre-modern agricultural activities in shaping contemporary gender roles and attitudes. My study complements this strand of literature. More importantly, the strict residence registration system, as well as the uniform central labor arrangement across the whole urban area before the reform provides a natural laboratory in which to examine the effectiveness of enforced gender-equality actions. The results from my paper suggest that a labor market intervention more than 40 years long, in addition to the movement of cultivating gender equality ideology cannot fully change people's attitudes toward the appropriate roles of men and women in society.

My paper also speaks to a vast literature on the effects of the reforms that happened in the transitional economies. The path that China followed in transforming from a central- planning, socialist economy to a market economy is similar to the reunification of Germany. Also, the SOE reform belongs to the worldwide privatization movement in the 1990s.[10] Many contributions to the existing literature have discussed the change of gender inequality after the reunification of Germany (Burda and Hunt, 2001; Hunt, 2002; Danthine and Hunt, 1994; Hunt, 2004). They find that although the gender wage gap narrowed after reunification, women were more likely to leave the labor market. My finding is consistent with these descriptive results. One advantage

[10] Refer to Megginson and Netter (2001) for a thorough review.

of my study is that the SOE reform introduces a sharp change to the labor market in China, which provides a good opportunity to identify the causal effect.

2.3 Institutional Background

In this section, I describe the history and change of women's status in Chinese society before and after the Communist Party came to power. I discuss the emergence, implementation and abolishment (SOE reform) of the enforced gender-equality actions in the urban areas. These historical events motivate my identification strategy as well as the explanations of the findings.

Pre-Communist Ruling Period Before the Communist Party came to power in 1949 China, the positions of women in marriage, family, and society had been mostly defined by Confucianism. For example, women were expected to follow the Three Bonds of Obedience: "To obey fathers when young, husbands when married, and adult sons when widowed." This outlook had fundamentally permeated Chinese culture and religion (Johnson, 2009). Hence, women were labeled as submissive, passive, and weak, and the appropriate role for them was to stay at home. Traditional gender norms also resulted in strong preferences for sons. However, the establishment of a Communist government in 1949 followed by a series of social, economic and political experiments under the Marxist ideas to create a socialist society, promoted women's rights and their position in the society (Entwisle and Henderson, 2000).

Communist Ruling Period One unique feature of China from 1950 to 1990 is that women's social status was strongly shaped by the political approach. Many pervasive reforms that were in favor of gender equality took place. The 1950 Marriage Law and the 1954 Constitution abolished polygamy, child betrothal, and interference in the remarriage of widows (Meijer, 1971). For the first time, the 1950 Marriage Law legalized that wife and husband enjoy equal status at home and marriage should be based on the complete willingness of the two parties. Later on, the Anti-Confucianism Cultural Revolution that happened between 1966 and 1976 denied all traditional ideas about women, and the central government used every possible method, including newspaper/TV/radio, school education, and books, to propagate Mao's "Women Can Hold Up Half the Sky" ideology. Besides implementing a new law to target gender equality within the marriage and cultivating the new ideology that women could contribute as much as men to the development of society, starting from 1950s, the Chinese government established a strict central planning system to arrange labor under the ideal of Marxist equality. As a result, the labor force participation rate of women was extremely high in urban areas, and the gender wage gap was kept at a level that was smaller than that of the United States and most OECD countries.[11]

The Marxist equality idea that was realized through the labor arrangement played a central role in redefining women's status in the society (Entwisle and Henderson, 2000). First, the labor arrangement and wage rate were completely centrally determined in the urban areas. One important reason for this system functioning is that the strict residence registration system, known as *hukou*, almost prohibited any migration between rural and urban areas. Until the late 1980s, China's economy was divided into two mutually exclusive parts . Each year, the State Ministry of Labor and Personnel assigned employment and wage quotas to each local government. Eventually, the labor quota would reach the educational institutions and the wage quota would be assigned to each state or collectively-owned firms or government departments. When an individual graduated, he/she would be assigned to a work unit mainly based on his/her educational attainment and political background.[12] No one would be allowed to search for a job themselves and no work unit could choose workers independently (Meng, 2000; Liu et al., 2008). Furthermore, individuals were not allowed to quit or change their jobs except for promotion. This was a life-time employment with an accurate fixed wage. There were 8 wage levels for factory workers and technicians, and 24 levels for administrative and managerial workers, with some variations across regions (Meng, 2000). The goal of the firms was not to maximize profit; Instead, they functioned as many independent small societies. They not only provided workers with employment, but also housing and medical treatment for family members, and child care and education for children. Due to the mandated equal labor attachment and the instituted equal pay for equal jobs for men and women, China has kept most formal institutions that guaranteed gender equality during that period. No doubt, these socialist policies had shrunk

[11]The female labor force participation rate was around 90% among the 19 to 54 age group, and the gender wage gap was, based on the author's calculation, about 12%, compared to about 30% in United States from the existing literature (Blau and Kahn, 2007, 2016).

[12]Generally speaking, political background indicates the length that an individual had been in the Communist Party.

the absolute size of the gender gap and transformed the gender norms to a large extent during this period.(Eichen and Zhang, 1993; Hannum and Xie, 1994; Yang, 1999).

Post-1990s, Period of SOE Reform The economic reform started in rural areas with the fast growing township enterprises, and later, on the set-up of four Special Economic Zones along the southeastern coast of China.[13] By the mid-1990s, the prosperous development of the private firms in the rural areas and the expanding foreign and export-oriented private firms in the coastal Economic Zones aggravated the failure of the SOE in most urban areas. Lacking a mechanism to incentive workers and the autonomy of market pricing, the main reason for SOE' barely surviving prior to market reform was the monopoly power created by the political intervention. Once the political intervention evaded, it was almost impossible for them to compete with other private firms (Lin et al., 1998; Lin and Tan, 1999; Perkins, 1994).

In the middle of the 1990s, about half of these SOE were experiencing losses, and the number of redundant workers was estimated to reach as high as 20%-30% of total workers (Xianguo, 2007). SOE reform was politically sensitive because life-time employment and equal pay with equal jobs were regarded as two key characteristics of socialist society. The central government did not endorse the SOE reform until the 15th Communist Party Congress in September 1997 (Frazier, 2006). The *zhuada fangxiao* ("grasping the big, enlivening the small") policy was announced at this Congress. The key component of the reform is to keep only a few large strategic sectors under the state ownership and merge, privatize or close most other medium-to-small firms. As a result, over 35 million workers were laid-off (Smyth et al., 2001; Chao, 2000; Zeqi and Yongnian, 1998).

The privatization movement started in 1996, featuring the appearance of early retirement. Massive laying off happened in 1997 and lasted about five years.[14] The data from the National Bureau of Statistics suggests that the employment of SOE peaked at about 109.5 million in 1995 before falling to 69.2 million at the end of 2002, a 36.8% decline (Yearbook, 1998, 2003). The most affected sectors were manufacturing, mining, and utilities, which fired 65% employees; the total number of employees in these sectors dropped from 44 million in 1995 to 15.5 in 2002

[13]Township enterprises are another form of collective-owned enterpreise, but the ownership belongs to farmers in the rural areas.

[14]There is another term to describe lay offs during that period in China: xiagang, which means that workers were forced to leave, but still had to maintain ties with their enterprise.

(Yearbook, 1998, 2003). Another stunning decrease is in the number of firms. The total number of industrial state-owned enterprises declined precipitously by 54.7%, from 110,000 in 1997 to 53,489 by late 2000 (Yearbook, 1998, 2003). The urban collective firms, which were owned by the local government, were also in the scope of the SOE reform. The shrinking of the collective firms shares a similar pattern with the state-owned enterprises. For example, the number of SOE industrial workers fell from 14.9 million in 1995 to 3.8 million in 2002 (Yearbook, 1998, 2003).

To summarize, within 5-6 years, the central planning labor arrangement was abolished. After the SOE reform, all firms worked toward the goal of profit maximization and were free to hire or fire workers from the growing labor market. New entrant workers no longer enjoyed the non-contract life-time employment, and their wages were determined by the market forces. Although the new SOE still have some monopolistic power in some specific sectors, they do not bear any other social responsibilities as before (Lee, 2000; Solinger, 2002). In Figure 2.1 and Figure A.2, I compare the change of gender gaps in employment, early retirement and earnings between 1988 and 2007 in China and U.S.. Overall, the gender gaps in employment and earnings have been decreasing across years in U.S., which has been extensively discussed in the literature. Studies also show the similar trend in many other developed countries. The gender gap in early retirement among the age 40 and 54 groups keeps constantly close to zero across years, which indicates few workers retire so early in U.S. In contrast, China shows a completely reversed trend in the change of gender gaps in these labor market outcomes. Gender gaps in all three interested outcomes started with very small numbers, and then increased dramatically at the year around 1995. For example, in 1988, the gender employment gap is about 20% in U.S. but only 0.05% in China, however, in 2007, these two countries converged to about 17%. Similarly, gender gaps in early retirement and earnings have also more than doubled in China after the SOE reform.

The SOE reform symbolizes the end of a special era when women were vigorously protected by the government in the labor market. Although laid-off workers were entitled to receive living allowances and unemployment benefits from the government to maintain a minimum living standard, current studies suggest that only about 34% of individuals experiencing job separations between January 1996 and November 2001 were employed again within 12 months of leaving their jobs (Cai et al., 2008). More importantly, the fact that the unpracticalness of the labor law and the absence of anti-discrimination law potentially disproportionately disadvantage women in the labor market given the possibility of resurgence of gender discrimination culture (Cooney, 2006; Lee, 2007; Yao and Xie, 2004). This would potentially counterbalance any contribution in decreasing the gender pay gap from the improvement in women's education attainment.

2.4 Conceptual Framework

Given the fact that China started with such low gender inequality in the labor market outcomes, it is not surprising to expect changes in the gender gaps. Therefore, the question is how and why.

First of all, SOE reform indicates that the overall wage structure transformed from a highly centralized form to a decentralized status. Existing studies have documented the importance of wage structure in the determination of gender inequality in the labor market outcomes. For example, Blau and Kahn (1996, 2000, 2003) find that more egalitarian pay structure systems are associated with lower gender pay gaps across over 20 countries between 1980s and 1990s. As a result, I would expect that the gender pay gap will increase because of the decentralization of the wage structure.

On the other hand, SOE reform is a shock to the gender-specific factors. For instance, SOE reform implies that personal preferences are no longer suppressed by the political power. Personal preferences are largely shaped by the traditional cultures, and as I have discussed in Section 3, before this reform, such preferences were strongly suppressed by the direct government intervention in employment and wage setting so employers did not have any freedom to express their willingness or preferences. In the post-reform period, employers may be discriminatory against females in recruiting and setting wages because of their personal preferences which are shaped by historical gender discrimination culture. Under this circumstance, I would expect that SOE reform leads to increased gender gaps in employment and wages and one of the important reasons is the resurgence of gender discrimination culture. Additionally, it is also possible that female and male workers have different personal preferences and this leads to the occupation or industry segregation. For example, extensive studies have suggested that women are less competitive or less risk averse than men and they may choose jobs with less income risk and lower mean wage rates (Gneezy et al., 2003; Dohmen et al., 2015). However, studies also suggest the influence of social norms on such preferences, therefore, even if women voluntarily choose to work in lower paying jobs or choose to leave the labor market, it is still a reflection of traditional gender norms (Gneezy et al., 2009; Zhang, 2013).

To sum up, the relationship between the SOE reform and the change of gender gaps in the labor market outcomes in such transitional economy is complex as various forces play together and whether traditional gender norms contribute to the change of gender gaps is a question that can only be answered by empirical analysis.

2.5 Data

This section provides detailed information about the dataset I have used and the construction of the intensity of the SOE reform.

2.5.1 China Household Income Project

The data used in this study comes from the survey of the China Household Income Project: 1988, 1995, 2002, and 2007. This project, including both rural and urban households, was designed by the Economics Institute of the Chinese Academy of Social Sciences(CASS) and a group of international economists. The provinces, number of households, and individuals covered by the project vary across years. In 1988, the research team surveyed 9,009 households with 31,827 individuals living in 10 provinces(Eichen and Zhang, 1993). In 1995, the numbers were 6,931 households and 21,698 individuals in 12 provinces (Li et al., 2008). The 2002 survey covered the same provinces as 1995, including 6,835 households and 20,632 individuals (Li et al., 2008). Sixteen provinces were surveyed in 2007, but only 9 provinces with a total of 14,683 individuals and 5,000 households are available for public use (Luo et al., 2013). Since the central labor arrangement was implemented in the urban areas and the SOE reform mainly targeted those firms that were located in the urban areas, I restrict my analysis to individuals who had urban *hukou* (Meng, 2000; Groves et al., 1995; Perkins, 1994). Also, the mandatory retirement age during that time period was 60 for men and 55 for women, so I focus my study on individuals between the ages of 19 and 54 (Du and Dong, 2009; Giles et al., 2006a).[15] The final dataset in my empirical analysis includes 24,706 households with 52,947 individuals across 14 provinces and 80 prefectures.

CHIP has detailed information about individuals' demography, working status, and income. Individuals were surveyed on their age, ethnicity, educational attainment, employment status, working industries, occupations, monthly earnings and other information related to income.[16] I use the answers to the employment status question to define both employed and retired dummy outcomes. Employed equals 1 if the individual reports he currently has a full time job and 0 otherwise; retired equals 1 if the person says he is retired and 0 otherwise.[17] I define earnings as the sum of regular wages, all kinds of bonuses and subsidies, and other income from the primary job.

Table 2.1 shows the summary statistics of key outcome variables and individual characteristics. Panel A shows the average labor market outcomes before (1988 and 1995) and after (2001 and 2007) the SOE reform. Monthly earnings were adjusted by CPI to the 2014 year. The real average monthly earnings had increased more than six-fold in post- than pre- reform period. This is not surprising, since China's economy was growing at an average of 9.91% per year during that time period. On the other hand, the employment rate declined from over 90% to just about 70% among working-age adults in the sample. Also, more people reported that they were "retired." As I discussed before, early retirement has been used as another method to lay off workers in the SOE restructuring process. Another worth noting feature about the change in the labor market is that very few people were working in the private sectors before the reform (about 2%). The proportion grew to more than 50% after the reform. Panel B compares the individual characteristics before and after the reform. Individuals are older and more educated in the post-reform period. Both age and education level could affect labor market outcomes;

[15] The retirement age for women varied across educational attainment and occupations. The youngest age at which women were legally allowed to retire was 45 (Du and Dong, 2009; Giles et al., 2006b). Later on, I expand my study group to age 60 as a robustness check.

[16] In 1995 and 2002, individuals were also asked their working hours per week.

[17] There are eight answers to the survey question: "What is your current employment status?" They are: (1) employed (full time job); (2) waiting for job; (3) unable to work; (4) retired; (5) currently a student; (6) pre-school children; (7) full time homemaker; (8) others. In the 2002 survey, there are several categories in addition: (1) officially off-duty (lixiu); (2) laid-off (xiagang); (3) ligang (left post); (4) early retirement; and (5) internal retirement. I group (1), (4) and (5) to "retired"; (2) and (3) as "others."

therefore, it is necessary to include these variables as controls in our analysis.[18]

2.5.2 SOE Reform Intensity Measurement

Since the SOE reform occurred at the national level, I cannot simply compare post-reform outcomes with pre-reform outcomes. The changes in labor market outcomes could be due to multiple reasons, other than SOE reform. Increasing retirement, for example, could be caused simply by an age-demographic shift or by an increase in household income. Therefore, in order to assess the causal effect of SOE reform on gender inequality in the labor market, I employ a difference-in-differences method. I compare outcomes before and after reform for individuals from the more affected areas to the less affected areas. I first define reform intensity using the change of the SOEs employment share in the urban areas. The higher the change in employment share, the higher the reform intensity.

The employment data comes from various official statistical publications and publicly-available databases.

- National and Provincial number of SOE workers in each industry and total number of workers in the urban areas are collected from the China Labor Statistical Yearbook 1996, 1995, and 2002, and Comprehensive Statistical Data and Materials on 50 Years of New China.

- Prefectural number of SOE workers by industry and total number of workers in the urban area are extracted from 14 Provincial Statistical Yearbook 1996 and, 1995 and China City Statistical Yearbook 1996, 1995, and 2002.

Figure 2.2 describes the change in the labor force in SOE sectors between 1988 and 2013. Before 1990, almost all urban workers were working in state-owned sectors. Then, the share gradually decreased. The sharp decrease started in 1997, when the SOE reform policy was officially announced and firms started to lay off workers. The massive laying off lasted about six years, between 1996 and 2001. After that, SOE no longer dominated the economy. In 2013,

[18] Not every wave of the surveys ask the working experience question; I define working experience = age - years of schooling - 6. Later on, in the regression analysis, due to the multicollinearity issue between years of schooling, working experience, and age, I only include age and years of schooling as controls.

fewer than 40% of total workers were working in any SOE sectors. Such change suggests that China was transforming from a central planned economy to a market-oriented economy.

In order to measure the SOE reform intensity, I collect the total number of workers (*Zhi gong*) and the number of workers in the SOE sectors in 80 prefectures (covered by the CHIP data set) for the years 1995 and 2001. Prefectures, which encompass all metropolitan areas in China, are logical geographic units for defining local labor market. I calculate the change of the SOE employment share as:

$$\Delta \text{ SOE Emp share}_p = \frac{L^{SOE}_{p,t_0}}{L_{p,t_0}} - \frac{L^{SOE}_{p,t}}{L_{p,t}}$$

In this expression, L^{SOE}_{p,t_0} ($L^{SOE}_{p,t}$) is the start (end) of period SOE employment in prefecture p and L_{p,t_0} ($L_{p,t}$) is the start (end) period total employment in prefecture p. A positive Δ SOE Emp share$_p$ suggests that the share of workers working in the SOE sectors is decreasing over the years and vice versa. [19]

Figure 2.3 shows the regional variation of the SOE reform intensity. The darker the color, the more workers left the public-owned sectors between 1996 and 2001. The average change of SOE employment share is 0.31 with the standard deviation 0.12.

The major concern of using the change of SOE employment share to measure intensity is that it could be correlated with some unobserved prefectural characteristics which affect the labor market outcomes. For instance, the outcome of employment and the local change of SOE employment share are simultaneously determined. Moreover, if private firms are more likely to enter the markets where the male workers were more productive, the firm entry could correlate with the change of SOE employment share and the gender gap in employment. To overcome this endogeneity issue, I collect the pre-reform number of SOE workers in each industry (at two-digit industry code) at the prefectural level from various provincial statistical yearbooks. By using the pre-reform prefectural industrial composition and national industry-specific shock to the SOE employment caused by the SOE reform, I develop a Bartik intensity index to instrument the prefectural change of the SOE employment share.[20] Due to the data availability and bounded

[19]Out of 293 prefectures, 89 were covered in CHIP. Due to the change of geocode, I am able to identify 80 prefectures.

[20]Bartik instrument was first introduced by Bartik (1994), and used in papers such as David et al. (2013), Card (2009), and Basso and Peri (2015).

by the CHIP surveyed prefectures, I compile an employment data set that covers 37 prefectures across all 14 provinces which have been surveyed in the CHIP. The Bartik intensity index is constructed as follows:

$$\text{Bartik Intensity Index}_p = \sum_{i=1}^{n} \text{SOE Emp share}_{i,p,t_0} \times \Delta \text{ SOE Emp share}_i$$

Where SOE Emp share$_{i,p,t_0}$ is the start of period SOE employment share in industry i and prefecture p. Δ SOE Emp share$_i$ is the aggregate change of SOE employment share in industry i between the start and the end period.

Figure A.3 shows histograms of the distribution of the change of SOE employment share both for all 80 prefectures and the subsample (37 prefectures) with the available pre-reform industry specific number of SOE workers. The mean of the subsample is 0.33 with standard deviation 0.13, which is marginally larger than the total sample mean (0.31 with the standard deviation 0.12).

Table 2.2 shows the summary statistics of the pre-reform share of SOE workers by industry across the 37 prefectures. First, manufacturing plays the most important role in the old central labor arrangement system, and wholesale and retail trade comes next. Second, the standard deviation is relatively small across most industries, which suggests the central government significantly intervened in the economy to build a homogenous market regardless of the local differential endowment. I report the national industry- specific change of SOE employment share in Table 2.3. Manufacturing lost most workers working in the SOE. The SOE also substantially shrank in the mining, construction, real estate and wholesale sectors. Lastly, Table A.4 summarizes both the change of SOE employment share and the Bartik intensity index. I will present the correlation between the change of SOE employment share and Bartik intensity index in the next section when I discuss my main identification strategy.

2.6 Empirical Analysis

2.6.1 Descriptive Analysis

I first document the change of women's economic activities, simply comparing before and after the SOE reform based on linear regressions and quantile regressions. One important contribution of my analysis is that I show an additional dimension of gender wage gap - the rank gap. Taken the gender wage gap in levels together, I provide a more complete picture of the change of gender gaps in the labor market outcomes.

The labor market outcome is specified as: For individual i in prefecture p and year t,

$$Y_{ipt} = \alpha + \beta_1 Female_i + \delta_t + \gamma_p + X'_{ipt} + \varepsilon_{ipt} \tag{2.1}$$

where Y_{ipt} is one of the following outcomes: (1) employed (1 or 0); (2) retired (1 or 0); (3) ln(monthly earnings) (4) work in private sectors (1 or 0). Monthly earnings include regular wage, bonus and all subsidies from current primary job. The price is deflated at the 2014 level. X'_{ipt} is a vector of individual characteristics controls, which include: age, age squared, years of schooling, and ethnicity. δ_t is year fixed effects. γ_p is prefecture fixed effects. Standard error is clustered at the prefecture level. I restrict my analysis to those individuals who report they currently have a full time job if outcome is ln(monthly earnings) or working in private sectors I run Equation 2.2 for before (1988 and 1995) and after (2002 and 2007) periods, separately.

Table 2.4 presents the results in four panels and two columns. Each panel represents one outcome in two columns, separately. Column (1) shows gender gap in the labor market before the reform while column (2) indicates gender gap in the labor market after the reform. The results suggest the gender gap in all interested outcomes significantly increases in the post-reform period. For example, panel A indicates that females are 5.6 percentage points less likely to be employed than males before the reform and 15.2 percentage points less likely to be employed than males after the reform. The gender gap in employment significantly increases. More females than males leave the labor market after the reform. Results in panel B suggests that some females leave the labor market in the form of early retirement. Females are 11.4 percentage points more likely to retire early than males in the post-reform period. As discussed before, many firms use early retirement to force workers to leave the job position during the labor market restructuring process. The results suggest that firms are more likely to employ this method on females. Panel C shows that gender earnings gap increases from 12.1% to 22.5% even after partialling out the effect of education, age, ethnicity, unobserved time-invariant prefectural characteristics and time trend. For each dollar that a working man earns, a working woman can only earn less than 80 cents in the post-reform period. Panel D reflects the dramatic change of the labor market structure. In the central planning system, very few (2%) people worked in the private sectors, and there is no difference between men and women working in different ownerships' firms. However, after the SOE reform, 49% of men and 57% of women work in the private sectors. women are 7.1 percentage points more likely to sort into private sectors, which offer less pay and worse welfare benefits than state-owned sectors during that time period (Meng, 2000).

I also estimate quantile regressions with log earnings and percentile rank as dependent variables, separately. The specification is as follows:

For individual i in prefecture p and year t,

$$Y_{ipt} = \alpha + \beta_1 Female_i + \gamma_p + X'_{ipt} + \varepsilon_{ipt} \tag{2.2}$$

where Y_{ipt} is either ln(monthly earnings) or an individual's percentile rank in the male earnings distribution for a corresponding year and prefecture. Other control variables include three categories of education attainment, three age categories, and prefecture fixed effects. Robust standard errors are clustered at the prefecture level.[21]

I estimate the quantile regressions decade-by-decade and present the results by selected quantiles in Table 2.5 and Table A.5. Overall, the results suggest that the gap has significantly increased in the post-reform period (2002 and 2007) across all quantiles, however, the dynamic process and the magnitude of change varies across years and across quantiles. For example, gender gap changed little between 1988 and 1995, however, women's status in the labor market

[21] Three education groups are:(1) equal to or less than middle school (2) high school and technical school (3) college and above; Three age groups are: (1) age>=19 and age<=30 (2) age>30 and age<=40 (3) age>=40 and age<=54. I use categories instead of linear expression of age and education as controls because this is consistent with later gender rank gap analysis. All of the results presented in the paper are qualitatively robust to using either linear controls of age and education or categories of age and education.

rapidly declined in the process of the labor market restructuring. Between 1988 and 2002, gender earnings differentials increased by 20% - 88%, with the top quantile lost least while the bottom suffered most. Surprisingly, between 2002 and 2007, the deterioration of women's status in the labor market is almost driven by the medium and above percentiles. For instance, the gap increased 20% at the 90th quantile between 1988 and 2002 but it increased more than 70% between 2002 and 2007.

Table A.5 illustrates another dimension of gender earnings differentials - the change of the rank gap. I follow Bayer and Charles (2016) to calculate hypothetical percentile positions for women if their wage distribution would have remained the same as men's wage distribution. The difference between this hypothetical rank and men's true position is defined as gender rank gap.[22]

There are at least two important reasons to present the change of the gender rank gap. First, the change of the rank gap does not usually move in the same direction as the level gap. For example, the 75th quantile females was positioned at 66th quantile in males' wage distribution in 1988 and their position increased to 69th in 1995. However, such improvement completely diminished and was reversed in the post-reform period with the median quantile women lost most. The earnings of the median working females equaled the earnings of the 43th quantile working males in 1988, but their position dropped to 39th in 2002 and 35th in 2007. Second, by studying the change of gender rank gap, I will be able to disentangle the influences of change of wage structure, which is gender-neutral, and other gender specific factors. For instance, the improvement of females' relative positions could be resulted from the decreased gender gap in education attainment or the worsened rank could be because of gender discrimination. By analyzing the change of the gender rank gap provides an opportunity to explore the importance of gender discrimination in contributing to the change of gender earnings gap.

While the results in Table 2.4, Table 2.5, and Table A.5 indicate that gender gaps in labor market outcomes have increased, they are not necessarily caused by the SOE reform. Many different reasons could be contributing to this pattern. Examples include increased household income, working in different industries, and/or occupations. Some other major events which happened between 2002 and 2007 could also result in the increased gender gaps, for instance,

[22]This exercise is done with the same age and education groups.

trade liberalization or massive internal migration. To investigate the casual impact of the SOE reform on gender inequality and the underlying mechanisms of the increased gender gap in the labor market, I next employ both difference-in-differences (DID) strategy to investigate whether and how men and women are affected differently in the labor market restructuring process. Then, I discuss the existence of gender discrimination by providing two pieces of suggestive evidence.

2.6.2 Main Strategy: Difference-in-Differences

My main strategy is the DID method. I exploit two sources variations, geographical intensity variation and the time of SOE reform officially implemented. I use the change of SOE employment share to measure the reform intensity. Because my objective is to study whether women are affected differently, I estimate a generalized DID model, as follows:

For individual i in prefecture p in year t,

$$\begin{aligned} Y_{ipt} = \alpha + \beta_1 Female_i \times After_t \times \Delta EmpShare_p + \beta_2 Female_i \times After_t + \beta_3 Female + \\ + \beta_4 Female_i \times \Delta EmpShare_p + \beta_5 \Delta EmpShare_p \times After_t + \delta_t + \gamma_p + X'_{ipt} + \varepsilon_{ipt} \quad (2.3) \end{aligned}$$

Where Y_{ipt} is one of the three outcomes: (1) employed (1 or 0); or (2) ln(real monthly earnings). $\Delta EmpShare_p$ is calculated by using the formula presented in section 4.2. $Female$ and $After$ are two dummy variables. X'_{ipt} is a vector of individual characteristics, including age, age squared, years of schooling, and ethnicity. δ_t is year fixed effects. γ_p is prefecture fixed effects. Standard errors are clustered at the prefecture level. $\Delta EmpShare_p$ has been standardized to have mean equal to zero and standard deviation equal to one to facilitate interpretation.[23]

β_1 is the main coefficient of interest. It describes the additional effect. In other words, a statistically significant β_1 suggests that the greater the exposure to the SOE reform, the larger the effect on females. Essentially, it captures whether the gender gap would change because

[23]The 2007 wave did not ask the question about Communist party membership; I include this as another control as robustness check in the Appendix, and the results do not change. Also, I try to define "employed" as both full employed and self-employed in the Appendix; the results do not change either. Among individuals who are employed, 2.7% are self-employed. Furthermore, only two waves (1995 and 2002) ask weekly working hours; I add this as another control to study the effect on earnings in the Appendix, the results do not change.

of greater exposure to the privatization movement. Besides, β_2 indicates the overall increased gender gap after the reform, and β_3 shows the gender gap before the reform. β_5 is another coefficient worth noting, since it captures the effect of the SOE reform on males.

One important underlying assumption to validate the DID strategy is that those more affected areas would have had the same trend in the gender gap as less affected areas had there been no SOE reform. In section 7.2, I will provide supporting evidence for this assumption. Another major concern of this strategy is that the change of SOE employment share may be correlated with some unobserved prefectural changes that may affect the outcome variables. In the robustness check section, I provide the IV strategy to overcome such concerns.

2.7 Main Results

The baseline results of OLS analysis are presented in Table 2.7. Columns (1) to (2) provide full sample estimates while columns (3) to (4) show the subsample results as described by Equation 2.3. I find similar results by studying different samples. Overall, I find that SOE reform is associated with women being less likely to be employed and it is associated with an increase in the gender earnings gap.

Specifically, β_2 across the first two columns suggests that gender gaps increase by 10.5 percentage points in employment during the post-reform period. Also, the gender monthly earnings gap increases 11.8% after the reform. The results are similar by using the subsample. The fact that gender gaps increase could be resulting from the improvement of males' labor market outcomes, or deterioration of females' labor market outcomes, or both. The insignificant β_5 across all four columns suggests that males are not affected by the SOE reform. In contrast, the statistically significant β_1 across four columns suggests that women are, disproportionately, negatively affected by the reform. In other words, the greater exposure to the SOE reform, the larger the effect on women. β_1 in column (1) indicates that a one standard deviation increase in the reform intensity (19% increase in the change of SOE employment share) is associated with a decrease in the likelihood of being employed by 1.4 percentage points. The effect is larger by using the subsample, which is shown in column (3). A simple back-to-the-envelope calculation suggests that the SOE reform can explain a 13% - 24.5% increase in the employment gender gap. And columns (2) and (4) show that the SOE reform can explain a 33% - 36.4% increase in the monthly earnings gender gap.

I present these results by controlling for a series of individual characteristics, prefectural fixed effects, time fixed effects and prefectural specific time trends. So any time-invariant unobserved prefectural characteristics, common shocks to all prefectures and prefectural specific time shocks that could affect the outcomes are all considered. These results suggest that women are more likely to leave the labor market than men in the restructuring process. Moreover, for those who are working, women earn much less than men even if they have the same pre-market individual characteristics as men.

2.7.1 Robustness Check

In the previous section, I have presented robust results on gender gaps by controlling for a series of individual characteristics, and labor market characteristics. In this section, I further discuss some other possible confounding events that happened between 1988 and 2007. Then, I will present two pieces of evidence to support the parallel trend assumption.

Falsification Test

The primary concern of using the difference-in-differences strategy is the failure of satisfying the parallel trend assumption, so I conduct a pseudo policy evaluation experiment and present the results in Table 2.14. The idea is to assume the SOE reform happened some time between 1988 and 1995. Hence, I should not find any effects by studying this pseudo SOE reform. The null results from Table 2.14 confirm this assumption and this experiment provides first evidence to support the validity of using DID in the true SOE reform.

As another placebo test, I conduct a permutation test in which I randomly permute treatment variables within the sample. For each permutation, the timing of the SOE reform and the intensity are randomly chosen. Individuals' exposure to different treatment variables are then assigned accordingly.[24]. Figure 2.6 displays the empirical distributions of the placebo treatment effects on three outcome variables from 1,000 permutation tests. The fact that the distribution

[24]Recently, permutation tests have been used in the following papers: (Agarwal et al., 2014), (Bloom et al., 2012) and (Chetty et al., 2011)

is centered at zero is comforting as these placebo tests are expected to find no impacts. In panel A, when I compare the treatment effects that are based on actual exposure, the results indicate that less than 1% of the time permutation estimates are larger than the estimates of actual treatment. In panel B and panel C, the results suggest that none of the 1,000 times of permutation estimates are larger than the estimates of actual treatment. This result based on permutation tests reassures that the effect of SOE reform is statistically significant.

Instrument Variable Strategy

As I discussed before, DID estimation may suffer from an endogeneity issue since change of SOE employment share may be correlated with some unobservable prefectural characteristics which could affect the interested outcomes. Therefore, I use IV strategy to address this concern.

I develop a Bartik intensity index by using the pre-reform prefectural industrial employment composition and national industry-specific shock to the SOE employment to instrument the change of SOE employment share. In this setting, the Bartik intensity index works as a negative labor demand shift.

The IV equation takes the form of the equation represented in Equation 2.3 above, but the variable of interest is replaced by predicted change of SOE employment share:

$$\begin{aligned} Y_{ipt} = \alpha + \beta_1 Female_i \times After_t \times \Delta \widehat{EmpShare}_p + \beta_2 Female_i \times After_t + \beta_3 Female + \\ + \beta_4 Female_i \times \Delta \widehat{EmpShare}_p + \beta_5 \Delta \widehat{EmpShare}_p \times After_t + \delta_t + \gamma_p + X'_{ipt} + \varepsilon_{ipt} \quad (2.4) \end{aligned}$$

The predicted value of $\Delta \widehat{EmpShare}_p$ is generated by the first stage specified as follows:

$$\Delta \widehat{EmpShare}_p = \pi + \pi_1 BartikIntensity_p + \delta_t + X'_{ipt} + \varepsilon_{ipt} \quad (2.5)$$

All other elements in these equations are the same as in Equation 2.3. For this IV approach to be valid, the instrument needs to satisfy the exclusion restriction such that, conditional on the controls in the models, subsequent trends in the gender gap in interested outcomes would not be correlated with the change of SOE employment share, except for a direct effect of the Bartik intensity index. I also need the Bartik intensity index to be a strong predictor of the change of SOE employment share. Figure 2.4 shows a simplified bivariate version of the first stage relationship in the IV approach. It presents a simple scatter plot depicting the relationship between the Bartik intensity index and change of SOE employment share. The pattern is clearly linear, demonstrating that this monotonicity requirement holds. A bivariate regression between these two variables yields a t- statistic of around 4. Given the strength of these relationships, it is not surprising that the data exhibit sufficient power in our first stage regression; the F-statistic on the instrument is 12.15.

Table 2.8 shows the IV results by Equation 2.4. First, column (1) suggests a strong first stage relationship between the change of SOE employment share and the Bartik intensity index. Second, by using the Bartik intensity index as an instrument, I find similar patterns in the increased gender gaps in employment, retirement and monthly earnings as the DID estimation. The magnitude is larger than DID estimation, which suggests DID may underestimate the effect of the SOE reform. One possible reason for the insignificant β_5 across column (2) to column (4) indicates that males are not affected in employment or monthly earnings by this reform. On the contrary, significant β_1 across these two columns suggests that females are substantially negatively affected by SOE reform. A one standard deviation (20%) increase in the reform intensity causes the gender gap in employment to increase by 6.9 percentage points and in the gender monthly earnings gap, to increase by 8.4%. The back-to-the-envelope calculation suggests that over 50% increased gender gaps can be explained by the SOE reform. This holds for all two outcomes: employment and monthly earnings.

In summary, I find that the SOE reform causes women to disproportionately leave the labor market. Moreover, I find that the gender earnings gap increases due to this privatization movement. Both DID and IV estimation have their own strength and weakness. The DID estimation results may be more representative and have more power than IV estimation results because of the larger sample size, and IV estimation tends to be unbiased.

Access to WTO and Migration

One possible confounding event is China's entry into the WTO in 2001. Staring from 2002, trade liberalization not only attracts a large number of foreign companies entering into the

Chinese market, but also results in a dramatic increase of export-driven private firms. However, my measurement of intensity of SOE reform does not capture the effect of trade liberalization because I restrict my calculation of change of employment share to the time period before 2002.

Another potential confounding factor is migration. From 1990 to 2000, more and more individuals from the rural areas of the west part of China migrated to the east to work. But they were not entitled to enjoy any benefits or rights which belonged to urban residents. For example, it was almost impossible for them to work in the SOE. Furthermore, they tend to do part-time jobs, jobs without any contract and they are usually not officially registered. In this paper, my definition of urban workers does not include any workers who are not registered officially, as a result, the existence of migrant workers will not affect my measurement of privatization intensity. To the extent which there may be any measurement error in the data collection process, I drop those prefectures where most migrant workers would go during that time period, and I present my results in Table A.9 in the Appendix. I find similar results as before.[25]. Due to this reform, women are negatively affected in employment and monthly earnings. Also, women are more likely to retire early.

2.8 Mechanisms

2.8.1 Employment

There could be many potential reasons that drive the increased gender employment gap. For instance, childbearing age women may choose to leave the labor market because they need to take care of the children which they did not have to because of the free childcare in the pre-reform period. To explore this possibility, I augment DID approach to examine whether the age path of the estimated impact on gender gaps changes as a function of exposure to the SOE reform.

The augmented regression model takes the following form:

[25] Most of these migrant workers were working in one particular province, *Guangdong*, during that time period.

$$
\begin{aligned}
Y_{igt} = \alpha + \sum_{g=1}^{11} \beta_{1g} Female_i \times After_t \times \Delta EmpShare_p \times g + \sum_{g=1}^{11} \beta_{2g} Female_i \times After_t \times g \\
+ \sum_{g=1}^{11} \beta_{3g} Female_i \times \Delta EmpShare_p \times g + \sum_{g=1}^{11} \beta_{4g} After_t \times \Delta EmpShare_p \times g + \tau_g \\
+ \delta_t + \gamma_p + X'_{ipt} + \varepsilon_{ipt}
\end{aligned}
\tag{2.6}
$$

where g represents 11 age categories between ages 22 and 54. The respective coefficients β_{1g} map out the age pattern in the gender gap in response to the SOE reform. τ_g is age fixed effects. All other variables in the equation are the same as in Equation 2.3.

I present the results in Figure 2.5. For each figure, the X-axis represents 11 age categories. Each point indicates the effect on a specific age group with 90% confidence intervals. In panel (a), I find the effect on employment is almost entirely driven by age between 43 and 54 groups. These women are already beyond the childbearing age. As a result, I can rule out the possibility that the existence of children is the reason for women disproportionately leaving the labor market.

The second important possible reason for women to leave the labor market is the increased household income. And the fact that women have left the labor market and that the gender gap has increased do not necessarily suggest that women are worse off than before. To investigate this possible explanation, I further present DID estimation results by studying different groups. Table 2.9 shows the effects of SOE reform on gender gaps in employment by different income and education groups. β_1 across column (1) and column (2) suggest that there is no significant difference between high income and low income groups in the gender gaps in employment. Furthermore, results from column (3) and column (4) indicate that the increased gender employed gap is almost entirely driven by those low educated groups, which suggest that the demand for low skilled women may have decreased because of the labor market restructuring. To sum up, it is highly unlikely that women choose to leave the labor market themselves and SOE reform has exacerbated the inequality between high skilled and low skilled groups.

2.8.2 Earnings

Panel b in Figure 2.5 suggests that only women between age 30 and 40 are negatively affected in earnings. To explore the potential reasons, column (1) to column (6) in Table 2.10 show the impact of SOE reform on gender gap in monthly earnings among young cohort, adding working industries, occupations, the existence of a child under age 6, increased household income and ownerships as other controls. For working individuals, the sizable gender monthly earnings gap may result from sorting into different industries or occupations. For example, nurses and teachers are regarded as female-dominant occupations which pay less than male-dominant financial sectors. The existence of a child at home may cause women to be less productive at work, which may explain their lower monthly earnings than men. Moreover, private sectors tend to offer lower wages and less welfare benefits than state-owned sectors in China. Thus, if women are more likely to sort into private sectors, their earnings could be lower.[26]

However, I find all these factors can only contribute to a small part of the widening gender gap. From column (1) to column (6), the coefficient of β_1 decreases from -0.045 to -0.030, with adjusted R-squared increases from 0.6192 to 0.6987. A one standard deviation increase in the SOE reform intensity causes the gender gap in monthly earnings to increase by 3.0%, after partialling out the effects of working industries, occupations, ownerships, increased household income and childcare cost. This suggests that the increased gender monthly earnings gap is driven by within ownership, industry and occupation variations. In the next section, I will focus on discussing another possibility of driving the increased gender earnings gap - discrimination.

2.8.3 Influence of Traditional Gender Norms

Current literature proposes three leading reasons to explain the persistent gender earnings gap in the labor market: gender differences in productivity and/or preferences, or labor market discrimination (Bertrand, 2011). It is hard to directly test the labor market discrimination hypothesis, either statistical or tasted-based, using survey data. I will provide some suggestive evidence from two different perspectives to shed light on the mechanism of discrimination.

The special feature about China is that it has a long history of gender discrimination culture,

[26] As an extra analysis, I do find working women are 2.4 percentage points more likely to sort into private sectors.

which is shown by the extremely unbalanced sex ratio at birth (Qian, 2008). Many existing literature has discussed the unique feature of extreme child sex ratios in China, South Korea, and Northwest India. Some studies argue that the patrilineal culture interlaying with the pre-modern political and administrative systems shapes the rigid son preferences (Das Gupta, 2009; Gupta, 2005). Also, there are studies suggesting that the regional variation of contemporary gender norms can be traced back to some historical factors. For example, Alesina et al. (2013) find that regions which traditionally practice plough agriculture have less equal gender norms.

I take advantage of the fact that the son preference culture might vary across regions, and that, as a result, sex ratio at birth might be different in different places (Jayachandran, 2015). In other words, sex ratio at birth could be used as a signal to proxy the traditional gender norms. Since I do not have access to birth registry, I use 1990 census data to calculate the sex ratio for those cohorts under age 10 to proxy for existing gender norms and divide prefectures into high and low sex ratio areas and I study whether women are affected differently in these two areas. It is crucial to use pre-reform data, because it will rule out the possibility that the variation of sex ratio is driven by the SOE reform.

The Impact of SOE Reform on Gender Wage Gap in Levels

Table 2.11, and Table 2.12 show the OLS and IV estimation results. First and foremost, I find that effects are almost entirely driven by the high sex ratio areas for employment and earnings. For example, β_1 in columns (2) and (4) in Table 2.11 suggests women are less likely to be employed in areas where there is a high sex ratio at birth. The additional effect is 3.4 percentage points from OLS estimation and 9 percentage points from IV estimation. Column (2) in Table 2.12 indicates that a one standard deviation (20%) increase in the SOE reform intensity is associated with an increase of 12% in gender earnings gap in the high sex ratio areas, while a negative but insignificant effect is detected in the low sex ratio areas. IV estimation suggests the similar results, while the magnitude is bigger.[27]

[27] One concern in using sex ratio to proxy gender discrimination is that the variation of sex ratio by regions may be driven by the availability of prenatal sex selection technology. However, the ultrasound machine was first introduced to China in the early 1980s and when it came to 1987, every county had been equipped with six machines, on average (Almond et al., 2010; Chen et al., 2013). And even if the sex ratio is driven by such supply side factor, it can still be a result of different tastes for boys and girls.

The Impact of SOE Reform on Gender Wage Gap in Rank Positions

As I discussed before, it is hard to conclude that traditional gender norms play an important role in the increased gender earnings gap by just investigating the level change. The reason is such level change can be driven by two fundamental factors and one of them is the change of wage structure, which has little relationship with gender norms. In order to evaluate the relative importance of the change of wage structure and traditional gender norms in the determination of increased gender wage gap, I also look at the impact of the SOE reform on gender wage gap in rank positions by different sex ratio areas. As I discussed before, any increased gender rank gap is driven by the effect of gender-specific factors, so I estimate the impact of the SOE reform on gender rank gap by using the percentile rank in the male earnings distribution as the dependent variable in quantile regressions and I look at whether these effects differ by high versus low sex ratio areas. Results are presented in Table A.14. Generally speaking, the coefficient of interest (β_1) is much larger in the high sex ratio areas than the low sex ratio areas, and they are statistically significantly different from each other. However, none of them are significant and the point estimate is small. For example, β_1 in column (3) suggests that the 50th quantile females' position has decreased from 43th quantile in males wage distribution to 41th as the layoff intensity increases by one standard deviation (20%).

In summary, it is not obvious that traditional gender norms have played an important role in the increased gender earnings gap and in my another paper, I conduct an nonparamentric decomposition analysis to show that the dominant factor for the aggregate increased gender earnings gap is the change of wage structure.

2.9 Conclusions

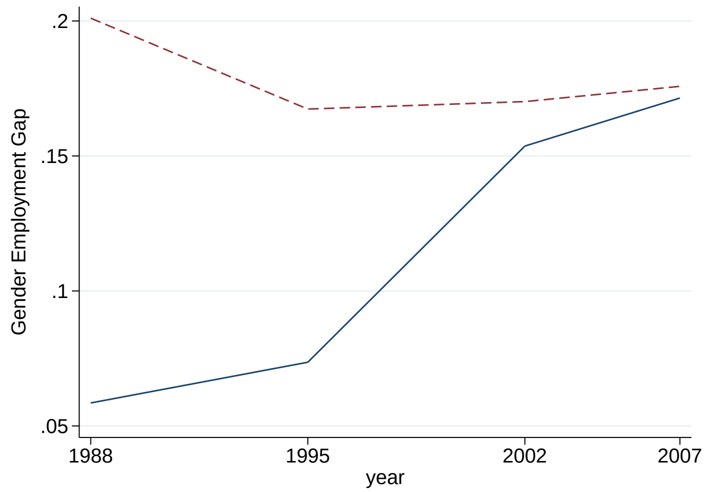
One of the most noticeable achievements in the past several decades in our society is the promotion of gender equality in almost every aspect of human activities. Recently, researchers have shifted the focus to understand the persistent sizable gender gap in politics, high earnings and high-status occupations. Many developed countries have implemented various policies to enforce gender equality in some specific occupations; however, current research does not find consistent significant, positive effects from these policies on narrowing gender gaps. This paper mainly contributes to this area and I study a unique historical period in China between 1950 and the 1990s in which the central government had been cultivating the gender equality ideology within its citizens, setting up laws to ensure gender equality in rights and, more importantly, implementing the central planning labor arrangement to guarantee an extremely high female labor force participation rate and low gender earnings gap.

Has this period of more than over 40 years of strict government intervention changed people's attitudes toward the appropriate roles between men and women in the society? My research suggests that the answer is No. I employ both difference-in-differences and instrumental variable strategies to study the causal effects of SOE reform or so called privatization movement that took place in the late 1990s on gender inequality in the labor market. My DID and IV strategies produce similar results. These two methods should be considered as complements to each other since DID estimates may have more power by using a larger sample while IV estimates tend to generate unbiased results. I find the privatization movement leads to a significant increase in the gender gaps in the labor market. The SOE reform negatively and disproportionately affects women. The greater exposure to the SOE reform, the larger the effect.

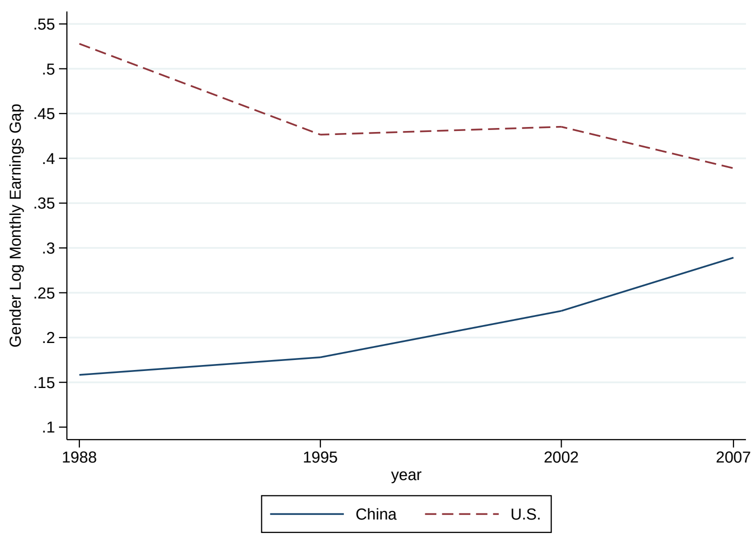
Both my DID and IV estimation results suggest that the increased gender gap is neither because of the increased household income nor because of women being less productive. By contrast, I provide suggestive evidence to shed light on the importance of gender discrimination in the increased gender gaps. The first method originates from the idea of wage decomposition and the results are not obvious for the increased gender earnings gap. I also use geographic variation in sex ratio to proxy historical gender norms, and my results suggest that gender gaps have increased in those places with high male-to-female sex ratio areas.

Figure 2.1: Change of Gender Gaps in Labor Market Outcomes: China VS U.S.

(a) Employment

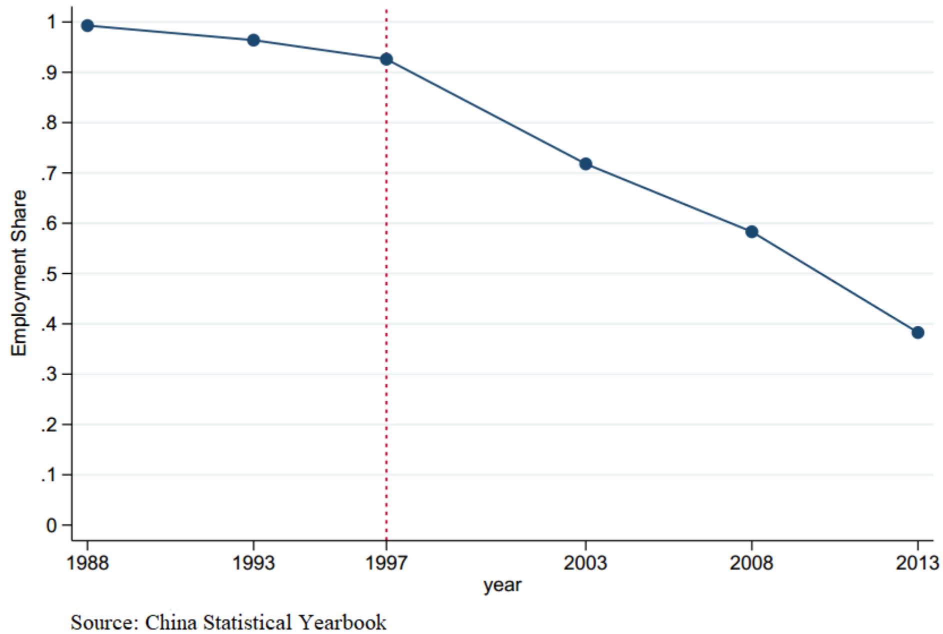


(b) Monthly Earnings



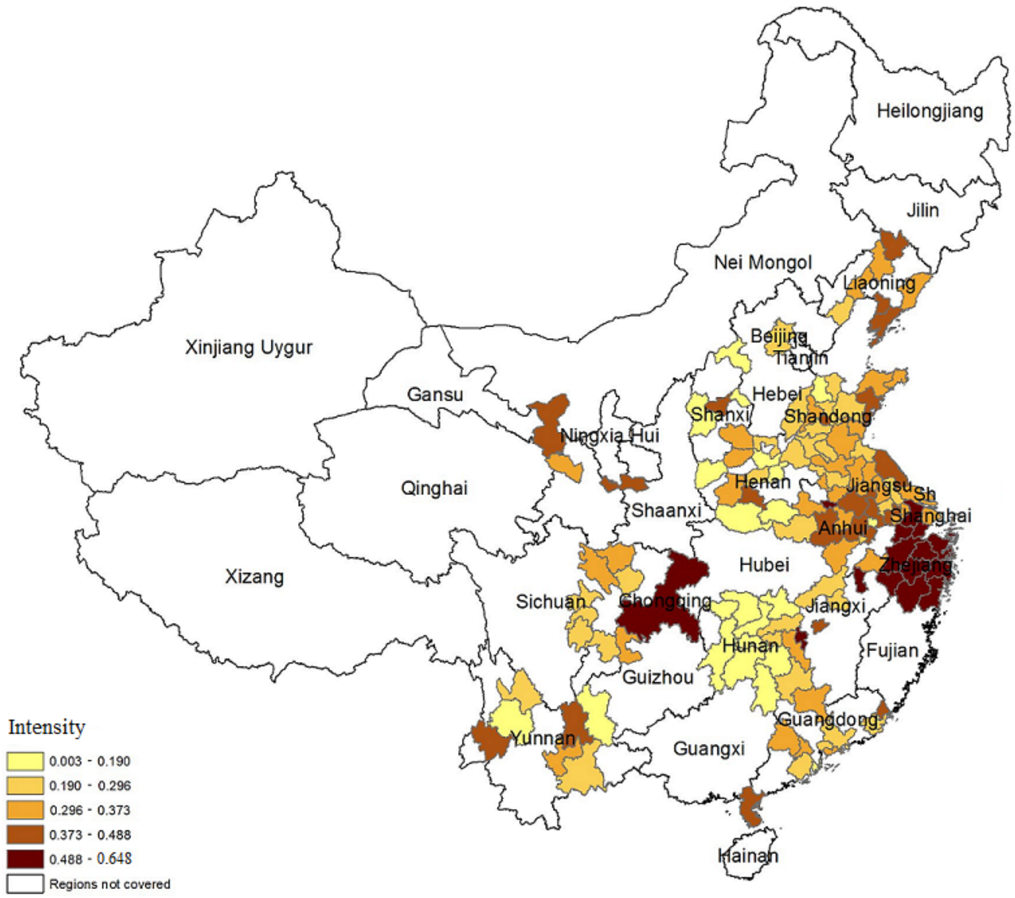
Notes: (1) Data source: China: Data comes from CHIP 1988, 1995, 2002, and 2007. I restrict to those individuals between age 19 and 54 in urban areas; U.S.: Data comes from Annual Social and Economic Supplement of the Current Population Survey. I restrict to those white individuals between age 19 and 54. When calculating monthly earnings, I further restrict to those individuals who are nonagricultural employees and report they worked 52 weeks with positive earnings in the past year. Real monthly earnings are deflated at the 2014 price level for both figures. (2) Gender gap in employment = (number of currently employed male workers/total number of male individuals) - (number of currently employed female workers/total number of female individuals);

Figure 2.2: Share of Urban Labor Force Working in SOE



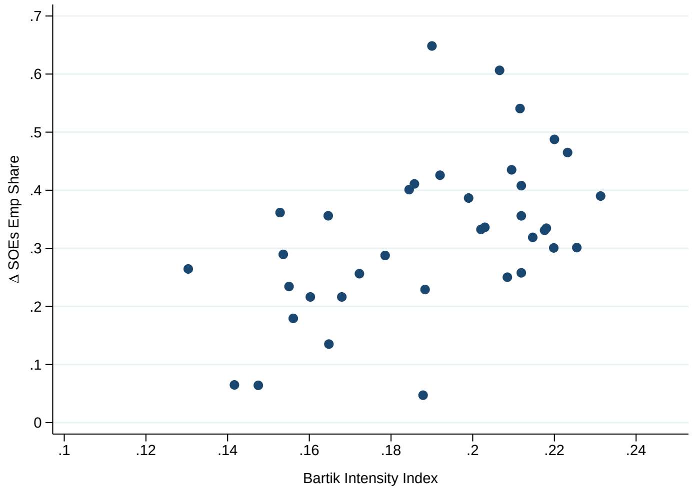
Notes: Data comes from Comprehensive Statistical Data and Materials on 50 Years of New China. China Statistic Yearbook 2004, 2009, 2014. SOE (State-Owned Enterprises) include central SOE, local SOE, and collective-owned firms in urban areas.

Figure 2.3: Regional Variation in Change of SOE Employment Share



Notes: Δ SOE Emp share = (number of workers in SOE_{1995}/total number of workers in a given urban prefecture area_{1995}) - (number of workers in SOE_{2001}/total number of workers in a given urban prefecture area_{2001}). SOE include central SOE, local SOE and collective-owned firms in the urban areas. Data comes from China Provincial Statistical Yearbook and China Labor Statistical Yearbook 1996, 2002. White color refers to the regions which are not covered by the CHIP survey or no data available.

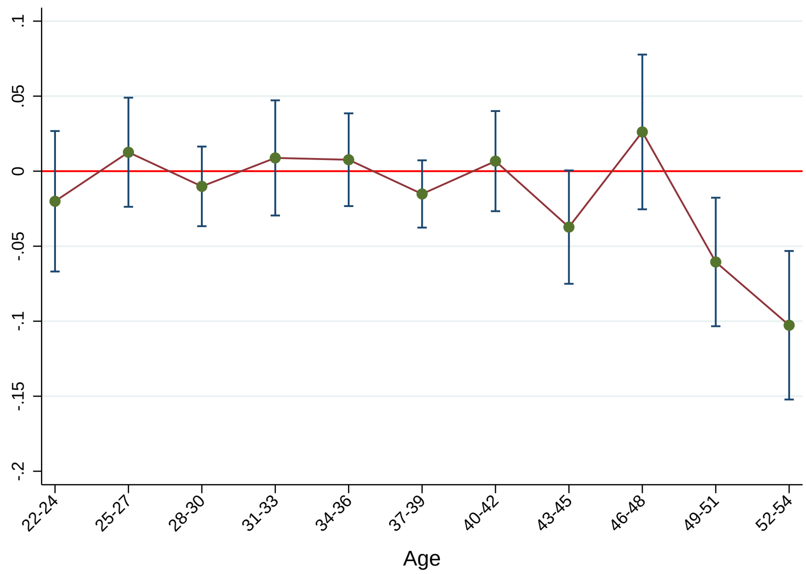
Figure 2.4: Relationship between Bartik Shift-share Intensity Index and Change of SOE Employment Share



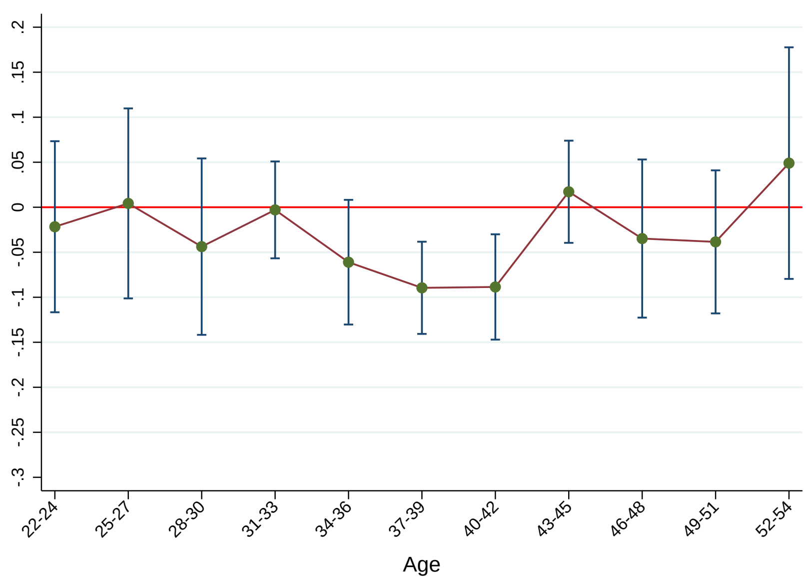
Notes: Data comes from China Provincial Statistical Yearbook, China Labor Statistical Yearbook, China Statistical Yearbook 1996, 2002. Bartik Intensity Index$_p$ = ($\sum_{i=1}^{n}$ pre-reform share of SOE workers in industry i at prefecture p * Δ SOE employment share in industry i at the national level). Pre-reform share of SOE workers in industry i at prefecture p = (number of SOE workers in industry i at prefecture p/ number of workers at prefecture p). Δ SOE employment share in industry i at the national level = ((national number of SOE workers in industry i_{1995}/national number of workers in industry i_{1995}) - (national number of SOE workers in industry i_{2001}/national number of workers in industry i_{2001})). Δ SOE Emp share = (number of workers in SOE_{1995}/total number of workers in a given urban prefecture $area_{1995}$) - (number of workers in SOE_{2001}/total number of workers in a given urban prefecture $area_{2001}$).

Figure 2.5: OLS Estimate Coefficients of the Impacts of SOE Reform

(a) Employment



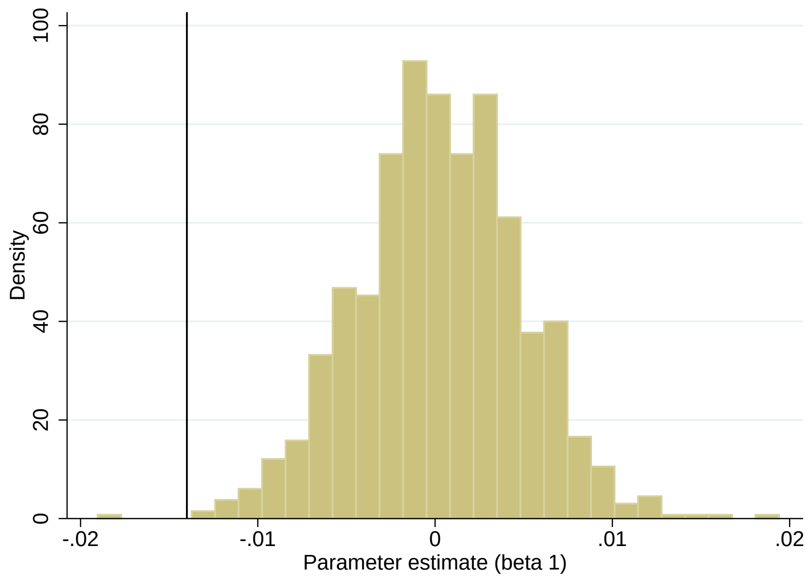
(b) Monthly Earnings



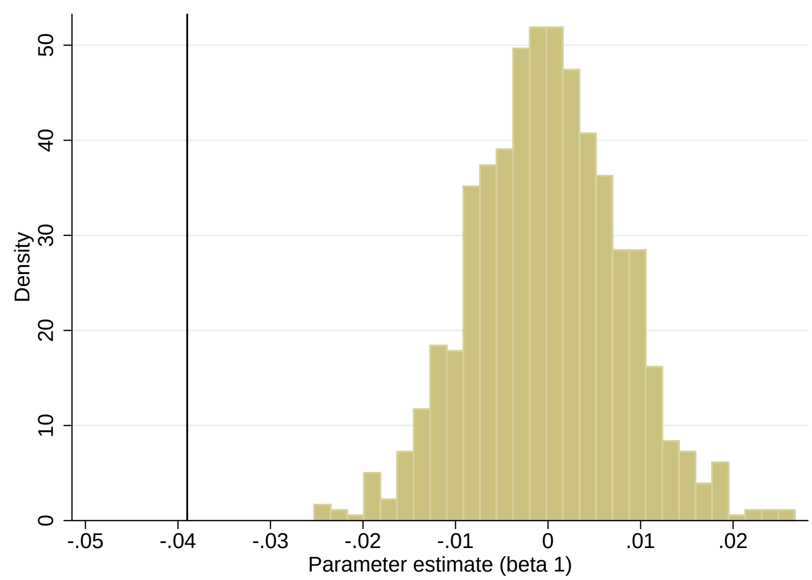
Notes: Regression estimates by Equation 2.6 are plotted. The dot and the bar correspond to the coefficient estimates with 90% confidence intervals.

Figure 2.6: Permutation Test Results, Coefficient of (female*after*Δ SOE emp share β_1)

(a) Employment



(b) Monthly Earnings



Notes: I assigned placebo treatment in randomly selected years and prefectures drawn without replacement. The histogram displays the coefficient estimates of a triple interaction term: female, after, and Δ SOE Emp share from 1,000 permutations. The vertical line shows the estimates of the actual treatment effect. Female and after are two dummy variables. Δ SOE Emp share = (number of workers in SOE_{1995}/total number of workers in a given urban prefecture area_{1995}) - (number of workers in SOE_{2001}/total number of workers in a given urban prefecture area_{2001}. Panel A shows that 1 out of 1,000 permutation estimates (absolute value) is greater than that of actual treatment. Panel B shows that 0 out of 1,000 permutation estimates (absolute value) is greater than that of actual treatment.

Table 2.1: Summary Statistics of Key Variables: 1988 - 2007

	Before (1988 and 1995)	After (2002 and 2007)
Panel A: selected labor market outcomes		
Monthly earnings (in year 2014 RMB)	587.29	1823.22
	(359.72)	(2080.53)
Currently employed	0.91	0.71
	(0.28)	(0.46)
Retired	0.04	0.09
	(0.19)	(0.29)
Work in private sectors	0.03	0.53
	(0.17)	(0.50)
Panel B: individual characteristics		
Female	0.51	0.51
	(0.50)	(0.50)
Age	36.76	38.93
	(9.74)	(9.89)
Minority	0.04	0.03
	(0.20)	(0.17)
Years of schooling	9.92	11.39
	(2.92)	(3.25)
Potential work experience	21.11	21.81
	(10.29)	(11.27)
Communist party membership	0.21	0.24
	(0.41)	(0.43)
Observations	31235	21135

Note: Unweighted means and standard deviations are presented. Standard deviations in parentheses. Individuals are between age 19 and 54. Not every wave asks the actual work experience, potential work experience equals years of schooling minus age and 6.

Table 2.2: Prefectural Pre-reform Share of SOE Workers, by Industry

Industry	Mean	St.Dev.	Min	Max
Mining	0.018	0.026	0.001	0.117
Manufacturing	0.418	0.086	0.233	0.542
Electricity, Gas and Water Production and Supply	0.015	0.006	0.008	0.032
Construction	0.057	0.025	0.031	0.135
Transport, Storage and Communications	0.050	0.018	0.021	0.116
Wholesale and Retail Trade, Restaurants	0.133	0.032	0.065	0.209
Financial Intermediation and Insurance	0.018	0.005	0.006	0.209
Real Estate Activities	0.006	0.005	0.001	0.029
Social Services	0.032	0.020	0.010	0.099
Scientific Research and Polytechnical Services	0.014	0.016	0.002	0.073

Notes: Data for 37 prefectures comes from Provincial Statistical Yearbook 1996, 1995. Pre-reform share of SOE workers = (number of SOE workers in industry i at a given prefecture p_{1995}/ total number of workers in prefecture area p_{1995}) in the urban areas.

Table 2.3: National Change of SOE Employment Share, by Industry

Industry	Δ SOE Employment Share
Mining	0.202
Manufacturing	0.334
Electricity, Gas and Water Production and Supply	0.126
Construction	0.190
Geological Prospecting and Water Conservancy	0.004
Transport, Storage and Communications	0.089
Wholesale and Retail Trade, Restaurants	0.170
Financial Intermediation and Insurance	0.096
Real Estate Activities	0.211
Social Services	0.122
Scientific Research and Polytechnical Services	0.070

Notes: Data comes from China Statistical Yearbook 1996, 2002. Δ SOE emp share at the national level = (national number of SOE workers in industry i_{1995}/ total number of workers in industry i_{1995}) - (national number of SOE workers in industry i_{2001}/ total number of workers in industry i_{2001}).

Table 2.4: Gender Gaps in the Labor Market

	Before	After
	1988 and 1995	2002 and 2007
Panel A: Employed		
Female	-.056***	-.152***
	(.004)	(.009)
Male mean	0.96	0.83
Female mean	0.89	0.65
Obs.	29693	20709
Panel B: Retired		
Female	.047***	.114***
	(.003)	(.008)
Male mean	0.01	0.03
Female mean	0.06	0.15
Obs.	29813	20709
Panel C: ln(Monthly earnings)		
Female	-.121***	-.225***
	(.009)	(.015)
Male mean	634.17	2113.47
Female mean	539.45	1513.63
Obs.	26483	14508
Panel D: Work in private sectors		
Female	.002	.071***
	(.002)	(.009)
Male mean	0.02	0.49
Female mean	0.02	0.57
Obs.	28093	14407

Notes: Sample includes all individuals between age 19 and 54. Monthly earnings are deflated at the 2014 price level. Robust standard errors are clustered at the prefecture level. Regression controls for age, age squared, years of schooling, ethnicity, prefecture, and year dummies. Panel C include those individuals who report they currently have a full-time job. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.5: Gender Differences in Log Monthly Earnings, Quantile Regressions- 1988-2007

	1988	1995	2002	2007
10th Quantile	-0.137***	-0.161***	-0.257***	-0.223***
	(0.016)	(0.023)	(0.023)	(0.029)
25th Quantile	-0.111***	-0.133***	-0.226***	-0.277***
	(0.011)	(0.013)	(0.027)	(0.028)
50th Quantile	-0.101***	-0.096***	-0.198***	-0.297***
	(0.010)	(0.012)	(0.019)	(0.011)
75th Quantile	-0.097***	-0.090***	-0.141***	-0.276***
	(0.009)	(0.010)	(0.018)	(0.022)
90th Quantile	-0.109***	-0.122***	-0.132***	-0.230***
	(0.012)	(0.012)	(0.018)	(0.026)
Obs.	16,222	10,743	8,775	5,811

Notes: Each main cell of the table reports the coefficient from 10th, 25th, 50th, 75th, and 90th quantile regressions of the individual's log real monthly earnings on female dummy variable. Additional controls include education categories, age categories, and prefecture fixed effects. Sample includes all individuals between age 19 and 54 who report they currently have a full-time job. Monthly earnings are deflated at the 2014 price level. Robust standard errors are clustered at the prefecture level.

Table 2.6: Gender Differences in Rank in Male Earnings Distribution, Quantile Regressions-1988-2007

	1988	1995	2002	2007
10th Quantile	-6.190***	-4.403***	-5.566***	-6.648***
	(0.554)	(0.681)	(0.754)	(0.978)
25th Quantile	-6.905***	-5.850***	-8.533***	-11.121***
	(0.563)	(0.600)	(0.930)	(1.678)
50th Quantile	-7.753***	-5.900***	-10.777***	-14.812***
	(0.725)	(0.662)	(1.147)	(1.205)
75th Quantile	-8.555***	-6.061***	-7.417***	-12.728***
	(0.857)	(0.740)	(0.921)	(1.091)
90th Quantile	-7.471***	-5.678***	-4.816***	-7.307***
	(0.942)	(0.822)	(0.767)	(0.976)
Obs.	16,222	10,743	8,775	5,811

Notes: Each main cell of the table reports the coefficient from 10th, 25th, 50th, 75th, and 90th quantile regressions of the individual's percentile rank in the male earnings distribution on female dummy variable. Additional controls include education categories, age categories, and prefecture fixed effects. Sample includes all individuals between age 19 and 54 who report they currently have a full-time job. Monthly earnings are deflated at the 2014 price level. Robust standard errors are clustered at the prefecture level.

Table 2.7: OLS Estimates of the Impacts of SOE Reform

	All		Subsample	
Dependent variable	Employed	ln(Monthly earnings)	Employed	ln(Monthly earnings)
Mean	0.83	RMB1068.25	0.82	RMB1294.63
St.dev.	0.38	RMB1458.28	0.39	RMB1752.63
	(1)	(2)	(3)	(4)
Female × after × Δemp share, β_1	-.014**	-.039**	-.024**	-.048***
	(.007)	(.016)	(.012)	(.014)
Female × after, β_2	-.105***	-.118***	-.098***	-.132***
	(.009)	(.014)	(.014)	(.019)
Female, β_3	-.054***	-.123***	-.049***	-.099***
	(.004)	(.010)	(.005)	(.013)
Female × Δemp share, β_4	.015***	.002	.010	-.008
	(.005)	(.019)	(.007)	(.012)
After × Δemp share, β_5	.002	-.017	.005	.034
	(.010)	(.025)	(.016)	(.054)
Obs.	47522	37829	25450	19427
Number of prefectures	80		37	

Notes: Individuals between age 19 and 54. Monthly earnings are deflated at the 2014 year level. Columns(3) and column (6) include individuals who report they currently have a full-time job. Subsample includes those 37 prefectures with pre-reform number of SOE workers by industry. All models include age, age squared, years of schooling, ethnic minority, prefecture fixed effects, year fixed effects and prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.8: 2sls Estimates of the Impacts of SOE Reform

	First stage	IV	
Dependent Variable	Δemp share	Employed	ln(Monthly earnings)
Mean	0.33	0.82	RMB1294.63
St.Dev.	0.13	0.39	RMB1752.63
	(1)	(2)	(3)
Bartik intensity	0.400***		
	(0.122)		
Female $\times$ after $\times$ Δemp share, β_1		-0.069**	-0.084*
		(0.028)	(0.044)
Female $\times$ after, β_2		-0.082***	-0.117***
		(0.016)	(0.019)
Female, β_3		-0.050***	-0.103***
		(0.005)	(0.011)
Female $\times$ Δemp share, β_4		0.011	0.014
		(0.013)	(0.033)
After $\times$ Δemp share, β_5		-0.038	-0.016
		(0.039)	(0.093)
F-statistics	12.15		
p-value	0.00		
Obs.	25502	25450	19427

Notes: Individuals between ages 19 and 54. Monthly earnings are deflated at the 2014 year price level. Column (3) includes individuals who report they currently have a full time job. Subsample includes those 37 prefectures with pre-reform number of SOE workers by industry. Column (2) and column (3) include age, age squared, years of schooling, ethnic minority, prefecture fixed effects, year fixed effects, and prefecture specific time trend. Column (1) includes age, age squared, years of schooling, ethnic minority, year fixed effects. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. Bartik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.9: Estimates of the Impact of SOE Reform on Employment, by Household Income and Education Attainment

	High Income	Low Income	Edu>=High School	Edu<High School
	(1)	(2)	(3)	(4)
Female × after × Δemp share, β_1	-.014	-.014**	.010	-.049***
	(.010)	(.007)	(.008)	(.013)
Female × after, β_2	-.066***	-.130***	-.089***	-.181***
	(.012)	(.010)	(.009)	(.015)
Female × Δemp share, β_3	.008	.017***	.0001	.035***
	(.006)	(.006)	(.004)	(.009)
After × Δemp share, β_4	.017	-.009	-.003	.005
	(.015)	(.009)	(.009)	(.022)
Female, β_4	-.039***	-.061***	-.012***	-.104***
	(.004)	(.005)	(.003)	(.007)
Obs.	19255	28267	29032	18490

Notes: Individuals between age 19 and 54. All regressions include age, age squared, years of schooling, ethnic minority, year fixed effects, prefecture fixed effects, and prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.10: OLS Estimates of the Impact of SOE Reform on Monthly Earnings (young cohort, age<=40), additional controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female $\times$ after $\times$ Δemp share, β_1	-.046***	-.040***	-.044***	-.043**	-.038**	-.041**	-.030**
	(.015)	(.015)	(.015)	(.017)	(.015)	(.016)	(.013)
Female $\times$ after, β_2	-.122***	-.120***	-.108***	-.047**	-.068***	-.076***	-.138***
	(.019)	(.018)	(.019)	(.023)	(.018)	(.023)	(.016)
Female, β_3	-.093***	-.089***	-.088***	-.118***	-.123***	-.114***	-.091***
	(.010)	(.010)	(.011)	(.012)	(.011)	(.011)	(.010)
Industry		Yes					Yes
Occupation			Yes				Yes
Child under age 6				Yes			Yes
Household income					Yes		Yes
Work in private sectors						Yes	Yes
Adj. R-squared	0.6192	0.6292	0.6303	0.6130	0.6756	0.6950	0.6987
Obs.	22603	22338	22351	22603	22601	22141	21802

Notes: monthly earnings are deflated at the 2014 price level. Samples include individuals who report that they currently have a full-time job. All models include age, age squared, years of schooling, ethnic minority, year fixed effects, prefecture fixed effects, and prefectural specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.11: Estimates of the Impact of SOE Reform on Employment, by Intensity of Male-to-Female Sex Ratio

	OLS		IV	
	Low(0.99-1.07)	High(1.08-1.26)	Low(0.99-1.07)	High(1.08-1.26)
	(1)	(2)	(3)	(4)
Female × after × Δemp share, β_1	-.008	-.034**	-.038	-.090***
	(.009)	(.015)	(.032)	(.033)
Female × after, β_2	-.103***	-.106***	-.095***	-.075***
	(.012)	(.015)	(.016)	(.021)
Female, β_3	-.051***	-.045***	-.055***	-.046***
	(.005)	(.005)	(.008)	(.006)
Obs.	22032	23076	11124	14326

Notes: Dependent variable is employed. Individuals between age 19 and 54 . First two columns include full sample. Column (3) and column (4) include 37 prefectures with pre-reform industry employment composition. Low refers to low male/female sex ratio at birth; High refers to high male-to-female sex ratio at birth. All models include age, age squared, years of schooling, ethnic minority, working industries, year fixed effects, prefecture fixed effects and prefecture specific time trend. Sex ratio at birth is calculated by using the 1990 census, I restrict to those individuals who are under age 10. The mean of sex ratio at birth is 1.09 with standard deviation 0.06. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. Bartik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.12: Estimates of the Impact of SOE Reform on Monthly Earnings, by Intensity of Male-to-Female Sex Ratio

	OLS		IV	
	Low(0.99-1.07)	High(1.08-1.26)	Low(0.99-1.07)	High(1.08-1.26)
	(1)	(2)	(3)	(4)
Female × after × Δemp share, β_1	-.007	-.067***	-.003	-.126**
	(.018)	(.013)	(.105)	(.059)
Female × after, β_2	-.093***	-.111***	-.136***	-.102***
	(.018)	(.018)	(.030)	(.027)
Female, β_3	-.127***	-.117***	-.079***	-.116***
	(.014)	(.011)	(.012)	(.015)
Obs.	19881	16918	8422	10761

Notes: Dependent variable is ln(monthly earnings). Individuals between age 19 and 54 who report they currently have a full time job. First two columns include full sample. Column (3) and column (4) include 37 prefectures with pre-reform industry employment composition. Monthly earnings are deflated at the 1988 price level. Low refers to low male/female sex ratio at birth; High refers to high male-to-female sex ratio at birth. All models include age, age squared, years of schooling, ethnic minority, working industries, year fixed effects, prefecture fixed effects and prefecture specific time trend. Sex ratio at birth is calculated by using the 1990 census, I restrict to those individuals who are under age 10. The mean of sex ratio at birth is 1.09 with standard deviation 0.06. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. Bartik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.13: Estimates of the Impact of SOE Reform on Earnings Rank Positions, by Intensity of Male-to-Female Sex Ratio

	Overall	Low(0.99-1.07)	High(1.08-1.26)
	(1)	(2)	(3)
Female $\times$ after $\times$ Δemp share, β_1	-1.408	-.981	-1.849
	(1.172)	(1.628)	(1.422)
Female $\times$ after, β_2	-1.337*	.543	-3.052***
	(.737)	(.886)	(.919)
Female, β_3	-7.807***	-8.559***	-6.966***
	(.574)	(.959)	(.509)
Female $\times$ Δemp share, β_4	-.227	-.478	-.075
	(1.133)	(1.870)	(.716)
After $\times$ Δemp share, β_5	-.547	-1.129	-.219
	(1.135)	(1.288)	(1.888)
Obs.	37829	18553	19276

Notes: Individuals include those between age 19 and 54 who report they currently have a full-time job. All models include age, age squared, years of schooling, and ethnicity. All models also include prefecture fixed effects, year fixed effects and prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.14: Placebo Test: 1988 and 1995

Dependent Variable	Employed	ln(Monthly earnings)
	(1)	(2)
pseuFemale $\times$ after $\times$ Δemp share, β_1	-.006	-.013
	(.008)	(.022)
pseuFemale $\times$ after, β_2	-.015*	-.035**
	(.008)	(.016)
Female, β_3	-.050***	-.113***
	(.004)	(.011)
Female $\times$ Δemp share	.018***	.004
	(.005)	(.022)
pseuAfter $\times$ Δemp share	.001	.021
	(.006)	(.031)
Obs.	27374	24074

Notes: I assume the SOE reform happened some time between 1988 and 1995. Thus, 1995 would be pseudo-after year. Sample includes all individuals between ages 19 and 54. Column (2) includes those individuals who report they currently have a full-time job. Monthly earnings are deflated at the 2014 price level. Robust standard errors are clustered at the prefecture level. All models control for age, age squared, years of schooling, ethnicity, prefecture, and year dummies. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Bibliography

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel, "Regulating consumer financial products: Evidence from credit cards," *The Quarterly Journal of Economics*, 2014, p. qju037.

Agranov, Marina and Pietro Ortoleva, "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 2017, *125* (1), 000–000.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough*.," *Quarterly Journal of Economics*, 2013, *128* (2).

Almond, Douglas, Hongbin Li, and Lingsheng Meng, "Son preference and early childhood investments in China," *manuscript, Tsinghua University*, 2010.

Altonji, Joseph G and David Card, "The effects of immigration on the labor market outcomes of less-skilled natives," in "Immigration, trade, and the labor market," University of Chicago Press, 1991, pp. 201–234.

Andersen, Steffen, Glenn W Harrison, Morten I Lau, and E Elisabet Rutström, "Eliciting risk and time preferences," *Econometrica*, 2008, *76* (3), 583–618.

_ , _ , Morten Igel Lau, and E Elisabet Rutström, "Elicitation using multiple price list formats," *Experimental Economics*, 2006, *9* (4), 383–405.

Andreoni, James and Charles Sprenger, "Uncertainty equivalents: Testing the limits of the independence axiom," Technical Report, National Bureau of Economic Research 2011.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva, "Does the Gender Composition of Scientific Committees Matter?," *Available at SSRN 2628176*, 2015.

Barr, Abigail and Garance Genicot, "Risk sharing, commitment, and information: an experimental analysis," *Journal of the European Economic Association*, 2008, *6* (6), 1151–1185.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?* number wbsle. In 'Books from Upjohn Press.', W.E. Upjohn Institute for Employment Research, November 1991.

Bartik, Timothy J, "The effects of metropolitan job growth on the size distribution of family income," *Journal of Regional Science*, 1994, *34* (4), 483–501.

Basso, Gaetano and Giovanni Peri, "The Association between Immigration and Labor Market Outcomes in the United States," IZA Discussion Papers 9436, Institute for the Study of Labor (IZA) October 2015.

Bauer, John, Wang Feng, Nancy E Riley, and Zhao Xiaohua, "Gender inequality in urban China: Education and employment," *Modern China*, 1992, *18* (3), 333–370.

Bauermeister, Golo and Oliver Musshoff, "Risk Aversion and Inconsistencies-Does the Choice of Risk Elicitation Method and Display Format Influence the Outcomes?," Technical Report, Agricultural and Applied Economics Association 2016.

Bayer, Patrick and Kerwin Kofi Charles, "Divergent Paths: Structural Change, Economic Rank, and the Evolution of Black-White Earnings Differences, 1940-2014," Technical Report, National Bureau of Economic Research 2016.

Bertrand, Marianne, "New perspectives on gender," *Handbook of labor economics*, 2011, *4*, 1543–1590.

_ , Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney, "Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway," Technical Report, National Bureau of Economic Research 2014.

Binswanger, Hans P, "Attitudes toward risk: Experimental measurement in rural India," *American journal of agricultural economics*, 1980, *62* (3), 395–407.

Blau, Francine D and Lawrence M Kahn, "Wage structure and gender earnings differentials: an international comparison," *Economica*, 1996, pp. S29–S62.

_ **and** _ , "Swimming upstream: Trends in the gender wage differential in the 1980s," *Journal of labor Economics*, 1997, *15* (1, Part 1), 1–42.

_ **and** _ , "Gender differences in pay," Technical Report, National bureau of economic research 2000.

_ **and** _ , "Understanding international differences in the gender pay gap," *Journal of Labor economics*, 2003, *21* (1), 106–144.

_ **and** _ , "The gender pay gap have women gone as far as they can?," *The Academy of Management Perspectives*, 2007, *21* (1), 7–23.

_ **and** _ , "The gender wage gap: Extent, trends, and explanations," Technical Report, National Bureau of Economic Research 2016.

Bloom, David, Elizabeth Cafiero, Eva Jané-Llopis, Shafika Abrahams-Gessel, Lakshmi Bloom, Sana Fathima, Andrea Feigl, Tom Gaziano, Ali Hamandi, Mona Mowafi et al., "The global economic burden of noncommunicable diseases," Technical Report, Program on the Global Demography of Aging 2012.

Booth, Alison L, Elliott Fan, Xin Meng, and Dandan Zhang, "Gender Differences in Willingness to Compete: The Role of Culture and Institutions," 2016.

Brick, Kerri, Martine Visser, and Justine Burns, "Risk aversion: experimental evidence from South African fishing communities," *American Journal of Agricultural Economics*, 2012, *94* (1), 133–152.

Bruner, David, "Multiple switching behavior in multiple price lists," *Applied Economics Letters*, 2011.

Burda, Michael C and Jennifer Hunt, "From reunification to economic integration: productivity and the labor market in Eastern Germany," *Brookings Papers on Economic Activity*, 2001, *2001* (2), 1–71.

Cai, Fang, Albert Park, and Yaohui Zhao, "The Chinese labor market in the reform era," *Chinas great economic transformation*, 2008, pp. 167–214.

Card, David, "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

_ , "Immigration and Inequality," *American Economic Review*, May 2009, *99* (2), 1–21.

Cassar, Alessandra, Feven Wordofa, and Y. Jane Zhang, "Competing for the benefit of offspring eliminates the gender gap in competitiveness," *Proceedings of the National Academy of Sciences*, 2016, p. 201520235.

Chao, S-C, "The Reform of State-owned Enterprises in Mainland China: Re-examining State-Society Relations," *New Zealand Journal of Asian Studies*, 2000, *2*, 78–96.

Charness, Gary and Angelino Viceisza, "Three risk-elicitation methods in the field: Evidence from rural Senegal," *Review of Behavioral Economics*, 2016.

_ **, Uri Gneezy, and Alex Imas**, "Experimental methods: Eliciting risk preferences," *Journal of Economic Behavior & Organization*, 2013, *87*, 43–51.

Chen, Yuyu, Hongbin Li, and Lingsheng Meng, "Prenatal sex selection and missing girls in China: Evidence from the diffusion of diagnostic ultrasound," *Journal of Human Resources*, 2013, *48* (1), 36–70.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 2011, *126* (4), 1593–1660.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman, "Who is (more) rational?," *The American Economic Review*, 2014, *104* (6), 1518–1550.

Cooney, Sean, "Making Chinese Labor Law Work: The Prospects for Regulatory Innovation in the People's Republic of China," *Fordham Int'l LJ*, 2006, *30*, 1050.

Danthine, Jean-Pierre and Jennifer Hunt, "Wage bargaining structure, employment and economic integration," *The Economic Journal*, 1994, pp. 528–541.

Dave, Chetan, Catherine Eckel, Cathleen Johnson, and Christian Rojas, "Eliciting risk preferences: When is simple better?," *Journal of Risk and Uncertainty*, December 2010, *41* (3), 219–243.

David, H, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *The American Economic Review*, 2013, *103* (6), 2121–2168.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner, "Individual risk attitudes: Measurement, determinants, and behavioral consequences," *Journal of the European Economic Association*, 2011, *9* (3), 522–550.

Dohmen, Thomas J, Falk Armin, Bart Golsteyn, David Huffman, and Uwe Sunde, "Risk attitudes across the life course," 2015.

Dong, Yilin, "A note on geographical constraints and housing markets in China," *Journal of Housing Economics*, 2016, *33*, 15–21.

Du, Fenglian and Xiao yuan Dong, "Why do women have longer durations of unemployment than men in post-restructuring urban China?," *Cambridge Journal of Economics*, 2009, *33* (2), 233–252.

Eckel, Catherine C and Philip J Grossman, "Sex differences and statistical stereotyping in attitudes toward financial risk," *Evolution and human behavior*, 2002, *23* (4), 281–295.

Eichen, Marc and Ming Zhang, "Annex: The 1988 household sample surveydata description and availability," *The distribution of income in china*, 1993, *3314346.*

Entwisle, Barbara and Gail Henderson, *Re-drawing boundaries: work, households, and gender in China*, Vol. 25, Univ of California Press, 2000.

Frazier, Mark W, "State-sector shrinkage and workforce reduction in China," *European Journal of Political Economy*, 2006, *22* (2), 435–451.

Gaudecker, Hans-Martin Von, Arthur Van Soest, and Erik Wengström, "Heterogeneity in risky choice behavior in a broad population," *The American Economic Review*, 2011, *101* (2), 664–694.

Giles, John, Albert Park, and Fang Cai, "How has economic restructuring affected China's urban workers?," *The China Quarterly*, 2006, *185*, 61–95.

_ , _ , and _ , "Reemployment of dislocated workers in urban China: The roles of information and incentives," *Journal of Comparative Economics*, 2006, *34* (3), 582–607.

Gneezy, Uri, Kenneth L Leonard, and John A List, "Gender differences in competition: Evidence from a matrilineal and a patriarchal society," *Econometrica*, 2009, *77* (5), 1637–1664.

_ , Muriel Niederle, and Aldo Rustichini, "Performance in competitive environments: Gender differences," *The Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *The American Economic Review*, 2014, *104* (4), 1091–1119.

Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton, "China's evolving managerial labor market," *Journal of Political Economy*, 1995, pp. 873–892.

Gupta, Monica Das, "Explaining Asia's missing women: a new look at the data," *Population and development review*, 2005, *31* (3), 529–535.

Gupta, Monica Das, "Family systems, political systems, and Asia's' missing girls': the construction of son preference and its unraveling," 2009.

Gustafsson, Björn and Shi Li, "Economic transformation and the gender earnings gap in urban China," *Journal of Population Economics*, 2000, *13* (2), 305–329.

Ha, Wei, Junjian Yi, and Junsen Zhang, "Brain drain, brain gain, and economic growth in China," *China Economic Review*, 2016, *38*, 322–337.

Hannum, Emily and Yu Xie, *Trends in educational gender inequality in China: 1949-1985*, University of Michigan, 1994.

Hansen, Casper Worm, Peter Sandholt Jensen, and Christian Volmar Skovsgaard, "Modern gender roles and agricultural history: the Neolithic inheritance," *Journal of Economic Growth*, 2015, *20* (4), 365–404.

Harrison, Glenn and Elisabet Rutström, "Risk aversion in the laboratory," in "Risk aversion in experiments," Emerald Group Publishing Limited, 2008, pp. 41–196.

Harrison, Glenn W, Morten I Lau, and Melonie B Williams, "Estimating individual discount rates in Denmark: A field experiment," *The American Economic Review*, 2002, *92* (5), 1606–1617.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, December 2002, *92* (5), 1644–1655.

Hsieh, Chang-Tai and Zheng Michael Song, "Grasp the large, let go of the small: the transformation of the state sector in China," Technical Report, National Bureau of Economic Research 2015.

Hunt, Jennifer, "The transition in East Germany: When is a ten-point fall in the gender wage gap bad news?," *Journal of Labor Economics*, 2002, *20* (1), 148–169.

_ , "Convergence and determinants of non-employment durations in Eastern and Western Germany," *Journal of Population Economics*, 2004, *17* (2), 249–266.

Jacobson, Sarah and Ragan Petrie, "Learning from mistakes: What do inconsistent choices over risk tell us?," *Journal of Risk and Uncertainty*, 2009, *38* (2), 143–158.

Jayachandran, Seema, "The Roots of Gender Inequality in Developing Countries," *economics*, 2015, *7* (1), 63–88.

Jenq, Christina, "Privatization, Retirement Policy, and Labor Market Gender Gaps in Urban China, 1990-2005," 2015.

Johnson, Kay Ann, *Women, the family, and peasant revolution in China*, University of Chicago Press, 2009.

Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce, "Accounting for the slowdown in black-white wage convergence," *Workers and their wages*, 1991, pp. 107–43.

— , — , and — , "Wage inequality and the rise in returns to skill," *Journal of political Economy*, 1993, *101* (3), 410–442.

Kahneman, Daniel, "Maps of bounded rationality: Psychology for behavioral economics," *The American economic review*, 2003, *93* (5), 1449–1475.

— , *Thinking, fast and slow*, Macmillan, 2011.

— , Jack L Knetsch, and Richard H Thaler, "Experimental tests of the endowment effect and the Coase theorem," *Journal of political Economy*, 1990, pp. 1325–1348.

Kidd, Michael P and Xin Meng, "The Chinese state enterprise sector: Labour market reform and the impact on male–female wage structure," *Asian Economic Journal*, 2001, *15* (4), 405–423.

King, Elizabeth M and Andrew David Mason, "Engendering development through gender equality in rights resources and voice. Summary.," 2001.

Lee, Ching Kwan, *Against the law: Labor protests in Chinas rustbelt and sunbelt*, Univ of California Press, 2007.

Lee, Hong Yung, "Xiagang, the Chinese style of laying off workers," *Asian Survey*, 2000, *40* (6), 914–937.

Lemieux, Thomas, "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?," *The American Economic Review*, 2006, *96* (3), 461–498.

Li, Shi, Chuliang Luo, Zhong Wei, and Ximing Yue, "The 1995 and 2002 household surveys: Sampling methods and data description," *Inequality and public policy in China*, 2008, pp. 337–353.

Lin, Justin Yifu and Guofu Tan, "Policy burdens, accountability, and the soft budget constraint," *The American Economic Review*, 1999, *89* (2), 426–431.

— , **Fang Cai, and Zhou Li**, "Competition, policy burdens, and state-owned enterprise reform," *The American Economic Review*, 1998, *88* (2), 422–427.

Liu, Elaine, Shu Zhang et al., "A Meta-analysis of the Estimates of Returns to Schooling in China," *Department of Economics*, 2008.

Liu, Haoming, "Economic reforms and gender inequality in urban China," *Economic Development and Cultural Change*, 2011, *59* (4), 839–876.

Luo, Chuliang, Shi Li, Terry Sicular, Quheng Deng, and Ximing Yue, "The 2007 household surveys: sampling methods and data description," *Rising Inequality in China: Challenges to a Harmonious Society. Cambridge University Press, Cambridge*, 2013, pp. 445–464.

Luo, Dongdong and Chunbing Xing, "Who Is More Mobile in Response to Local Demand Shifts in China?," 2015.

Megginson, William L and Jeffry M Netter, "From state to market: A survey of empirical studies on privatization," *Journal of economic literature*, 2001, *39* (2), 321–389.

Meier, Stephan and Charles D Sprenger, "Discounting financial literacy: Time preferences and participation in financial education programs," *Journal of Economic Behavior & Organization*, 2013, *95*, 159–174.

Meijer, Marinus Johan, *Marriage law and policy in the Chinese People's Republic*, Hong Kong University Press, 1971.

Meng, Xin, *Labour market reform in China*, Cambridge University Press, 2000.

— , "Labor market outcomes and reforms in China," *The Journal of Economic Perspectives*, 2012, *26* (4), 75–101.

Millimet, Daniel L and Le Wang, "A distributional analysis of the gender earnings gap in urban China," *The BE Journal of Economic Analysis & Policy*, 2006, *5* (1).

Ni, Yuan et al., "Gender Wage Gap in Urban China," *Ekonomia journal*, 2005, *17.*

Niederle, Muriel, "Gender," in "Handbook in Experimental Economics, second edition," Princeton University Press, 2016.

_ , Carmit Segal, and Lise Vesterlund, "How costly is diversity? Affirmative action in light of gender differences in competitiveness," *Management Science*, 2013, *59* (1), 1–16.

Perkins, Dwight, "Completing China's move to the market," *The Journal of Economic Perspectives*, 1994, *8* (2), 23–46.

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*, 2008, *123* (3), 1251–1285.

Shi, Li, Song Jin, and Liu Xiaochuan, "Evolution of the gender wage gap among Chinas urban employees," *Social Sciences in china*, 2011, *32* (3), 161–180.

Shu, Xiaoling and Yanjie Bian, "Market transition and gender gap in earnings in urban China," *Social Forces*, 2003, *81* (4), 1107–1145.

Smyth, Russell, Zhai Qingguo, and Wang Jing, "Labour Market Reform in China's State-Owned Enterprises: A Case Study of Post-Deng Fushun in Liaoning Province," *New Zealand Journal of Asian Studies*, 2001, *3*, 42–72.

Solinger, Dorothy J, "Labour market reform and the plight of the laid-off proletariat," *The China Quarterly*, 2002, *170*, 304–326.

Su, Biwei and Almas Heshmati, "Analysis of gender wage differential in China's urban labor market," 2011.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen, "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *American Economic Review*, March 2010, *100* (1), 557–71.

Thaler, Richard H. and Cass R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press, 2008.

UN, Women et al., "Progress of the World's Women 2015-2016: Transforming Economies, Realizing Rights," Technical Report 2015.

Whalley, John and Chunbing Xing, "The regional distribution of skill premia in urban China: Implications for growth and inequality," *International Labour Review*, 2014, *153* (3), 395–419.

Wong, Yen Nee, "World Development Report 2012: Gender equality and development," in "Forum for Development Studies," Vol. 39 Taylor & Francis 2012, pp. 435–444.

Wu, Xiaogang and Yu Xie, "Does the market pay off? Earnings returns to education in urban China," *American Sociological Review*, 2003, pp. 425–442.

Xianguo, Yao, "The Evolution of China Labor Market and Government Behavior [J]," *Journal of Public Management*, 2007, *3*, 003.

Yang, Mayfair Mei-Hui, *Spaces of their own: women's public sphere in transnational China*, Vol. 4, U of Minnesota Press, 1999.

Yao and Xie, "Western theories of discrimination and the problems of discrimination in Chinese labor market," *LEPP WORKING PAPERS SERIES*, 2004, (6), 148–157.

Yearbook, China Labour Statistical, "Zhongguo laodong tongji nianjian," *China Labor Statistical Yearbook*, 1998.

_ , "Zhongguo laodong tongji nianjian," *China Labor Statistical Yearbook*, 2003.

Zeqi, Qiu and Zheng Yongnian, *Xia-Gang and its sociological implications of reducing labour redundancy in China's SOEs*, Singapore University Press & World Scientific Singapore, 1998.

Zhang, Y Jane, "Culture and the gender gap in competitive inclination: Evidence from the Communist experiment in China," 2013.

Appendix A

A.1 Figures

Figure A.1: Histogram of the Covariance of the Responses on the MPL and LS Tasks using Bootstrap MPL Samples.

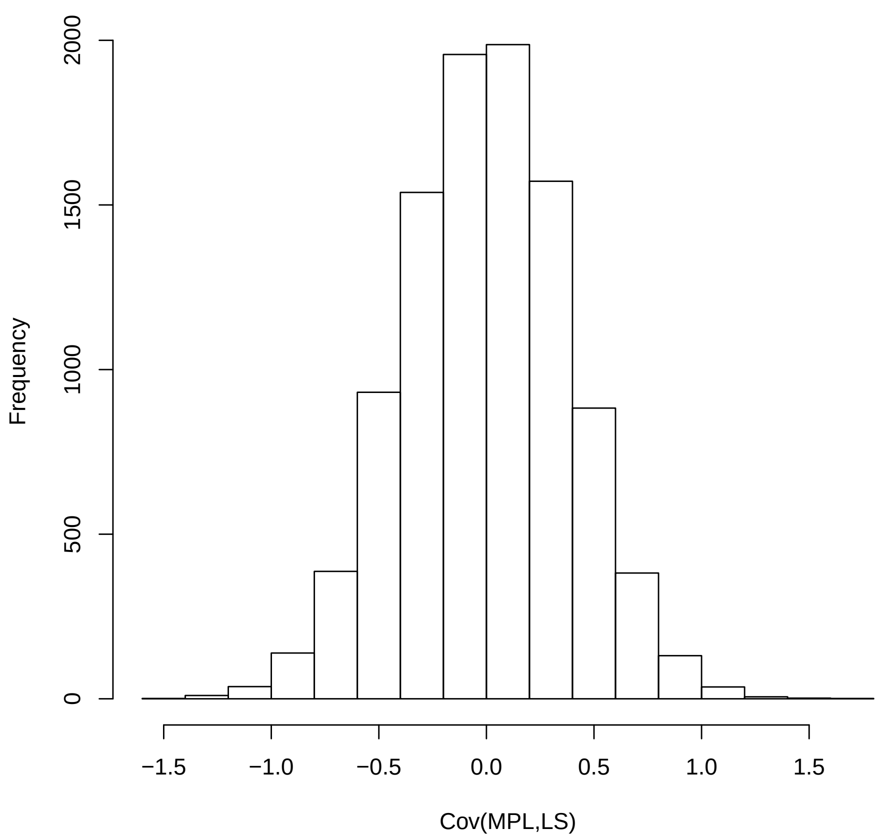
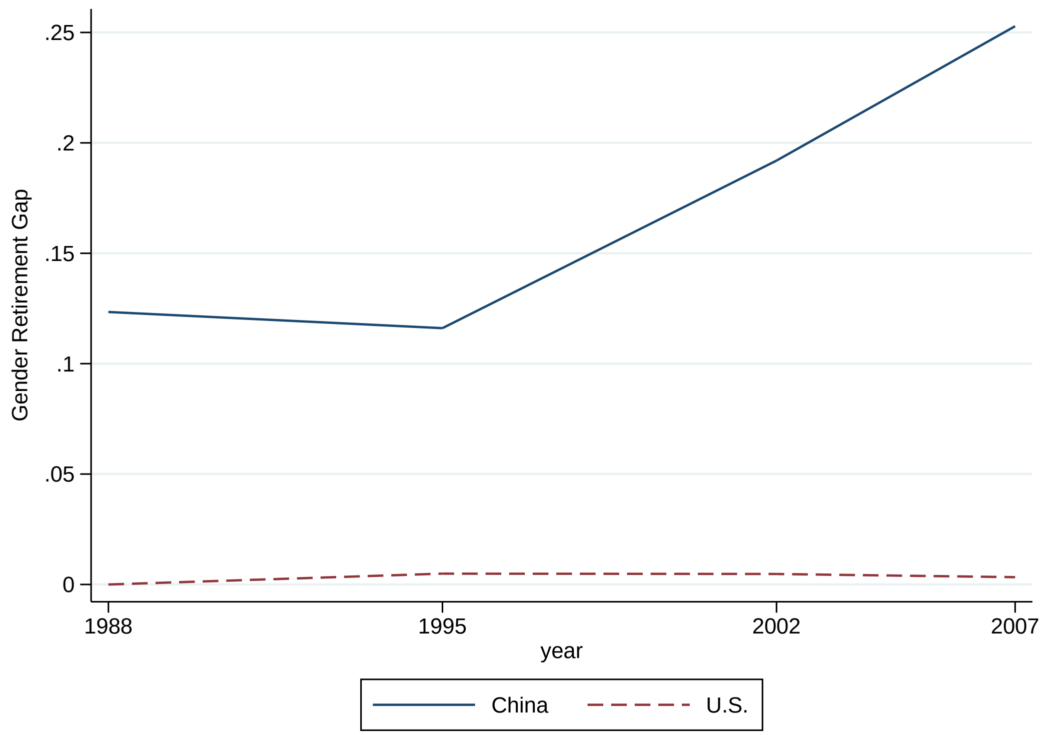


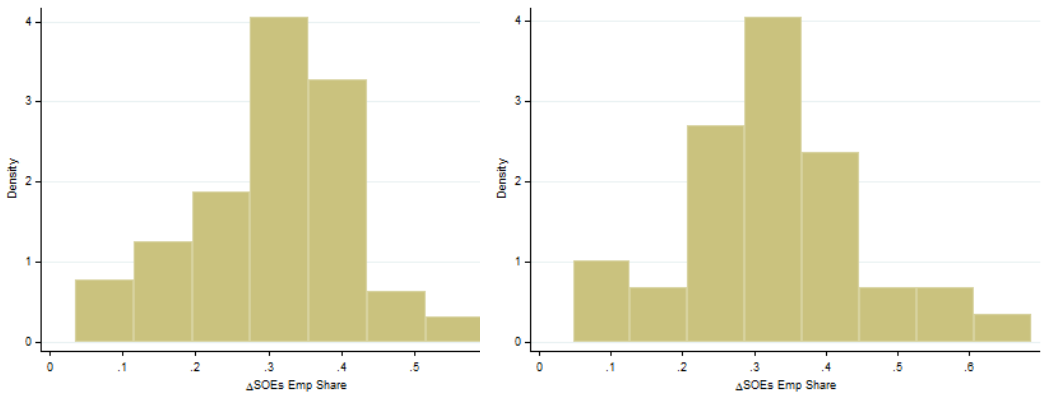
Figure A.2: Change of Gender Early Retirement Gap: China VS U.S.



Notes: (a) China: Data comes from CHIP 1988, 1995, 2002, and 2007. I restrict to those individuals between age 40 and 54 in urban areas. (b) U.S.: Data comes from Annual Social and Economic Supplement of the Current Population Survey. I restrict to those white individuals between age 40 and 54. Gender gap in retirement = (number of retired female individuals/total number of female individuals) - (number of retired male individuals/total number of male individuals).

Figure A.3: Change of SOE Employment Share

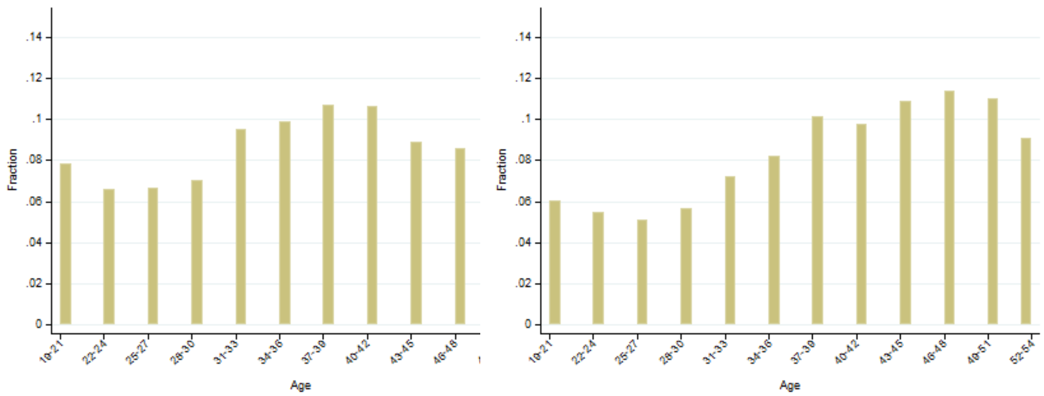
(a) All (b) Subsample



Notes: Bin width: 0.08. (a)Sample includes all 80 prefectures. The mean is 0.31 with standard deviation 0.12. (b) Sample includes 37 prefectures with pre-reform number of SOE workers by industry. The mean is 0.33 with standard deviation 0.13. Δ SOE Emp share = (number of workers in SOE_{1995}/ total number of workers in a given urban prefecture $area_{1995}$) - (number of workers in SOE_{2001}/total number of workers in a given urban prefecture $area_{2001}$). SOE include central SOE, local SOE and collective-owned firms in the urban areas. Data comes from China Provincial Statistical Yearbook and China Labor Statistical Yearbook 1996, 2002.

Figure A.4: Distribution of Different Age Groups

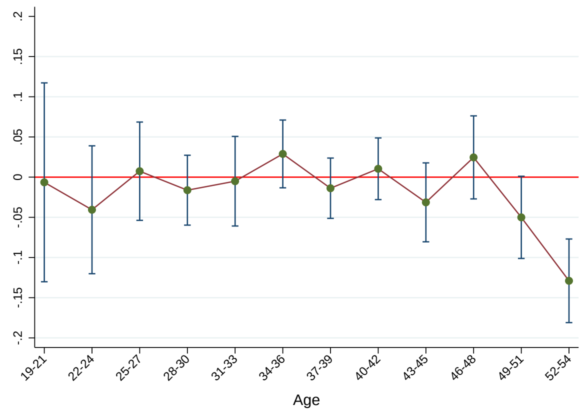
(a) Before (1988 and 1995) (b) After (2002 and 2007)



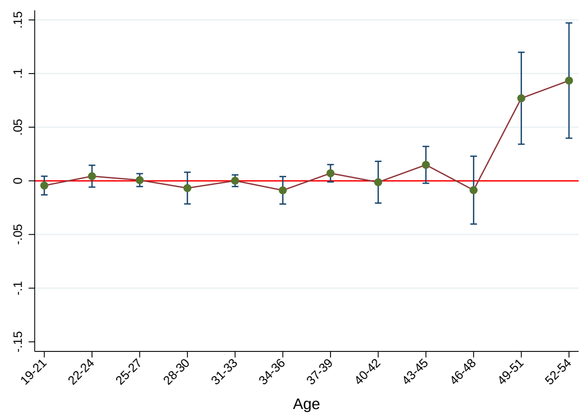
Notes: Bin width: 0.25. Data comes from CHIP 1988, 1995, 2002 and 2007.

Figure A.5: Robustness check: OLS Estimate Coefficients of the Impacts of SOE reform, drop 2007

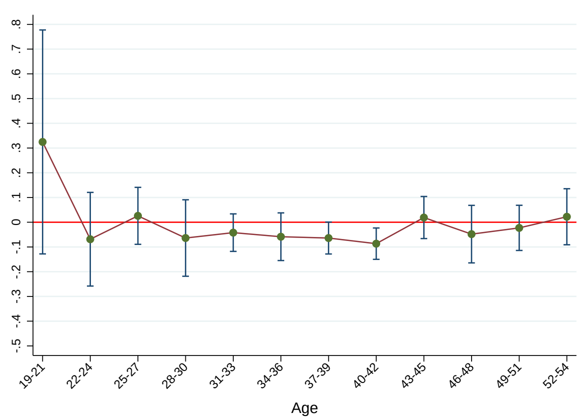
(a) Employment



(b) Retirement



(c) Monthly Earnings



Notes: Regression estimates by Equation 2.6 are plotted. The dot and the bar correspond to the coefficient estimates with 90% confidence intervals.

A.2 Tables

Table A.1: Testing the Balance of Selected Observables by MPL II

	Control		Treatment		P-values
	LS first	MPL first	LS first	MPL first	for
	(1)	(2)	(3)	(4)	H0: (1)=(2)=(3)=(4)
Female	.49	.54	.65	.57	0.451
	(.51)	(.5)	(.48)	(.5)	
Age	14.23	14.55	14.3	14.38	0.568
	(1.3)	(1.11)	(.99)	(1.02)	
Number of family members in the household	5.17	5.5	5.33	6.09	0.291
	(2.11)	(2)	(2.13)	(3.29)	
Distance from home to school (=1 if less than or equal to 30min walk)	.43	.33	.41	.46	0.589
	(.5)	(.47)	(.5)	(.5)	
Mother's educational attainment (=1 if less than or equal to primary)	.65	.71	.61	.76	0.418
	(.48)	(.46)	(.49)	(.43)	
Mother's occupation (=1 if agricultural)	.73	.69	.78	.76	0.766
	(.45)	(.47)	(.42)	(.43)	
Monthly allowance (=1 if less than or equal to 300RMB)	.8	.85	.85	.72	0.346
	(.41)	(.36)	(.36)	(.46)	
Test score	.06	.1	-.01	-.17	0.455
	(.82)	(.83)	(.86)	(.99)	
Observations	49	52		46	46

Notes: Means and standard deviations are presented. Standard deviations in parentheses. Exchange rate: 1RMB = 0.16 US dollars.

Table A.2: Cognitive Ability and MSB in the MPL - Control group

Dep.Var.= 1 if multiple switcher in MPL task

	(1)	(2)	(3)	(4)
Math test score	-0.091**	-0.112**		
	(0.044)	(0.048)		
Verbal test score			-0.117**	-0.125**
			(0.049)	(0.056)
Other controls	No	Yes	No	Yes
N	99	95	99	95

Notes: Students' test scores are standardized within each grade. Robust standard errors are in parentheses. Other controls include: gender, monthly household income, mother's educational attainment, mother's occupation, number of children in the household, and grade fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.3: Cognitive Ability and MSB in the MPL - Treatment group

Dep.Var.= 1 if multiple switcher in MPL task

	(1)	(2)	(3)	(4)
Math test score	-0.024	-0.035		
	(0.031)	(0.033)		
Verbal test score			-0.033	-0.045
			(0.021)	(0.028)
Other controls	No	Yes	No	Yes
N	92	87	92	87

Notes: Students' test scores are standardized within each grade. Robust standard errors are in parentheses. Other controls include: gender, monthly household income, mother's educational attainment, mother's occupation, number of children in the household, and grade fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.4: Change of SOE Employment Share and Bartik Shift-share Intensity Index

	Δ SOE Emp Share$_p$		Bartik Intensity Index$_p$
	All	Subsample	Subsample
Mean	0.31	0.33	0.19
St.Dev.	0.12	0.13	0.03
Min	0.04	0.05	0.13
Max	0.65	0.65	0.23
Number of prefectures	80	37	37

Notes: Author's calculation. Data comes from China Provincial Statistical Yearbook, China Labor Statistical Yearbook, China Statistical Yearbook 1996, 2002. p denotes prefecture.

Table A.5: Gender Differences in Rank in Male Earnings Distribution, Quantile Regressions-1988-2007(including non-employed individuals)

	1988	1995	2002	2007
10th Quantile	-6.940***	-4.407***	-3.673***	-1.098***
	(0.285)	(0.190)	(0.200)	(0.978)
25th Quantile	-8.451***	-7.383***	-8.533***	-12.359***
	(0.396)	(0.437)	(0.323)	(1.678)
50th Quantile	-9.483***	-7.378***	-11.255***	-19.104***
	(0.412)	(0.581)	(0.560)	(1.205)
75th Quantile	-9.940***	-6.823***	-10.680***	-17.298***
	(0.460)	(0.590)	(0.691)	(1.091)
90th Quantile	-8.489***	-5.920***	-7.223***	-12.054***
	(0.522)	(0.587)	(0.676)	(0.976)
Obs.	17,255	12,342	12,457	8,072

Notes: Each main cell of the table reports the coefficient from 10th, 25th, 50th, 75th, and 90th quantile regressions of the individual's percentile rank in the male earnings distribution on female dummy variable. Additional controls include education categories, age categories, and prefecture fixed effects. Sample includes all individuals between age 19 and 54. Monthly earnings are deflated at the 2014 price level. Robust standard errors are clustered at the prefecture level.

Table A.6: Reduced Form Estimates of the Impacts of SOE Reform

	Employed	Retired	ln(Monthly earnings)
	(1)	(2)	(3)
Female × after × Bartik intensity, b1	-.031***	.029***	-.038**
	(.011)	(.009)	(.019)
Female × after, b2	-.105***	.064***	-.119***
	(.010)	(.009)	(.018)
Female, b3	-.049***	.059***	-.097***
	(.006)	(.004)	(.011)
Female × Bartik intensity, b4	.004	-.003	.010
	(.006)	(.005)	(.011)
After × Bartik intensity, b5	-.013	.005	.0004
	(.008)	(.005)	(.037)
Obs.	25450	25494	19265

Notes: Sample includes all individuals between ages 19 and 54. Monthly earnings are deflated at the 1988 price level. All regressions control for age, age squared, years of schooling, working experience, working experience squared, ethnicity, prefecture dummies, year dummies, and prefecture specific time trend. Column(3) include individuals who report they currently have a full-time job and also control for working industries. Bartik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.7: OLS Estimates of the Impacts of SOE Reform, by Demographic Group (full sample)

	Age<=40			Age>40		
Dependent Variable	Employed	Retired	ln(Monthly earnings)	Employed	Retired	ln(Monthly earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled						
Female × after × Δemp share,β_1	-.002	.0007	-.039***	-.035***	.029***	-.019
	(.007)	(.001)	(.015)	(.014)	(.007)	(.027)
Female × after, β_2	-.058***	-.0009	-.122***	-.120***	.095***	-.097***
	(.008)	(.002)	(.019)	(.015)	(.013)	(.024)
Female, β_3	.005	.0009	-.089***	-.142***	.130***	-.169***
	(.004)	(.0007)	(.010)	(.007)	(.007)	(.015)
Obs.	27096	27166	22338	20426	20481	15023
Panel B: Edu< High School						
Female × after × Δemp share,β_1	-.015	-.0008	-.061**	-.044***	.035***	-.037
	(.015)	(.004)	(.024)	(.016)	(.011)	(.035)
Female × after, β_2	-.076***	-.0006	-.158***	-.125***	.098***	-.068**
	(.015)	(.004)	(.030)	(.017)	(.017)	(.029)
Female, β_3	-.004	.0006	-.107***	-.188***	.164***	-.217***
	(.004)	(.0008)	(.011)	(.009)	(.010)	(.022)
Obs.	15097	15161	12864	13389	13443	9182
Panel C: Edu>= High School						
Female × after × Δemp share, β_1	.003	.001	-.032*	-.004	.012	.010
	(.008)	(.002)	(.018)	(.010)	(.008)	(.025)
Female × after, β_2	-.049***	-.002	-.115***	-.107***	.105***	-.144***
	(.011)	(.002)	(.022)	(.015)	(.010)	(.031)
Female, β_3	.017**	.002	-.059***	-.061***	.060***	-.090***
	(.008)	(.002)	(.016)	(.010)	(.009)	(.010)
Obs.	11999	12005	9474	7037	7038	5841

Notes: Individuals between ages 19 and 54. Each model also includes age, age squared, years of schooling, ethnic minority, year fixed effects, prefecture fixed effects, and prefecture specific time trend. Column (3) and column(6) include those individuals who report they currently have a full-time job and also control for working industry. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.8: 2sls Estimates of the Impacts of SOE Reform, by Demographic Group (Subsample)

	Age<=40			Age>40		
Dependent Variable	Employed	Retired	ln(Monthly earnings)	Employed	Retired	ln(Monthly earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled						
Female × after × Δemp share, β_1	-.013	.001	-.201**	-.121***	.119**	.009
	(.030)	(.004)	(.101)	(.045)	(.050)	(.075)
Female × after, β_2	-.052***	-.003	-.099**	-.079***	.059**	-.135***
	(.019)	(.003)	(.042)	(.025)	(.025)	(.030)
Female, β_3	.015***	.002***	-.076***	-.142***	.138***	-.133***
	(.005)	(.0007)	(.015)	(.009)	(.010)	(.017)
Obs.	13935	13954	11228	11515	11540	8199
Panel B: Edu< High School						
Female × after × Δemp share, β_1	.007	.002	-.308*	-.147***	.145**	.010
	(.046)	(.009)	(.168)	(.055)	(.064)	(.084)
Female × after, β_2	-.091***	-.002	-.120*	-.076**	.053	-.094**
	(.029)	(.004)	(.062)	(.033)	(.032)	(.045)
Female, β_3	.007	.002**	-.087***	-.193***	.179***	-.164***
	(.006)	(.0008)	(.015)	(.014)	(.015)	(.027)
Obs.	7154	7172	6109	7296	7320	4798
Panel C: Edu>= High School						
Female × after × Δemp share, β_1	-.027	-.002	-.198**	-.058	.057	-.003
	(.028)	(.005)	(.095)	(.049)	(.039)	(.089)
Female × after, β_2	-.039*	-.004	-.101**	-.088***	.093***	-.185***
	(.021)	(.003)	(.043)	(.025)	(.018)	(.029)
Female, β_3	-.008	.003	.113**	-.012	.002	-.003
	(.013)	(.004)	(.048)	(.027)	(.025)	(.031)
Obs.	6781	6782	5119	4219	4220	3401

Notes: Individuals between ages 19 and 54. Each model also includes age, age squared, years of schooling, ethnic minority, year fixed effects, prefecture fixed effects, and prefecture specific time trend. Column (3) and column(6) include those individuals who report they currently have a full time job and also control for working industry. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. Bartik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.9: Estimates of the Impacts of SOE Reform, drop Guangdong province

	OLS			IV		
Dependent variable	Employed	Retired	ln(Monthly earnings)	Employed	Retired	ln(Monthly earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Female × after × Δemp share, β_1	-.016**	.015***	-.039**	-.063**	.061**	-.074*
	(.007)	(.004)	(.017)	(.027)	(.026)	(.041)
Female × after, β_2	-.110***	.064***	-.116***	-.082***	.047***	-.121***
	(.010)	(.008)	(.016)	(.017)	(.015)	(.018)
Female, β_3	-.054***	.052***	-.125***	-.053***	.061***	-.101***
	(.004)	(.003)	(.011)	(.005)	(.005)	(.013)
Obs.	42098	42203	33367	22249	22288	16824

Notes: Individuals between ages 19 and 54. Monthly earnings are deflated at the 2014 year level. column (3) includes individuals who report they currently have a full time job. All models include age, age squared, years of schooling, ethnic minority, prefecture fixed effects, year fixed effects, and prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. Batik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.10: OLS Estimates of the Impacts of SOE Reform (young cohort, age<=40)

Dependent variable	ln(Monthly earnings) (1)	(2)	Work in private sectors (3)
Female $\times$ after $\times$ Δemp share, β_1	-.041**	-.048*	.024***
	(.017)	(.026)	(.006)
Female $\times$ after, β_2	-.121***	-.088***	.005
	(.023)	(.026)	(.010)
Female, β_3	-.094***	-.080***	-.004*
	(.016)	(.010)	(.002)
Weekly working hours	Yes		
Communist party membership		Yes	
Obs.	12962	19502	21306

Notes: Individuals between ages 19 and 40 who report they currently have a full-time job. Monthly earnings are deflated at the 2014 year level. Column (1) include 1995 and 2002 waves because only these two waves ask the weekly working hours question. Column (2) includes 1988, 1995, and 2002 waves because 2007 wave did not ask the communist party membership question. All models include age, age squared, years of schooling, ethnic minority, prefecture fixed effects, year fixed effects. Column (2) includes prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.11: OLS Estimates of the Impacts of SOE Reform (full sample), alternative definition of employment

Dependent variable	Employed	ln(Monthly earnings)
	(1)	(2)
Female × after × Δemp share, β_1	-.017**	-.037**
	(.007)	(.017)
Female × after, β_2	-.115***	-.121***
	(.009)	(.014)
Female, β_3	-.060***	-.122***
	(.003)	(.010)
Female × Δemp share, β_4	.014***	.004
	(.005)	(.020)
After × Δemp share, β_5	.005	-.011
	(.011)	(.027)
Obs.	47661	38754

Notes: Individuals between ages 19 and 54. Monthly earnings are deflated at the 2014 year level. Employed equals to 1 if individuals report they currently have a full time job or self-employed, and 0 otherwise. Column(2) includes individuals who report they currently have a full-time job or are self-employed. All models include age, age squared, years of schooling, ethnic minority, prefecture fixed effects, and year fixed effects. Column (2) includes prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.12: DID Estimates of the Impacts of SOE Reform on Employment and Retirement (old cohort, age >=40), additional controls

Dependent variable	Employed				Retired			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female $\times$ after $\times$ Δemp share, β_1	-.036**	-.034**	-.035**	-.034**	.030***	.029***	.030***	.029***
	(.014)	(.014)	(.014)	(.014)	(.008)	(.008)	(.008)	(.008)
Female $\times$ after, β_2	-.123***	-.119***	-.123***	-.120***	.092***	.090***	.092***	.090***
	(.015)	(.015)	(.015)	(.015)	(.013)	(.013)	(.013)	(.013)
Female, β_3	-.130***	-.131***	-.130***	-.131***	.120***	.120***	.120***	.120***
	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)	(.006)
Household income		Yes		Yes		Yes		Yes
Child under age 6			Yes	Yes			Yes	Yes
Adj. R-squared	0.2663	0.2724	0.2671	0.2734	0.2540	0.2583	0.2541	0.2586
Obs.	22069	22062	22069	22062	22126	22119	22126	22119

Notes: Individuals between ages 40 and 54. All models include age, age squared, years of schooling, ethnic minority, prefecture fixed effects, year fixed effects and prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.13: Estimates of the Impact of SOE Reform on Retirement, by Intensity of Male-to-Female Sex Ratio

	DID		IV	
	Low(0.99-1.07)	High(1.08-1.26)	Low(0.99-1.07)	High(1.08-1.26)
	(1)	(2)	(3)	(4)
Female $\times$ after $\times$ Δemp share, β_1	.014***	.015**	.051	.068*
	(.004)	(.006)	(.042)	(.039)
Female $\times$ after, β_2	.043***	.068***	.040**	.041**
	(.008)	(.011)	(.017)	(.020)
Female, β_3	.050***	.056***	.059***	.061***
	(.004)	(.005)	(.007)	(.004)
Obs.	22473	25174	11144	14350

Notes: Dependent variable is retirement. Individuals between age 19 and 54 . First two columns include full sample. Column (3) and column (4) include 37 prefectures with pre-reform industry employment composition. Low refers to low male/female sex ratio at birth; High refers to high male-to-female sex ratio at birth. All models include age, age squared, years of schooling, ethnic minority, working industries, year fixed effects, prefecture fixed effects and prefecture specific time trend. Sex ratio at birth is calculated by using the 1990 census, I restrict to those individuals who are under age 10. The mean of sex ratio at birth is 1.09 with standard deviation 0.06. Reported robust standard errors are clustered at the prefecture level. Change of SOE employment share has been standardized to have mean equal to 0 and standard deviation equal to 1. Bartik intensity has been standardized to have mean equal to 0 and standard deviation equal to 1. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.14: Estimates of the Impact of SOE Reform on Earnings Rank Positions, by Intensity of Male-to-Female Sex Ratio(including non-employed individuals)

	Overall	Low(0.99-1.07)	High(1.08-1.26)
	(1)	(2)	(3)
Female × after × Δemp share, β_1	-1.531	-1.487	-1.596
	(.961)	(1.353)	(1.235)
Female × after, β_2	-1.147	.215	-2.389**
	(.787)	(.913)	(1.084)
Female, β_3	-12.277***	-12.847***	-11.626***
	(.580)	(1.002)	(.453)
Female × Δemp share, β_4	.113	.241	-.135
	(1.015)	(1.712)	(.537)
After × Δemp share, β_5	-1.092	-1.753	-.612
	(.976)	(1.865)	(1.416)
Obs.	46174	22409	23765

Notes: Individuals include those between age 19 and 54. All models include age, age squared, years of schooling, and ethnicity. All models also include prefecture fixed effects, year fixed effects and prefecture specific time trend. Reported robust standard errors are clustered at the prefecture level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix B

B.1 Proofs

B.1.1 Conceptual Framework

Scenario II In the general case under Scenario II, the nudge treatment unconfuses a proper subset of individuals. To describe this scenario we introduce a third type. Type 3 individuals, occuring with probability p_3, are confused using the MPL even when nudged. Type 1 and Type 2 occur with probability p_1 and p_2, respectively, with $p_1 + p_2 + p_3 = 1$.

Under this general Scenario II, for the control group using the MPL, the response of Type 1 individuals is $X_{1s} = S + \eta_{1s}$; the response of Type 2 individuals is $X_{2s} = \eta_{2s}$; and the response of Type 3 individuals is $X_{3s} = \eta_{3s}$, $\eta_{2s} \overset{d}{=} \eta_{3s}$. For the control group using the LS task, because the confusion is specific to the MPL, the response of Type 1, Type 2 and Type 3 individuals are $Y_{1s} = S + \nu_{1s}$, $Y_{2s} = S + \nu_{2s}$, and $Y_{3s} = S + \nu_{3s}$, respectively, where $\nu_{1s} \overset{d}{=} \nu_{2s} \overset{d}{=} \nu_{3s}$.

For the treatment group using the MPL, the response of Type 1 individuals is $X_{1n} = S + \eta_{1n}$; the response of Type 2 individuals is $X_{2n} = S + \eta_{2n}$; and the response of Type 3 individuals is $X_{3n} = \eta_{3n}$. $\eta_{1n} \overset{d}{=} \eta_{2n} \overset{d}{=} \eta_{1s}$; $\eta_{3n} \overset{d}{=} \eta_{2s} \overset{d}{=} \eta_{3s}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status, for all three types of individuals. That is to say, $Y_{1n} = S + \nu_{1n}$, $Y_{2n} = \nu_{2n}$, and $Y_{3n} = \nu_{3n}$, where $\nu_{1n} \overset{d}{=} \nu_{1s}$, $\nu_{2n} \overset{d}{=} \nu_{2s}$, and $\nu_{3n} \overset{d}{=} \nu_{3s}$.

Then, for the control group, $Cov(MPL_s, LS_s) = p_1 Cov(X_{1s}, Y_{1s}) + p_2 Cov(X_{2s}, Y_{2s}) + p_3 Cov(X_{3s}, Y_{3s}) = p_1 \sigma_s^2$, and for treatment group, $Cov(MPL_n, LS_n) = (p_1 + p_2)\sigma_s^2$. This yields the result that $Cov(MPL_n, LS_n) > Cov(MPL_s, LS_s)$ if $p_2 > 0$.

Scenario III In the general case under Scenario III, the nudge treatment unconfuses a proper subset of individuals. As in the section above, we introduce a third type. Type 3 individuals, occuring with probability p_3, are confused using the MPL even when nudged. Type 1 and Type 2 occur with probability p_1 and p_2, respectively, with $p_1 + p_2 + p_3 = 1$.

Under this general Scenario III, for the control group using the MPL, the response of Type 1 individuals is $X_{1s} = S + \eta_{1s}$; the response of Type 2 individuals is $X_{2s} = \eta_{2s}$; and the response of Type 3 individuals is $X_{3s} = \eta_{3s}$. $\eta_{2s} \stackrel{d}{=} \eta_{3s}$.

For the control group using the LS task, the response of Type 1 individuals is $Y_{1s} = S + \nu_{1s}$; the response of Type 2 individuals is $Y_{2s} = \nu_{2s}$; and the response of Type 3 individuals is $Y_{3s} = \nu_{3s}$. $\nu_{2s} \stackrel{d}{=} \nu_{3s}$.

For the treatment group using the MPL, the response of Type 1 individuals is $X_{1n} = S + \eta_{1n}$; the response of Type 2 individuals is $X_{2n} = S + \eta_{2n}$; and the response of Type 3 individuals is $X_{3n} = \eta_{3n}$. $\eta_{1n} \stackrel{d}{=} \eta_{2n} \stackrel{d}{=} \eta_{1s}$; $\eta_{3n} \stackrel{d}{=} \eta_{2s} \stackrel{d}{=} \eta_{3s}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status. For the treatment group using the LS task, the response of Type 1 individuals is $Y_{1n} = S + \nu_{1n}$; the response of Type 2 individuals is $Y_{2n} = \nu_{2n}$; and the response of Type 3 individuals is $Y_{3n} = \nu_{3n}$. $\nu_{1n} \stackrel{d}{=} \nu_{1s}$, $\nu_{2n} \stackrel{d}{=} \nu_{2s}$, and $\nu_{3n} \stackrel{d}{=} \nu_{3s}$.

For the control group, we have $Cov(MPL_s, LS_s) = \sum_{j=1}^{3} p_i Cov(X_{is}, Y_{is}) + p_1\mu_{X1s}\mu_{Y1s} + p_2\mu_{X2s}\mu_{Y2s} + p_3\mu_{X3s}\mu_{Y3s} - (\mu_{Xs})(\mu_{Ys})$, while for the treatment group, we have $Cov(MPL_n, LS_n) = \sum_{j=1}^{3} p_i Cov(X_{in}, Y_{in}) + p_1\mu_{X1n}\mu_{Y1n} + p_2\mu_{X2n}\mu_{Y2n} + p_3\mu_{X3n}\mu_{Y3n} - (\mu_{Xn}\mu_{Yn})$.

Note that $\sum_{i=1}^{3} p_i Cov(X_{is}, Y_{is}) = \sum_{i=1}^{3} p_i Cov(X_{in}, Y_{in}) = p_1\sigma_s^2$, so we can rearrange to obtain $Cov(MPL_s, LS_s) = Cov(MPL_n, LS_n) + p_2\mu_{X2s}\mu_{Y2s} - p_2\mu_{X2n}\mu_{Y2n} + \mu_{Xn}\mu_{Yn} - \mu_{Xs}\mu_{Ys} = Cov(MPL_n, LS_n) + p_2\mu_{Y2s}(\mu_{X2s} - \mu_{X2n}) - \mu_{Ys}(\mu_{Xs} - \mu_{Xn}) = Cov(MPL_n, LS_n) + p_2\mu_{Y2s}(\mu_{X2s} - \mu_{X2n}) - p_2\mu_{Ys}(\mu_{X2s} - \mu_{X2n}) = Cov(MPL_n, LS_n) + p_2(\mu_{X2s} - \mu_{X2n})(\mu_{Y2s} - \mu_{Ys})$. Thus, as long as the difference in the expected value of measured risk tolerance in the control group for Type 1 and Type 2 is not in the opposite direction for the two tasks, i.e., either $E(X_{1s}) - E(X_{2s})$ and $E(Y_{1s}) - E(Y_{2s})$ have the same sign, or one or both of the differences is 0, then $(\mu_{X2s} - \mu_{X2n})(\mu_{Y2n} - \mu_{Yn}) \geq 0$, and we have $Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$.

B.1.2 Test Statistic

To test the explanations in our conceptual framework, we need to compare the covariances between responses on the MPL and LS tasks for the control group which uses the standard protocol and the treatment group which uses the nudge protocol. To the best of our knowledge, covariance comparison tests have not been previously considered. Tests for correlations cannot be used because the variances in our case are indeterminate.

For notational convenience, let $M_s = MPL_s$ and $L_s = LS_s$ for the control group, and let $M_n = MPL_n$ and $L_n = LS_n$ for the treatment group.

Denote the sample covariance between M_s and L_s by

$$S_{M_sL_s} = \frac{1}{n_s - 1}\sum_{i=1}^{n_s}\Big[(M_{s_i} - \overline{M}_s)(L_{s_i} - \overline{L}_s)\Big], \tag{B.1}$$

where $\overline{M}_s$ is the average of the data $\{M_{s_i}, i = 1, \ldots, n_s\}$ for MPL, and $\overline{L}_s$ is defined for LS in a similar way.

The sample covariance (B.1) is used because it is unbiased, i.e. $E(S_{M_sL_s}) = Cov(M_s, L_s) = \sigma_{M_sL_s}$. It is easy to verify, so we omit the proof here. In what follows, we derive the variance of $S_{M_sL_s}$ to construct our test.

First, consider

$$\begin{aligned} E\Big[(n_s-1)^2 S^2_{M_sL_s}\Big] &= E\Big[\sum_{i,j=1}^{n_s}\big[(M_{s_i} - \overline{M}_s)(L_{s_i} - \overline{L}_s)(M_{s_j} - \overline{M}_s)(L_{s_j} - \overline{L}_s)\big]\Big] \\ &= E\Big[\sum_{i,j=1}^{n_s}\big[(M_{s_i} - \mu_{M_s})(L_{s_i} - \mu_{L_s})(M_{s_j} - \mu_{M_s})(L_{s_j} - \mu_{L_s})\big]\Big] \\ &\quad - 2n_s E\Big[(\overline{M}_s - \mu_{M_s})(\overline{L}_s - \mu_{L_s})\sum_{j=1}^{n_s}(M_{s_j} - \mu_{M_s})(L_{s_j} - \mu_{L_s})\Big] \\ &\quad + n_s^2 E\Big[(\overline{M}_s - \mu_{M_s})^2(\overline{L}_s - \mu_{L_s})^2\Big], \end{aligned} \tag{B.2}$$

where $\mu_{M_s} = E(M_s)$ and $\mu_{L_s} = E(L_s)$.

For the last term on the right hand side in (B.2), we have

$$\begin{aligned} n_s^2 E\Big[(\overline{M}_s - \mu_{M_s})^2(\overline{L}_s - \mu_{L_s})^2\Big] &= \frac{1}{n_s^2} E\Big[\sum_{i,j,k,l=1}^{n_s}\big[(M_{s_i} - \mu_{M_s})(M_{s_j} - \mu_{M_s})(L_{s_l} - \mu_{L_s})(L_{s_k} - \mu_{L_s})\big]\Big] \\ &= \frac{1}{n_s^2}\sum_{i=1}^{n_s} E\Big[(M_{s_i} - \mu_{M_s})^2(L_{s_i} - \mu_{L_s})^2\Big] + \frac{n_s - 1}{n_s}\sigma^2_{M_s}\sigma^2_{L_s} + \frac{2(n_s-1)}{n_s}\sigma^2_{M_sL_s}, \end{aligned} \tag{B.3}$$

where $\sigma^2_{M_s} = Var(M_s)$ and $\sigma^2_{L_s} = Var(L_s)$.

Similarly, for the first and second terms on the right in (B.2), we have

$$E\Big[\sum_{i,j=1}^{n_s}\big[(M_{s_i}-\mu_{M_s})(L_{s_i}-\mu_{L_s})(M_{s_j}-\mu_{M_s})(L_{s_j}-\mu_{L_s})\big]\Big]=\sum_{i=1}^{n_s}E\Big[(M_{s_i}-\mu_{M_s})^2(L_{s_i}-\mu_{L_s})^2\Big]+n_s(n_s-1)\sigma^2_{M_sL_s}, \tag{B.4}$$

and

$$\begin{aligned}&-2n_sE\Big[(\overline{M}_s-\mu_{M_s})(\overline{L}_s-\mu_{L_s})\sum_{j=1}^{n_s}(M_{s_j}-\mu_{M_s})(L_{s_j}-\mu_{L_s})\Big]\\ =&\frac{-2}{n_s}\sum_{i=1}^{n_s}E\Big[(M_{s_i}-\mu_{M_s})^2(L_{s_i}-\mu_{L_s})^2\Big]-2(n_s-1)\sigma^2_{M_sL_s}.\end{aligned} \tag{B.5}$$

Substituting (B.3)-(B.5) into (B.2) and then dividing both sides by $(n_s-1)^2$ yields

$$E\Big[S^2_{M_sL_s}\Big]=\frac{1}{n_s}E\big[(M_s-\mu_{M_s})^2(L_s-\mu_{L_s})^2\big]+\frac{1}{n_s(n_s-1)}\sigma^2_{M_s}\sigma^2_{L_s}+\frac{(n_s-1)^2+1}{n_s(n_s-1)}\sigma^2_{M_sL_s}, \tag{B.6}$$

where $E\big[(M_s-\mu_{M_s})^2(L_s-\mu_{L_s})^2\big]$ is the common value of $E\Big[(M_{s_i}-\mu_{M_s})^2(L_{s_i}-\mu_{L_s})^2\Big]$ for $i=1,\dots,n_s$.

Finally, the variance of $S_{M_sL_s}$ is given by

$$Var(S_{M_sL_s})=\frac{1}{n_s}Var(W_s)+\frac{\sigma^2_{M_s}\sigma^2_{L_s}}{n_s(n_s-1)}(\rho^2_{M_sL_s}+1), \tag{B.7}$$

where $W_s=(M_s-\mu_{M_s})(L_s-\mu_{L_s})$ and $\rho_{M_sL_s}$ is the correlation between M_s and L_s.

By the central limit theorem and the independence of standard and nudge groups, we have the result that

$$\frac{(S_{M_sL_s}-S_{M_nL_n})-(\sigma_{M_sL_s}-\sigma_{M_nL_n})}{\sqrt{Var(S_{M_sL_s})+Var(S_{M_nL_n})}} \tag{B.8}$$

will have a limiting standard normal distribution.

Note that $Var(S_{M_sL_s})$ and $Var(S_{M_nL_n})$ are unknown, and need to be estimated in order to give a p-value for our test. We use

$$\frac{1}{n_s}\widehat{Var}(W_s)+\frac{\sigma^2_{M_s}\sigma^2_{L_s}}{n_s(n_s-1)}(\hat{\rho}^2_{M_sL_s}+1) \tag{B.9}$$

to estimate $Var(S_{M_sL_s})$, where

$$\widehat{Var}(W_s)=\frac{1}{n_s}\sum_{i=1}^{n_s}(W_{s_i}-S_{M_sL_s})^2-(\overline{W}_s-S_{M_sL_s})^2,$$

$W_{s_i}=(M_{s_i}-\overline{M}_s)(L_{s_i}-\overline{L}_s)$ for $i=1,\dots,n_s$, $\overline{W}_s=\frac{1}{n_s}\sum_{i=1}^{n_s}W_{s_i}=\frac{n-1}{n}S_{M_sL_s}$, and $\hat{\rho}_{M_sL_s}$ is the sample Pearson's correlation. Similar estimates are used for $Var(S_{M_nL_n})$.

B.2 Experimental Protocol

Verbal Instructions (Translation) – Control Group

Thank you for participating in our game today!

First, we want you to understand that this game has no influence on your academic performance, and there is no right or wrong answer. Your choice is only related to your personal preference. How much cash you can earn in the end essentially depends on your choices and luck.

Do you have any questions?

Now, we are going to explain these two games. You are going to play both of them. After that, we randomly choose one to realize your payment. (We are going to toss a coin, if the heads shows up, you will get your payment based on the results of game one; if the tail shows up, and you will get your payment based on the results of game two.)

Since your ID number is odd, you will play game one first. (If your id is even, you will play game two first). Please read the instructions carefully, and write your choice on the paper. In this game, there are six lotteries; you are allowed to choose one of them. For each lottery, there is 50/50 chance to win each payoff. For example, if you choose lottery 2 and this game is chosen for payout at the end, and the ping pang ball you draw is yellow, you will get 5 yuan.

Now, we are going to play game two and I will explain the instructions now. In this game, there are six groups and you will need to make a choice in each group. There are two choices – A and B—in each group, your task is to pick either A or B. If you pick A, it means you would like to enter into a lottery, there is a 50% probability you will win 10 Yuan and 50% probability you will get nothing. If you choose B, it means you don’t want to enter the lottery and would like to get the cash directly. If this game is chosen for payout at the end, we will randomly choose one of these six groups and realize your payment based on your choice (A or B). Now, you can read these six groups one by one, if you would like to choose A, put a mark in the blank; if you would like to choose B, put a mark on the other blank.

Verbal Instructions (Translation) – Treatment Group

Thank you for participating in our game today!

First, we want you to understand that this game has no influence on your academic performance, and there is no right or wrong answer. Your choice is only related to your personal preference. How much cash you can earn in the end essentially depends on your choices and luck.

Do you have any questions?

Now, we are going to explain these two games. You are going to play both of them. After that, we randomly choose one to realize your payment. (We are going to toss a coin, if the heads shows up, you will get your payment based on the results of game one; if the tail shows up, and you will get your payment based on the results of game two.)

Since your ID number is odd, you will play game one first. (If your id is even, you will play game two first). Please read the instructions carefully, and write your choice on the paper. In this game, there are six lotteries; you are allowed to choose one of them. For each lottery, there is 50/50 chance to win each payoff. For example, if you choose lottery 2 and this game is chosen for payout at the end, and the ping pang ball you draw is yellow, you will get 5 yuan.

Now, we are going to play game two and I will explain the instructions now. In this game, there are six groups and you will need to make a choice in each group. There are two choices – A and B—in each group, your task is to pick either A or B. If you pick A, it means you would like to enter into a lottery, there is a 50% probability you will win 10 Yuan and 50% probability you will get nothing. If you choose B, it means you don't want to enter the lottery and would like to get the cash directly. If this game is chosen for payout at the end, we will randomly choose one of these six groups and realize your payment based on your choice (A or B). Now, you can read these six groups one by one, if you would like to choose A, put a mark in the blank; if you would like to choose B, put a mark on the other blank.

Have you decided? You can think about your choices again carefully and can change your choices. If you would like, we can explain this game one more time.

ID: _______________

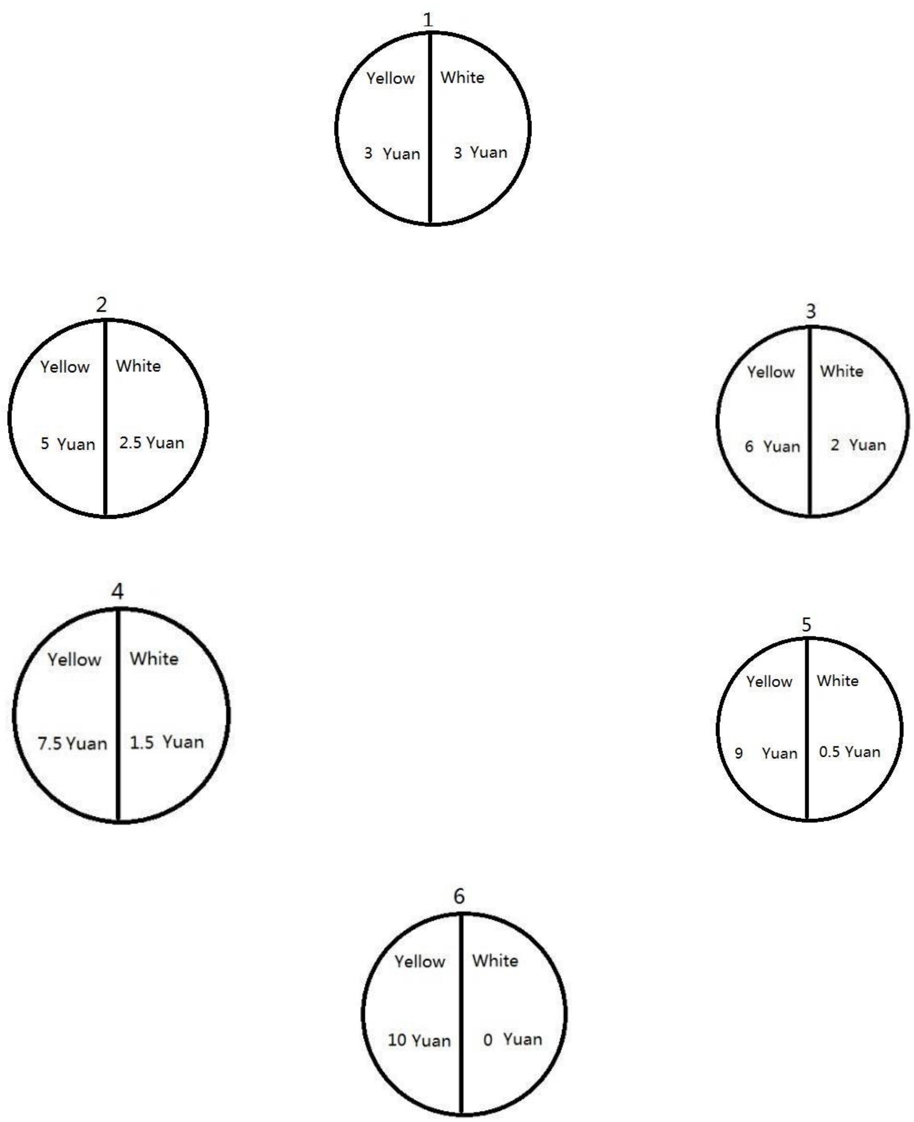
Experimental Instructions (Translation)

Hello everyone, welcome to today's game. Today's game has no impact whatsoever on your academic performance. Depending on your choices in the game, you will have a chance to win a cash prize, so please make your decisions carefully. This game is simple, but please listen carefully to our instructions, and if you have any questions, please raise your hand.

In this game, there are six groups of prizes, each group has two payouts. You can choose any one of the groups. After you've made your choice, how much money you can win depends on the color of the ping pong ball you draw. There are 2 ping pong balls in this black plastic bag, one is yellow and the other is white.

Please write down the number of the group you choose in the blank space below.

I would like to choose number_______.



Experimental Instructions (Translation)

Hello everyone, welcome to today's game. Today's game has no impact whatsoever on your academic performance. Depending on your choices in the game, you will have a chance to win a cash prize, so please make your decisions carefully. This game is simple, but please listen carefully to our instructions, and if you have any questions, please raise your hand.

In each of the 6 questions that we ask you, you must choose: Option A – lottery, or Option B – fixed payment. If you choose A, each outcome of Option A has 50% chance of occurring.

We will record your six choices, and then randomly choose a number between 1-6 to determine which choice will decide your payment.

ID:__________

1	**☐ Option A: Lottery**				**☐ Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 1 RMB
2	**☐ Option A: Lottery**				**☐ Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 2 RMB
3	**☐ Option A: Lottery**				**☐ Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 3 RMB
4	**☐ Option A: Lottery**				**☐ Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 4 RMB
5	**☐ Option A: Lottery**				**☐ Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 5 RMB
6	**☐ Option A: Lottery**				**☐ Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 6 RMB