# Gender Inequality, Household Bargaining, and Social Welfare: A Structural Analysis of China's Two-Child Policy Using Machine Learning

Link to the latest version

Fangzhu Yang *

November 13, 2023

**Abstract**

China relaxed its strict One-Child Policy to universally allow couples to have two children in 2016. Although the new policy suggests an improvement in welfare for couples, as they now have more freedom to achieve their desired fertility levels, it has the drawback of possibly increasing gender inequality both in the labor market and within the household. This paper starts with a difference-in-difference method to show that the new policy increased the gender wage gap between women and men and negatively affected the intrahousehold bargaining power of women. Motivated by this empirical pattern, I then build and estimate a dynamic collective household model to quantify the welfare impact of the new policy on both genders using a novel machine learning method and indirect inference. The results suggest that the welfare cost of the Two-Child Policy for women is equivalent to 6.00% of lifetime consumption, while the welfare benefit of the policy for men is equivalent to 7.23% of lifetime consumption. Policy experiments suggest that implementing anti-discrimination laws for women in the labor market significantly improves women's welfare while providing public childcare subsidies is most effective in stimulating fertility in the post-policy era.

**Keywords**: gender inequality, intra-household bargaining, fertility, family planning policies, labor market discrimination, social welfare, machine learning, reinforcement learning

*Department of Economics, Johns Hopkins University, 3100 Wyman Park Dr, Baltimore, MD 21211. Email: fangzhu@jhu.edu. I extend my heartfelt gratitude to my advisors, Prof. Yingyao Hu, Prof. Yujung Hwang, and Prof. Robert Moffitt, for their invaluable insights, guidance, and numerous revisions of my paper. I am deeply thankful to the many individuals who shared their expertise at various seminars, enriching my work. Special thanks to Prof. Jian Ni, Prof. Michael Keane, Prof. Lixiong Li, Prof. Jonathan Elliott, Prof. Stelios Fourakis, Prof. Mohammed Ali Khan, Prof. Elena Krasnokutskaya, Prof. Nicholas Papageorge, Prof. Matthew Wiswall, Prof. Adam Isen, Prof. Juan Pantano, and Prof. Naoki Aizawa. I also want to thank the applied microeconomics group from my cohort and the department for their help and encouragement, including Huan Deng, Silin Huang, Qingyang Han, Yusuf Kulu, Matt Zahn, Chuhan Liu, Xinyu Zhao, Strahil Lepoev, and Wonsik Ko. Lastly, I would like to thank Wei Jin for his support and assistance throughout this journey.

# 1 Introduction

China announced the Two-Child Policy nationwide in 2016, which allows couples to have two children without any restrictions, and therefore, put an end to the well-known One-Child Policy. Such a significant policy shift not only leaves a notable gap in research examining its effects on women's welfare and gender inequality, but also provides a unique opportunity for us to better understand fertility decisions, household bargaining, and labor supply. This paper adopts a structural framework to study how China's universal relaxation of the One-Child Policy in 2016 affected gender inequality and social welfare. Specifically, I analyze the interplay between policy impacts on the women's labor market and the subsequent transmission of these effects into intra-household bargaining dynamics and resource allocations. I start with generalized Difference-in-Differences methods to show that the Two-Child Policy negatively affected women's wage rate and at the same time decreased women's intra-household bargaining power. Then, motivated by the empirical evidence, I develop and estimate a dynamic collective household model with limited commitment to quantify the welfare impact of the new policy on both genders. This model integrates the reduced-form findings that the policy change negatively influenced women's intra-household bargaining power, primarily stemming from a reduced wage rate relative to their husbands. It also contains the policy feature that relaxes the fertility constraints imposed on couples.

In the estimation of the model, I adopt the novel policy gradient methods from the reinforcement learning literature for solving the dynamic model and estimating the structural parameters. The estimates suggest that the Two-Child Policy has a welfare cost for women while a welfare gain for men, thus enlarging the gender inequality in China. The policy effect exhibits rich heterogeneity, with the most pronounced adverse effects observed among highly educated women who desired to have only one child. Policy experiments indicate that the enforcement of anti-discrimination laws in the labor market yields substantial improvements in women's welfare. Meanwhile, the provision of public childcare subsidies is most effective for boosting fertility and enhancing welfare during the post-policy era.

Beyond its policy implications in China, this paper utilizes the unique policy shift that relaxes people's fertility constraints to provide new evidence on the adverse impact of shifts in anticipated and actual fertility on women's labor market circumstances. It demonstrates that the worse outside options can lead to changes in couples' intra-household resource allocations, hurting women's welfare. This paper suggests that policies seemingly designed to improve welfare, such as the Two-Child Policy, may inadvertently exacerbate gender inequality when viewed through a household lens that takes into account gender disparities within the household.

**Background** Since China's implementation of the One-Child Policy in 1979, there has been a rich stream of literature studying the impact of this extreme family planning policy on various economic and social outcomes. After nearly four decades of implementation, this controversial policy was replaced by the Two-Child Policy nationwide in 2016, allowing couples to have two children unconditionally. The Two-Child Policy relaxes the rigid fertility constraint placed on couples, suggesting an improvement in overall population welfare as couples with a high fertility intention can now have their ideal number of children. Therefore, one positive channel for this policy to increase people's welfare is to expand their choice set of fertility and make them reach their desired number of children if the number is higher than one. However, it also has the potential to increase gender inequality in China from two major channels.

One channel is that, since the relaxation of the One-Child Policy, gender discrimination in the labor market against women has been increasingly reported (He et al., 2022). Due to higher potential and realized fertility, women may face worse labor market conditions, including lower hourly wages than their husbands compared to before (Li, 2022). The literature has well documented that changes in outside options will affect couples' intrahousehold behaviors (Voena, 2015; Calvi, 2020). In this case, the reduction in outside options, such as lower relative wages compared to the husbands', decreases wives' bargaining power within the household, resulting in less preferred intrahousehold resource allocations. Moreover, when women and men differ in fertility preferences, women's lower bargaining power will also negatively affect their fertility autonomy (Doepke and Kindermann, 2019; Rasul, 2008; Ashraf et al., 2014). Therefore, worse labor market conditions and the resulting decreased bargaining power for women may exacerbate the gender inequality problem in China.

The other channel is that, the "double burden" problem became more severe for Chinese women after the Two-Child policy was implemented. Due to social norms, more of the childcare burden is traditionally levied on the wife than the husband, regardless of their labor market conditions (Myong et al., 2021). After the relaxation of fertility constraints, it has become increasingly difficult for women to balance work and life due to a higher request for childcare time. Therefore, when analyzing the welfare impact of this new policy on individuals, it is necessary to view the wife and the husband as separate agents since the policy has heterogeneous effects on their labor market outcomes, within-household bargaining power, and childcare time input. The interplay between the policy's negative impact on women in public and private spheres will reinforce each other and may lead to more significant gender inequality within the household and in society (Qian and Jin, 2018).

**My Contributions** With these multiple channels in play simultaneously, it is unclear what the overall welfare impact of this policy is on both genders. This paper makes contributions to the literature as follows:

First, I establish a structural framework aimed at quantifying the effects of the policy reform on gender inequality and overall social welfare. Instead of treating the household as a whole, I employ a collective life-cycle model that considers the husband and wife as separate agents engaged in cooperative bargaining regarding fertility outcomes. In each period, they choose private consumption, leisure, and a common birth decision. The Two-Child Policy affected the couple by having an impact on both the penalty term for having excess birth and the wage process. Moreover, recognizing that changes in outside options can influence individuals' bargaining power, I incorporate a limited-commitment feature into the collective model, where Pareto weights update in each period contingent on the relative wages of the wife and husband, therefore also affected by the policy reform. Consequently, the policy reform's adverse impact on the wife's labor market prospects is transmitted to her disadvantaged intra-household resource allocations, including fertility outcomes. Utilizing the structural model, I can systematically measure the policy reform's welfare impact on both genders through distinct channels: the relaxation of fertility restrictions and its influence on intra-household resource allocations.

Second, this paper presents solid empirical findings to motivate my structural model. Employing a generalized difference-in-differences approach, I utilize the 2014 excess fertility rate residual (Zhang, 2017) as a proxy for the policy intensity of the One-Child Policy. This measure allows me to identify provinces that experienced substantial policy shocks upon the introduction of the Two-Child Policy. These provinces are then categorized into treatment and control groups based on the magnitude of these policy shocks. I then run a triple-difference regression on the impact of the policy on the hourly wage rate for women, using men as another control group. The results suggest that the policy reform negatively affected women's wage rate compared to men, providing evidence for the enlarged gender wage gap brought about by the policy relaxation. Next, I run a difference-in-differences regression on the impact of the policy on the alcohol consumption for men and women. I use alcohol consumption as a proxy for men's bargaining power within the household since alcohol is traditionally considered to be a male-favored good. The result shows a significant positive coefficient for the policy effect parameter for men, suggesting that men's intra-household bargaining power has been positively affected by the policy reform. In other words, women's bargaining power has been negatively affected by the policy reform. The two sets of reduced-form analysis shed light on how the Two-Child Policy has increased the welfare gap between women and men through the enlarged gender wage gap, which is then transferred into women's lower bargaining power and thus worse intra-household allocations. Since the policy also has a welfare-enhancing channel as it relaxes the rigid fertility constraints, the reduced-form analysis is not able to argue whether women's overall welfare has been improved or not, thus motivating my structural model.

Third, on the technical front, I adopt the policy gradient methods from the reinforcement learning literature (Sutton et al., 1999) to tackle the large state space problems when solving for the optimal policy. This method helps deal with the large state space problem since it directly works with the policy function without discretizing the state space or calculating the expected value function over the entire state space. It has been widely applied to many fields such as robotics and medicine. Nonetheless, its applications in economics remain limited, despite its growing popularity in recent times [1]. Using this method, I directly parameterize the policy as a stochastic function of the state variables. Utilizing the policy gradient theorem (Sutton et al., 1999), I can then update the policy function parameters using the gradient of the policy function in a stochastic gradient descent framework. After sufficient iterations, the algorithm outputs the optimal policy function that gives the highest lifetime reward for individuals.

In solving the structural parameters, I combine the inner loop of solving the policy function using policy gradient methods with the outer loop of using indirect inference to search for the optimal deep parameters. The identification relies on the data features that I can observe the private leisure choices of the husband and the wife. At the same time, I can also observe the separate ideal number of children for men and women, plus the actual number of children in the household. Using these two statistics, I can identify the parameters in the Pareto weight equation, which is often considered to be the most challenging part of identification in collective household models. For the moments of estimation, I use a total of 58 statistics, consisting of both first-order moments and regression coefficients. The model fits are reasonably good, for both the training set and the validation set.

Fourth, employing the estimated structural model, this paper reveals that the policy has exacerbated gender inequality by adversely affecting women's welfare while positively impacting men's welfare. Specifically, the Two-Child Policy imposes a welfare cost on women equivalent to 6% of their lifetime consumption, while benefiting men equivalent to 7.23% of their lifetime consumption. Moreover, the estimated bargaining power for women has experienced a sharp decline following the policy change, from approximately 0.39 in 2010 to around 0.27 in 2020. These findings demonstrate how the policy reform has widened the

[1] Charpentier (2023) provides a detailed introduction to the state-of-the-art of reinforcement learning techniques and their applications in economics. Hughes (2014) discusses how reinforcement learning can be applied to complex multi-agent problems such as stochastic games.

welfare disparity between women and men, intensifying gender inequality both in the labor market and within the household. In addition, there exists substantial heterogeneity in the policy's impact, with highly educated women and women desiring only one child mostly affected by the policy change. This finding is intuitive, as these women do not benefit from the policy's relaxation of fertility constraints. Instead, they may face higher fertility outcomes if their husbands desire more children, while simultaneously suffering from worse labor market conditions and reduced intra-household bargaining power.

Fifth, I perform a series of counterfactual policy experiments to explore how different policies could enhance social welfare and promote fertility. Firstly, I shut down the adverse impact on women's wage rates, preventing the widening of the gender wage gap post-policy reform. The result reveals that, in the absence of this wage effect, women, on average, benefit from the Two-Child Policy. The second set of policy experiments shows that compared to reducing children's educational costs, providing public childcare services is more effective in stimulating fertility and improving women's welfare in the post-policy era. Finally, a more equitable distribution of the childcare burden will improve women's welfare while decreasing the number of children couples have, since men now bear a larger share of the childcare burden and are less inclined to have a second child. In summary, the results of these policy experiments suggest that, to mitigate the negative impact of the Two-Child Policy on gender inequality, it is advisable to promote anti-discrimination policies safeguarding women in the labor market and to advocate for policies that facilitate access to public childcare services.

**This Paper in the Related Literature** My paper is mainly connected to four streams of literature: (1) endogenized fertility decisions within households, (2) motherhood penalty on labor market outcomes, (3) women's bargaining power and intra-household resource allocations, (4) effects of family planning policies on various socioeconomic outcomes in China, and (5) machine learning and dynamic discrete choice estimation.

There has been a large stream of literature studying endogenized fertility decisions in a dynamic setting, such as Gayle et al. (2021), Francesconi (2002), Keane and Wolpin (2010). My work is, in particular, related to the papers that view the husband and the wife as separate agents and employ a bargaining framework for fertility decisions (Doepke and Kindermann, 2019; Ashraf et al., 2014, Baudin et al., 2015; Myong et al., 2021). My paper differs from the existing literature by developing and estimating a dynamic life-cycle model where the husband and the wife bargain over fertility using a cooperative framework with limited commitment. Furthermore, my research utilizes a unique fertility policy reform in a developing country, a departure from the prevailing literature which predominantly concentrates on fertility patterns in developed nations where fertility policy changes are less prevalent or non-existent.

My paper is also closely related to the extensive body of literature investigating the motherhood penalty on women's labor market outcomes. To name a few, Adda et al. (2017) estimate a dynamic life cycle model of labor supply, fertility, and savings. They quantify the career costs associated with children, focusing on sorting into the labor market and across occupations. Kleven et al. (2019) study the impacts of children on gender inequality in the labor market using Danish administrative data. They found that the earnings penalty comes from labor force participation, hours of work, as well as the wage rate. Chen et al. (2020) study the impact of motherhood on the within-firm gender gap in China. They found that reduced working hours and women's sorting into lower-level jobs together contribute to the overall gender gap in the labor market. My paper distinguishes itself from the existing literature by not solely examining the motherhood penalty on women's labor market outcomes but also by considering how this penalty leads to diminished intra-household bargaining power and less favorable resource allocations. To achieve this, I integrate the reduced-form analysis of the policy's impact with a structural model of household bargaining, demonstrating the interplay between penalties experienced in both the public and private spheres.

My work also builds on the literature in family economics studying the effect of legal reforms on intra-household allocations through changing the bargaining power. For instance, Voena (2015) investigates the effect of changes in divorce laws on couples' intertemporal choices and well-being. Calvi (2020) studies how changes in the heritage law affected Indian women's intrahousehold bargaining power and thus affected their mortality risk at an older age. Huang et al. (2023) investigates how changes in China's property division laws affected couples' intrahousehold decisions. They found that a modification in the property division laws in favor of men resulted in a better intrahousehold allocation for them. Low et al. (2020) looked at how a U.S. welfare reform that introduced limits on years of welfare receipt affected the intrahousehold allocations of women. My paper departs from these papers by examining the unexplored topic of how shifts in fertility laws can influence individuals' bargaining power and, consequently, reshape the allocation of resources and fertility outcomes within households.

Furthermore, my paper is related to the existing body of literature on the impact of family planning policies in China. Since the implementation of the One-Child Policy in China, many scholars have studied its impact on Chinese society from a variety of perspectives. To name a few, Yu et al. (2020) argue that the One-Child Policy amplified economic inequality in China since poor families from rural areas were less constrained by the OCP than rich ones and had more children but invested less in human capital per child. Wang et al. (2017) examine the labor market consequences of China's family planning policies, focusing on the aging problem of the working population. They find that because of internal migration, the aging problem is more severe in rural areas than in urban areas. Apart from the One-Child Policy, some other papers study the impact of China's "Later, Longer, Fewer" (LLF) campaign on the whole society [2]. For instance, Chen and Huang (2020) argue that this policy can explain about half of the decline in China's total fertility rate from 5.7 in 1969 to 2.7 in 1978. Chen and Fang (2021) study the long-term impact of the LLF policy in old ages and find that parents more exposed to this policy report more severe depressive symptoms.

Ever since the announcement of the Two-Child Policy in 2016, a growing body of literature has emerged, seeking to evaluate the social consequences of the relaxation of China's previously stringent family planning policies. For instance, He et al. (2022) send out fictitious resumes to investigate labor market discrimination against expected family responsibilities. They find that women, but not men, are subject to labor market discrimination after the policy change. Li (2022) finds that women who become eligible to have a second child are having more new birth, and are less likely to work, work fewer hours per week, and earn less. My paper differs from existing research by developing and estimating a dynamic structural model. This model allows for a comprehensive formal welfare analysis of women after the policy shift, taking into account multiple channels through which the policy may influence women's well-being. These channels include the relaxed fertility constraints, the adverse effects of the policy on women's labor market outcomes, and the subsequent implications for intra-household resource allocations.

Finally, to the point of estimation, my paper builds on the computer science literature on reinforcement learning. Specifically, I adopt the popular policy gradient method from this literature to solve for the optimal policy function for my dynamic model and estimate the structural parameters. Sutton et al. (1999) first propose the policy gradient method, which has become widely applicable in many fields such as robotics and motor skills (Peters and Schaal, 2006; Peters and Schaal, 2008), medical applications (Jin et al., 2023; Hua et al., 2022), and many other empirical practices. My paper differs from the existing literature by applying the policy gradient method to economics and combine this method with indirect inference to estimate the structural parameters, which yields more fast and robust results.

The rest of the paper is organized as follows. I first provide the institutional background and summarize the data pattern in Section 2. I then show reduced-form evidence for the policy impact on the gender wage gap and the intrahousehold bargaining power for women

[2] The "Later, Longer, Fewer" (LLF) campaign belongs to a series of family planning policies carried out in the early 1970s in order to bring down the fertility rate. The LLF policy was implemented earlier than the one-child policy. "Later" means marriage at a later age – 23 years for women and 25 years for men. "Longer" refers to the norm of waiting more than three years between births. "Fewer" means that one couple could not have more than two children.

in Section 3. I present the structural model in Section 4, followed by the identification and estimation results in Section 5. In Section 6, I conduct welfare analysis and counterfactual experiments. Finally, Section 7 concludes.

# 2 Institutional Background and Data

## 2.1 Institutional Background

In response to the rapid population growth, the Chinese government implemented a series of family planning policies during the 1970s to control birth rates. These policies began with a relatively lenient version that encouraged couples to limit their family size before 1979. Subsequently, in 1979, the policy took a much stricter form, known as the One-Child Policy, which mandated that couples could have only one child. This policy was progressively enforced rigorously nationwide in 1980. To ensure its effectiveness, local governments undertook a series of measures and actions. As an example, if parents decline to pay the fine imposed for having a second child, the unauthorized children may be placed on a blacklist for hukou birth registration. This, in turn, would deprive them of access to a range of essential public services, including schooling.

Although the One-Child Policy was implemented nationwide in China, variations in its strictness emerged across provinces and among individuals with differing characteristics. To illustrate, in 1981, the fine for violating the policy was 1.23 times the annual household income in Beijing, yet it was only 0.647 times the annual household income in the province of Shaanxi (Ebenstein, 2010). Policy enforcement also varied based on personal attributes. For instance, parents who exceeded the birth limit could lose their jobs if they were employed by the government or state-owned enterprises. Furthermore, disparities existed between urban and rural areas, with urban areas generally witnessing more stringent and rigorous policy enforcement compared to their rural counterparts. Remarkably, in many provinces, rural residents were allowed to have a second child if their firstborn was a girl, a policy informally referred to as the "1.5-child policy." Lastly, it's important to note that the policy primarily targeted Han ethnic couples, which constitute approximately 92% of China's total population, while ethnic minority groups were not subject to the same policy restrictions.

The One-Child Policy underwent a series of relaxations beginning in the early 2000s due to concerns about declining fertility rates. In 2011, all provinces began permitting couples who were both only children to have two children. Subsequently, in November 2013, a revised policy allowed couples to have two children if at least one of the marital partners was an only child. However, even eligible couples under this policy still had to apply for permission to have a second child. A significant change occurred in December 2015 with the introduction of the universal Two-Child Policy, which allowed all couples to have two children, starting on January 1, 2016. Notably, unlike the earlier relaxations, this universal Two-Child Policy no longer required couples to seek a birth permit to legally have a second child. This dramatic policy shift aimed to address the compounding issues of a declining fertility rate, skewed sex ratio, and a rapidly aging population that resulted from the stringent One-Child Policy [3].

Although people had expected that the One-Child Policy would eventually be removed, the timing of the termination of the policy, however, was unanticipated by the general public (He et al., 2022). The radical policy change came as a big shock to the citizens in China at that time, putting an end to the almost four decades old, highly controversial One-Child Policy. It was estimated that about 100 million couples in China would be affected by the policy change. The universal Two-Child Policy will undoubtedly have a significant impact on Chinese society in many aspects. One crucial aspect is the effect on women's labor market conditions. Due to the potentially higher fertility of women and traditional gender roles in childbearing, employers may not treat them as favorably in the labor market as men. Combined with the fact that the labor market in China was not tight such that it is easy for employers to find qualified workers, employers will be willing to hire a male counterpart to replace the female candidates. Therefore, a worse labor market condition featuring a lower wage rate for women than before is an unintended consequence of the seemingly good Two-Child Policy.

## 2.2 Data

The primary data set I use is the China Family Panel Study (CFPS), a nationally representative, bi-annual longitudinal survey data [4]. CFPS employs a household-based survey design, having a survey for the household as a whole and one survey for each member living in this household. I take advantage of this structure of the survey to link the husband and the wife and track them using all six waves of data spanning from 2010 to 2020, as these waves cover the critical policy change that occurred in 2016. In the dataset, I can observe individuals' job histories, fertility histories, and time-use data. Within the job history survey section, respondents provide information about their primary employment, including details such as job start and end dates, wages, income, working hours, industry, and occupation. Regarding fertility histories, all children residing with the survey respondents are recorded in the household-linkage data, which includes basic information like gender and age. By leveraging this information, I construct complete fertility histories for each couple beginning

[3] Recently, the Three-Child Policy was announced in China, basically relaxing all the constraint put on couples' fertility behaviors. Since this new policy was carried out in 2021, which is out of my data range, I do not consider this further relaxation of the family planning policy in my study.

[4] The website for the data is https://www.isss.pku.edu.cn/cfps/en/.

from the time of their marriage, including the timing of births and the gender of their children. Furthermore, the 2010 wave of data includes a comprehensive time-use module that inquired about respondents' daily activities, including sleep, meals, leisure, work, home production, and more. Although this module was omitted in subsequent waves, the later waves still provide information on working hours and hours spent watching TV. In Appendix B, I outline the procedure used to impute leisure hours for the later years, utilizing the hours spent watching TV as a proxy.

Another noteworthy feature of the dataset is that in waves 2014 and 2018, respondents were asked about their ideal number of children. The question is as follows:

*"How many children do you think you would ideally have?"*

This question was asked separately to the wife and the husband in both waves. I interpret the answer to these questions as showing the intrinsic fertility preferences that the person has. In other words, by knowing both the wife's and the husband's ideal number of children, we can separately observe their preferences for fertility, which might be different in some households [5]. This variable plays an important role in my analysis as it reveals information on the extent to which people's fertility intentions were constrained before the policy change, and how closely the actual fertility aligns with both the husband's and wife's ideal number of children after the policy relaxation.

Having described the variables in the data that are key to my study, I proceed to discuss how the sample of interest is constructed. As my research focuses on examining the interaction between the husband and wife and their joint decisions within the household, I focus on couples who got married prior to 2016 and follow them across all waves in which they remained married. Since I choose to not model divorce, I exclude any periods where the couples got divorced. Around 3% of all couples ended up divorcing in the end, which is a reasonably small proportion of the whole population. Therefore, it is safe to abstract away from modeling divorce, since it rarely happened in my data. Furthermore, I specifically target couples where the wife's age fell within the range of 20 to 40 years old in 2016. This criterion ensures that the wife remained within the fertile age range when the Two-Child Policy was implemented, making the fertility decision still relevant for the couple. To reduce the computational burden, I narrow my focus to couples with a total number of children

[5] The question posed in 2014 was phrased somewhat differently: "*Regardless of policy constraints, how many children do you think would be ideal?*" This variation in wording reflects the ongoing implementation of the One-Child Policy at that time. In my analysis, I will exclusively consider the responses to the question asked in 2018, as they can more accurately capture individuals' fertility preferences without any actual policy constraints. The responses to the 2014 question will serve the purpose of calculating the excess fertility rate residual, a topic that I will discuss in detail in the empirical evidence section.

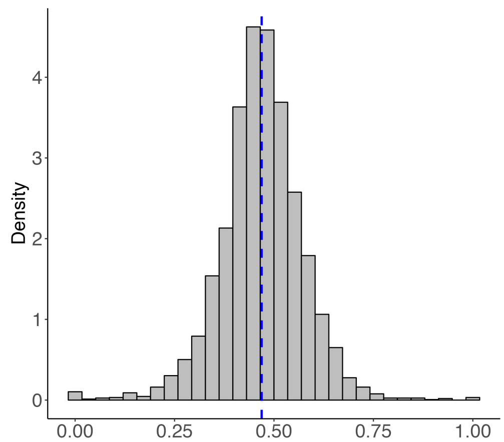
ranging from 0 to 3, and whose ideal number of children also falls within the 0 to 3 range. In this step, around 3% of couples were deleted, having either more than three children or wanting to have more than three children [6]. Meanwhile, I exclude couples in which the husband's age exceeds that of the wife's by more than 10 years. This is because, during the estimation phase, I make a simplifying assumption that the husband's age is two years older than the wife's age. In this step, 2% of couples were deleted, having an age difference greater than 10. Lastly, my analysis focuses exclusively on urban couples, as the Two-Child Policy primarily targeted this demographic. Additionally, the majority of the rural population is engaged in the agricultural sector, which falls outside the scope of my sample, which is centered on the employed population. Collectively, these criteria yield a final dataset for estimation comprising 1,583 couples and 10,744 observations. This dataset allows for the tracking of all these couples over multiple periods. Table 1 presents the summary statistics for key variables.

As shown in the table, the average age of women in the sample is 29.91, whereas men's average age is 31.76, indicating an average age difference of two years, with husbands typically being older than their wives. Regarding educational levels, the majority of both men and women have completed education levels below high school, with less than 30% of individuals having attained a college degree or higher. Husbands and wives exhibit similar profiles, with husbands generally having slightly higher educational qualifications. The average hourly wage rate for women stands at 18.64 Yuan (equivalent to approximately 2.57 USD), while men earn an average hourly wage rate of 28.01 Yuan (around 3.86 USD). This implies that, on average, men earn approximately 10 Yuan more per hour than women. In terms of time allocation, men, on average, enjoy 3 more hours of leisure time compared to women. The average working hours for men amount to 48.88, suggesting that the majority are engaged in full-time employment. Conversely, women have significantly fewer market work hours, averaging at 34.52. When it comes to home production hours, encompassing both housework and family caregiving, women predominantly shoulder these responsibilities, dedicating an average of 29.72 hours. In contrast, men contribute significantly fewer hours to home production, averaging at 11.28.

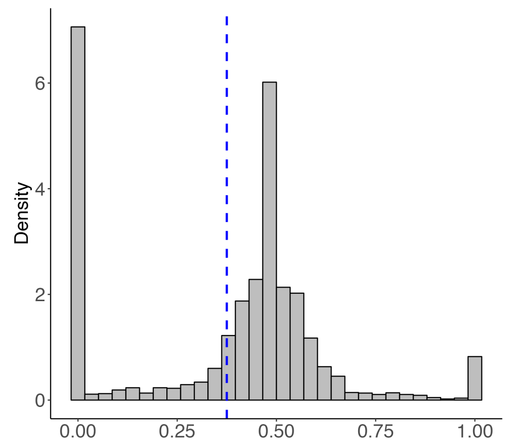
In terms of fertility-related statistics, the average ideal number of children for women in 2018 is 1.750, whereas men express an average ideal of 1.821 children in the same year. This discrepancy indicates that, on average, men have a higher desired number of children

[6] I choose to exclude those couples that have more than three children or wanting to have more than three children both because I would like to reduce the computational burden of the structural model later and because having more than three children is out of the scope of the Two-Child Policy. Since in the data, only 3% of couples fall into this category, I will be less concerned about any potential sample selection issues in this step.

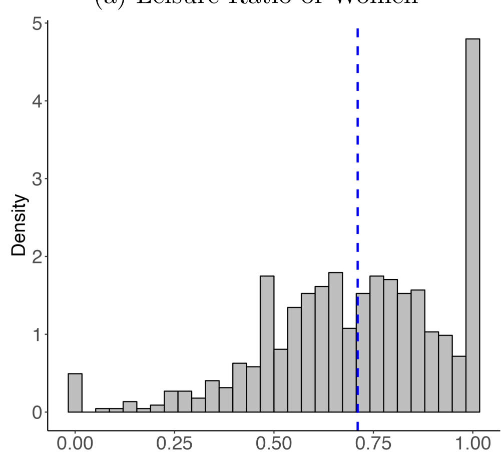
compared to women, a trend consistent with fertility preference patterns observed in other countries. Furthermore, the average number of children that couples had before the policy change is 1.131, whereas after the policy relaxation, the average increases to 1.495 children. Notably, there is a significant jump in the fraction of couples with exactly two children after the policy change. This signals a fertility response to the policy relaxation, resulting in more second births among couples. Figure 1 presents histograms showing the ratio of the wife's hours to the total hours of the husband and wife across different activities. The results confirm that the wives' timeshares for childcare and total home production are both greater than 0.5, while their timeshare for market work is way less than 0.5. In addition, their timeshare for leisure is also smaller than 0.5, suggesting that they are enjoying less leisure than their husbands.



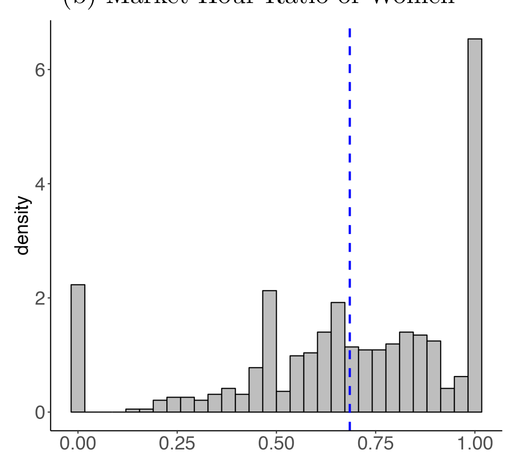

(a) Leisure Ratio of Women




(b) Market Hour Ratio of Women




(c) Home Production Ratio Hour of Women




(d) Childcare Hour Ratio of Women

Figure 1: Time Use Shares for Wives Within the Household

Table 1: Summary Statistics for Key Variables (CFPS 2010-2020)

| | Wife | Husband | Household |
|---|---|---|---|
| Characteristics | | | |
| Age | 29.91 [5.14] | 31.76 [5.56] | |
| Education | | | |
| - Less than high school | 0.468 [0.499] | 0.460 [0.499] | |
| - High school graduate | 0.250 [0.433] | 0.241 [0.428] | |
| - College and higher | 0.282 [0.450] | 0.299 [0.458] | |
| Hourly wage rate (CNY) | 18.64 [65.51] | 28.01 [217.61] | |
| Time use, hours per week | | | |
| Leisure | 25.76 [10.61] | 28.75 [10.47] | |
| | (28.63%) | (32.34%) | |
| Market work | 34.52 [26.27] | 48.88 [21.22] | |
| | (38.36%) | (54.98%) | |
| Home production | 29.72 [22.51] | 11.28 [11.04] | |
| | (33.03%) | (12.69%) | |
| Fertility | | | |
| Ideal number of children (2018) | 1.750 [0.521] | 1.821 [0.523] | |
| Number of children | | | |
| - Before the policy change | | | 1.131 [0.651] |
| - Children number=0 | | | 0.135 [0.342] |
| - Children number=1 | | | 0.618 [0.486] |
| - Children number=2 | | | 0.227 [0.419] |
| - Children number=3 | | | 0.020 [0.139] |
| - After the policy change | | | 1.495 [0.612] |
| - Children number=0 | | | 0.019 [0.137] |
| - Children number=1 | | | 0.510 [0.500] |
| - Children number=2 | | | 0.427 [0.495] |
| - Children number=3 | | | 0.043 [0.204] |

Note: This table presents the summary statistics for the key variables in the dataset that I use. The numbers in the square brackets denote the standard deviations for each variable. There are a total of 1583 couples tracked from 2010 to 2020, resulting in 10,744 observations. Age, hourly wage rate, and time use are averaged over all waves of the data. The hourly wage rate is in Chinese Yuan (CNY). The time use variables exclude hours for basic activities such as sleep, eating, and personal care. Leisure, market work, and home production hours therefore sum up to around 89 hours per week (an individual needs around 89 hours for basic activities weekly). The average market work hours include the unemployed population, whose market work hours are coded as 0. The ideal number of children is averaged using the 2018 value, and the actual number of children is averaged over individuals using their last periods before the policy change and their most recent periods after the policy change. The fraction of actual number of children is calculated using the same method.

# 3 Reduced-Form Analysis for the Policy Impact

In this section, I provide empirical evidence on the effect of the Two-Child Policy on both the gender wage gap in the labor market and the bargaining power of women within the household. In Section 3.1, I discuss the generalized Difference-in-Differences method I adopt for conducting the empirical analysis and how I categorize provinces into treatment and control groups using excess fertility rate residuals. Section 3.2 contains the summarize of the Difference-in-Differences results for the policy impact on birth rates, gender wage gap in the labor market, and women's intrahousehold bargaining power.

## 3.1 Generalized Difference-in-Differences Analysis

To quantify the effect of the Two-Child Policy on the gender wage gap in the labor market and household outcomes, I employ a generalized Difference-in-Differences method. Although the new policy was carried out nationwide, there were variations in the policy strictness of the One-Child Policy across provinces pre-2016. I define provinces with a strict One-Child Policy as the treatment group, as they went through a larger policy shock than those with a loose One-Child Policy before. Following Zhang (2017), I use excess fertility rate residuals in 2014 to measure the strictness of the One-Child Policy for each province. The excess fertility rate residuals are variations in excess fertility rates that cannot be explained by observed fertility preferences, thus serving as a proxy for the level of policy stringency associated with the One-Child Policy.

The excess fertility rate is defined as the proportion of women with a second child in year $t$, out of all the women with one child in year $t-1$. It measures the degree of violation of the One-Child Policy. Under perfect compliance, the excess fertility rate should just equal zero. I look at women from 2010-2014 and assess the proportion of these women who had a second child in the following year (year plus one). Subsequently, I calculate the average excess fertility rate just prior to the implementation of the Two-Child Policy by averaging these proportions. To obtain the excess fertility rate residuals, I run the following province-level regression using the ideal number of children and other provincial characteristics as regressors:

$$\begin{aligned} ExcessRate_k = \beta_0 &+ \beta_1 HighSchoolabove_k + \beta_2 Agriculture_k \\ &+ \beta_3 IdealNum.ofChild_k + \beta_4 Age_k \\ &+ \beta_5 Log(Houseprice)_k + \beta_6 Log(GDP)_k + \varepsilon_k, \end{aligned}$$

where $k \in 1, 2, ..., 25$ stands for the 25 provinces in my data, $ExcessRate_k$ is the average excess fertility rate for province $k$, and $IdealNum.ofChild_j$ is the average ideal number of children couples want in province $k$ in 2014. Other characteristics include proportion of people having high school degrees or above, fraction of people working in agriculture, average age of the population, log of housing prices, and log of GDP per capita. These covariates are assumed to be contributing to people's willingness to have a second child, and thus the violation of the One-Child Policy. Table D.4 shows the result for the fertility regression. The result suggests that the coefficient for the ideal number of children is significantly positive, while all the other coefficients are insignificant.

It is intuitive that the ideal number of children plays the most crucial role in deciding the excess fertility rate. The higher the average ideal number of children is, the stronger the preference for fertility is, and the more likely that the couple is going to violate the one-child restriction, thus resulting in a higher excess fertility rate. After controlling for the intrinsic preferences for fertility, other characteristics become insignificant. The residuals $\varepsilon_k$, therefore, are variations in the excess fertility rate that cannot be explained by all the observed fertility-related factors in that province. For example, if two provinces have similar ideal numbers of children and other characteristics for couples living there, while one province has a smaller excess fertility rate than the other, then it must be the case that this province had a stricter One-Child Policy than the other province, which resulted in fewer births. With this argument, the residual $\varepsilon_k$ can serve as a proxy for the degree of policy strictness in each province before the Two-Child Policy was carried out.

After obtaining the Excess Fertility Rate Residual $\varepsilon_k$ for each province, I rank provinces by their residuals from low to high, the lowest being those provinces that have a strict One-Child Policy and the highest being those provinces that have a lenient One-Child Policy before. Subsequently, I employ the median of the Excess Fertility Rate Residuals to categorize provinces into treatment and control groups. Figure 2 shows each province's EFR (Excess Fertility Rate) residual, where the central red dashed line representing the median point. Provinces to the left of the red line are those having a low EFR residual, thus were strictly regulated under the One-Child Policy and went through a large policy shock when the policy reform happened in 2016. Therefore, they are in the treatment group. Provinces to the right of the red line, on the other hand, have a relatively high EFR residual, indicating a loose regulation under the One-Child Policy, and there are put in the control group [7] [8].

## 3.2 Policy Effect on Fertility, Wage Gap, and Bargaining Power

WIth the treatment and control grouped defined, I am able to identify and estimate the policy effect on various outcome of interests. Firstly, I study how the Two-Child Policy affected couples' fertility choices by running a difference-in-differences regression on the provincial-level birth rates. Secondly, I run a triple-difference regression on the hourly wage rate, using males as an additional control group for women to study how the policy affected women's wage rate in the labor market. Finally, following the family economics literature, where

[7]Provinces include Inner Mongolia, Qinghai, Xinjiang, Ningxia, Tibet, and Hainan are excluded from my sample due to lack of enough samples.

[8]I have also tried to directly use the continuous measure in the Difference-in-Differences regressions, which resulted in less significant results. Therefore, I choose to keep the binary specification here.

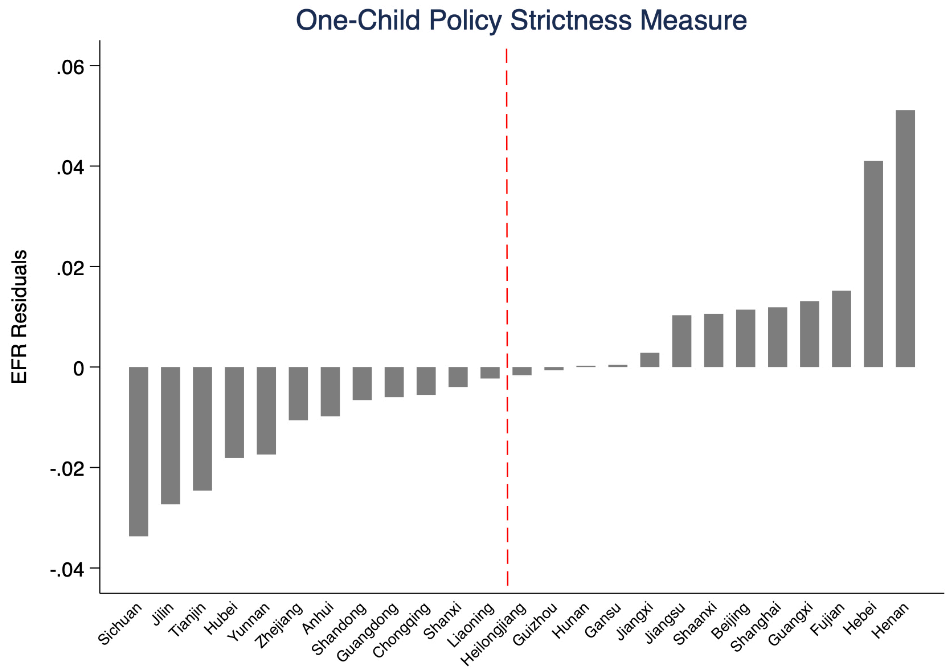



Figure 2: Excess Fertility Rate Residuals of Provinces

people use assignable private consumption goods to back out the resource-sharing rule in the household (Calvi,(2020); Huang et al.,(2023)), I use the alcohol consumption as a proxy for individual's bargaining power and run a difference-in-differences regression on this proxy. Equation 3.1, 3.2, and 3.3 summarize the regressions for the three analyses, where $\alpha_3$, $\beta_7$, and $\gamma_3$ capture the policy impact on birth rates, gender wage gap, and intra-household bargaining power, respectively.

$$\begin{aligned} Birth_{kt} = \alpha_0 + \alpha_1 Post_{2016} + \alpha_2 PolicyShock_k + \alpha_3 Post_{2016} \times PolicyShock_k \\ + f_t + f_k + \epsilon_{kt}. \end{aligned} \tag{3.1}$$

$$\begin{aligned} Wage_{ikt} = \beta_0 + \beta_1 Post_{2016} + \beta_2 Female_i + \beta_3 PolicyShock_k + \beta_4 Post_{2016} \times Female_i \\ + \beta_5 Post_{2016} \times PolicyShock_k + \beta_6 Female_i \times PolicyShock_k \\ + \beta_7 Post_{2016} \times Female_i \times PolicyShock_k + \boldsymbol{X}_{ikt}\boldsymbol{\alpha} + f_t + f_k + f_i + \epsilon_{ikt}. \end{aligned} \tag{3.2}$$

$$
\begin{aligned}
Alco_{ikt} = \gamma_0 + \gamma_1 Post_{2016} + \gamma_2 PolicyShock_k + \gamma_3 Post_{2016} \times PolicyShock_k \\
+ \boldsymbol{X}_{ikt}\boldsymbol{\alpha} + f_t + f_k + f_j + \epsilon_{ikt}.
\end{aligned} \tag{3.3}
$$

Table 2 summarizes the results for the three regressions and Figure 3 shows the corresponding event study graphs for the three analyses. The result in Column (1) suggests that the Two-Child Policy had a significantly positive affect on the provincial birth rates, which proves that there was a significantly positive fertility response to the policy relaxation. The estimated coefficient in Column (2) shows that the policy had a significantly negative impact on women's hourly wage compared to men's, enlarging the gender wage gap as a result. This result is consistent with the hypothesis that, due to a potentially higher fertility after the policy relaxation, employers prefer to hire men instead of women, resulting in a larger gender wage gap. Finally, the results in Column (3)–(4) suggest that the policy increased men's bargaining power within the household, since it significantly increased men's alcohol consumption, which is viewed as traditionally a male-favored good. On the other hand, the coefficient for women's alcohol consumption is insignificant. The reason is that women's demand for alcohol is inelastic, therefore an increase in their bargaining power does not result in significant changes in the alcohol consumption. The three event study graphs suggest that the parallel trend assumptions hold for all of the analyses, confirming the validity of the difference-in-differences regression results. More details regarding the reduced-form analyses can be found in Appendix A, where more robustness checks are also provided.

In summary, this section uses the excess fertility rate residuals to classify the provinces into treatment and control groups and adopts the generalized Difference-in-Differences methods to show that (1) the Two-Child Policy increased the fertility rate of couples and (2) had a negative impact on women's wage rate compared to men's and hence enlarged the gender wage gap in the labor market. I then show that (3) the policy also positively affected men's bargaining power while decreasing women's bargaining power within the household. These empirical results hint at the negative consequences of the Two-Child Policy on women's labor market outcomes, which then led to a decrease in the bargaining power of women within the household. The interaction between the private and public spheres makes it nontrivial to conduct welfare analysis in a reduced-form way. Therefore, starting from Section 4, I begin to build a structural model for a formal analysis of the policy impact on women's and men's welfare and on gender inequality.

Table 2: Difference-in-Differences Results for Birth Rates, Hourly Wage Rate, and Alcohol Consumption

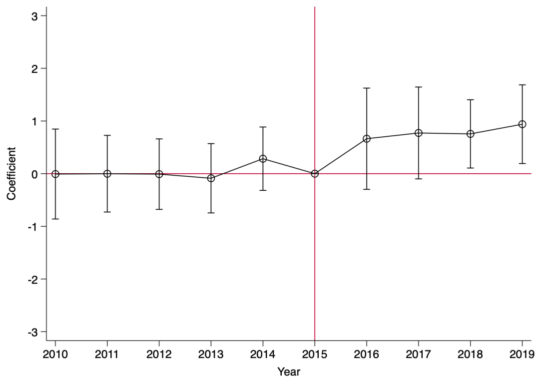
| VARIABLES | (1) Birth Rate (‰) | (2) Wage (CNY) | (3) Alcohol (Husband) | (4) Alcohol (Wife) |
|---|---|---|---|---|
| $\text{Post}_{2016} \times \text{Female} \times \text{Policy Shock}$ | | -2.357*** | | |
| | | (0.876) | | |
| $\text{Post}_{2016} \times \text{Policy Shock}$ | 0.6985*** | | 0.0435*** | 0.0028 |
| | (0.2022) | | (0.0167) | (0.0052) |
| Observations | 250 | 16,082 | 11,144 | 13,506 |
| R-squared | 0.926 | 0.198 | 0.0518 | 0.0122 |
| Individual Characteristics | N/A | Y | Y | Y |
| Year FE | Y | Y | Y | Y |
| Province FE | Y | Y | Y | Y |

Note: This table presents the results for the difference-in-differences estimation of the impact of the Two-Child Policy on the birth rate, hourly wage rate, and alcohol consumption frequency. The dependent variable in Column (1) is the average birth rate (‰) for each province and the dependent variable in Column (2) is the hourly wage rate of women in Chinese Yuan (CNY). Finally, the dependent variable in Column (3)–(4) is a binary indicator of whether consuming alcohol for at least three times last week, where Column (3) contains the result for husbands and Column (4) contains the result for wives.

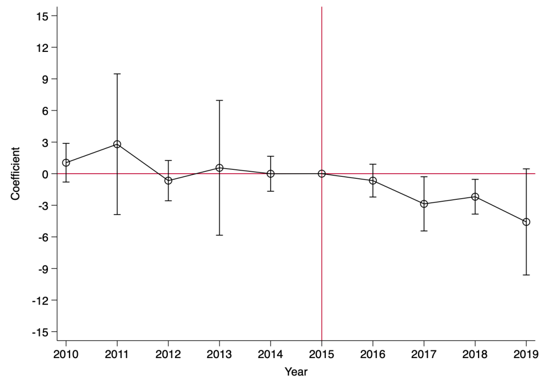
# 4 A Life-Cycle Model of Household Bargaining

To structurally quantify the impact of the Two-Child Policy on women's and men's welfare, I develop and estimate a dynamic household model. This model incorporates the two empirical findings that the gender wage gaps became larger after the Two-Child Policy and husbands have a higher bargaining power than before within the household. I assume couples interact using a cooperative framework with limited commitment. In each period, they choose private consumption, working hours, and a binary birth decision. The Two-Child Policy came as a shock for the agents, having an impact on both the utility function through the penalty term for excess birth and the wage process. The Pareto weight updates in a reduced-form way depending on the relative wage and therefore is affected by the policy change. This model helps identify the welfare impact of the policy change from various channels, focusing especially on within-household inequalities. Figure 4 depicts the timeline of the model.

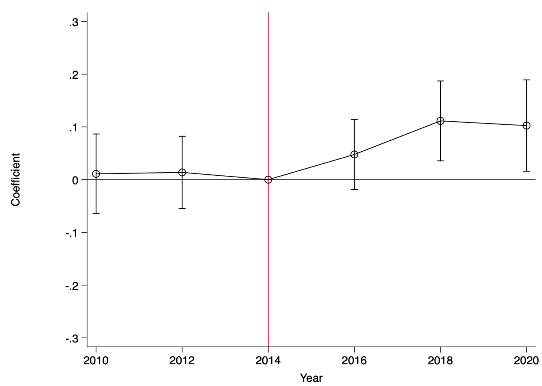
When entering the time period $t$, the couple holds the number of children $N_{j,t-1}$ and the Pareto weight $\theta_{j,t-1}$ from the last period. At the beginning of this period, the wage shocks for both the wife and the husband are realized. In the year 2016, the policy shock




(a) Provincial Birth Rate




(b) Hourly Wage Rate Gap




(c) Alcohol Consumption (Husband)

Figure 3: Event Study Graphs for the Policy Impact

is also realized. The couple then updates the Pareto weight $\theta_{jt}$ conditional on the realized wage shocks. After that, the fertility preference shock $\epsilon_{jbt}$ is realized, and the couple makes the binary birth decision $b_{ji}$. The couple then transits to the second sub-period, where the number of children $N_{jt}$ in the household is updated. This would affect the time needed for childcare and therefore decide the total time and monetary budget available for the wife and the husband. In this sub-period, the couple then chooses private consumption and leisure. Finally, the couple moves to the next period $t+1$ with the updated state variables $N_{jt}$ and $\theta_{jt}$. After describing the general timeline of the dynamic model, we move to the subsections below to discuss the details of the model.

## 4.1 Preferences

Let $j = 1, 2, ..., N$ be the index of each household in the data. Let $t = 1, 2, ..., T$ denote each time period starting from $age = 20$ to $age = 40$ for women. Each time period equals two

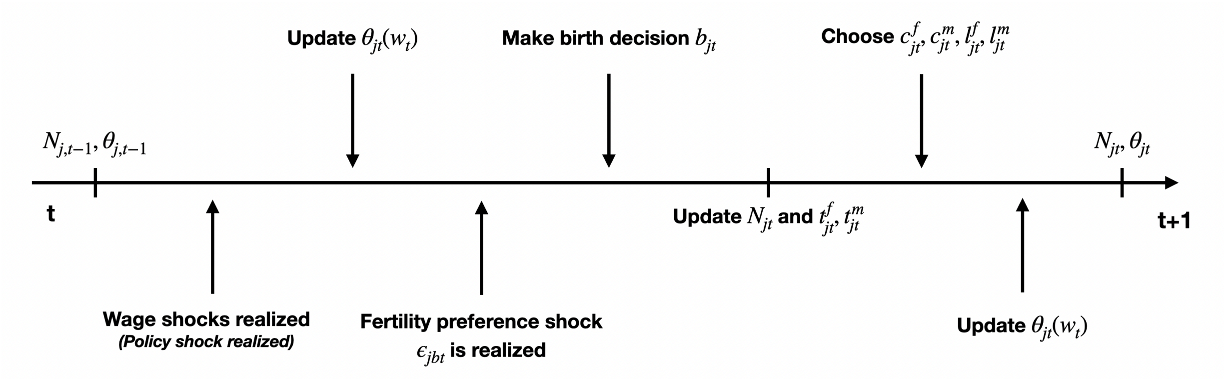



Figure 4: Timeline of the Dynamic Model

years in the data. Let $g \in \{f, m\}$ denote the wife and the husband in the household. The couple's individual flow utility depends on private consumption, leisure, and two fertility-related terms as shown in the following equation.

$$u(c^g_{jt}, l^g_{jt}, n_{jt}, \widetilde{n}^g_{jt}) = \gamma_1 \ln c^g_{jt} + \gamma_2 \ln l^g_{jt} - \gamma_3((n_{jt} - \widetilde{n}^g_{jt}))^2 - p_{jt} \times \mathbb{I}(n_{jt} > 1),$$

where

$$p_{jt} = \begin{cases} p_1 & \text{if } P_{jt} = 0 \ \& \ S_j = 1 \\ p_2 & \text{if } P_{jt} = 0 \ \& \ S_j = 2 \\ 0 & \text{if } P_{jt} = 1 \end{cases} .$$

A utility loss occurs in a quadratic form of the difference between the actual number of children in the household and the ideal number of children for this individual. This is captured by the third term in the utility function, where $n_{jt}$ is the actual number of children and $\widetilde{n}^g_{jt}$ is the ideal number of children for the wife and the husband separately [9]. I use this term to capture the couple's desire to reach their own ideal fertility outcomes.

Before the Two-Child Policy in 2016, having more than one child would bring an additional psychic cost into the utility function, the amount of which depends on whether the household is in a province that imposed a strict or relatively loose One-Child Policy. This penalty term for having excess birth is captured by the last part of the utility function, where the parameter $P_{jt}$ denotes the degree of penalty. This term is normalized to 0 after the year 2016, as the fertility restriction has been removed since then. Finally, I assume that this penalty term is paid in each period after the excess birth was given.

[9] Out of the two waves where people were asked about their ideal number of children, I use the answer in wave 2018 since I assume the after-policy-change answer would more credibly reflect people's fertility preferences than the earlier one in 2014.

The household flow utility is

$$U(c_{jt}^f, l_{jt}^f, c_{jt}^m, l_{jt}^m, b_{jt}, n_{jt}, \widetilde{n}_{jt}^f, \widetilde{n}_{jt}^m, \theta_{jt}) = \theta_{jt} u(c_{jt}^f, l_{jt}^f, n_{jt}, \widetilde{n}_{jt}^f) + (1-\theta_{jt}) u(c_{jt}^m, l_{jt}^m, n_{jt}, \widetilde{n}_{jt}^m) + \epsilon_{jbt},$$

which is a weighted sum of the individual utility by the Pareto weight that updates in each period. $\theta_{jt}$ here refers to the wife's Pareto weight within the family in decision-making. A discrete-choice-specific random preference shock is associated with the binary birth choice, which is assumed to be following the type I extreme value distribution.

## 4.2 Wage Process

I model an exogenous wage process for men and women separately, where everyone's income depends on a set of observable demographics, a permanent income process that follows an AR(1) process, and a measurement error. The equations below summarize the wage process.

$$\begin{aligned}
\log W_{it} &= \boldsymbol{Z}_{it}\boldsymbol{\beta} + u_{it} \\
u_{it} &= v_{it} + \epsilon_{it} \\
v_{it} &= \rho \cdot v_{it-1} + \xi_{it}, \quad \xi_{it} \sim N(0, \sigma_\xi^2) \\
\epsilon_{it} &\sim N(0, \sigma_\epsilon^2).
\end{aligned}$$

$\boldsymbol{Z}_{it}\boldsymbol{\beta}$ denotes the explained part of the wage process and $u_{it}$ stands for the unexplained part of the wage process. $v_{it}$ is the permanent component of the unexplained part that follows an AR(1) process. Finally, $\epsilon_{it}$ is a measurement error for the wage that follows a normal distribution. For women, I consider selection into work by assuming that the error term in the permanent component is correlated with the labor force participation decision as in the following equation.

$$\begin{aligned}
&Pr(L_{it} = 1) = Pr(\boldsymbol{X}_{it}\boldsymbol{\mu} + \eta_{it} > 0) = Pr(\eta_{it} > \alpha_{it} = -\boldsymbol{X}_{it}\boldsymbol{\mu}) \\
&\begin{pmatrix} \xi_{it} \\ \eta_{it} \end{pmatrix} \sim N(0, \begin{bmatrix} \sigma_\xi^2 & \sigma_{\xi\eta} \\ \sigma_{\xi\eta} & 1 \end{bmatrix}),
\end{aligned}$$

where $\eta_{it}$ is the error term for the labor force participation decision, and I assume it has a joint normal distribution with the error term in the permanent component of the wage process. The covariance term is specified using $\sigma_{\xi\eta}$, while $\sigma_\xi^2$ stands for the variance of the error term for the permanent component of the wage process. I normalize the variance of $\eta_{it}$ to be 1. $\boldsymbol{X}_{it}$ stands for the set of observed characteristics for deciding the labor participation for women. It is a superset of $\boldsymbol{Z}_{it}$ for the wage process, containing an additional characteristic for labor participation only that does not affect the wage determination.

For the wage equation specification, I assume the observed part includes age, age squared, and educational attainment. In addition, to capture the effect of the policy change on the wage level, I include a post-policy term, a term indicator of whether the couple is in a strict or loose province, and an intercept term between the post-policy and the strictness measure. This specification means to mimic the reduced-form evidence for the impact of the Two-Child Policy on the wage level in the empirical part. The equation below presents the detailed wage equation. $\beta_7$, therefore, captures the policy impact on the wage level.

$$
\begin{aligned}
\log W_{it} = \beta_1 t + \beta_2 t^2 + \beta_3 HighSchool_i + \beta_4 CollegeandHigher_i \\
+ \beta_5 Strict_{jt} + \beta_6 PostPolicy_{jt} \\
+ \beta_7 Strict_{jt} \times PostPolicy_{jt} + u_{it}.
\end{aligned}
$$

## 4.3 Childcare Costs

Having children entails both a time cost and a monetary cost in this model. First of all, a time cost is required to take care of a specific number of children in the household. I assume a nonparametric function of the childcare time cost in the number of children, which is assumed to be both increasing and concave. The couple shares the childcare time burden between them according to a certain rule. When allocating the childcare time burden between spouses, instead of assuming the couple optimizes the time input according to their wages, I assume the share of the burden is a constant in the population, which follows the social norm in China.

Let $\boldsymbol{X}_t = [x_t^f(n), x_t^m(n)]$ denote the childcare time input for raising $n$ children:

$$\boldsymbol{X}_t = [\tau Q(n), (1-\tau) Q(n)],$$

where $Q(n)$ is the childcare time needed for raising $n$ children, and $\tau$ is the share of the burden levied on women, which is a constant across the population in my model. Both the childcare time demand function $Q(n)$ and the share of burden parameter $\tau$ are pre-estimated directly using the time use data.

Secondly, there will be a monetary cost for raising $n$ children in the household, denoted by $C_t(n)$. In my model, I explicitly interpret this cost as the educational costs of raising children since educational costs are the main part of the monetary costs of raising children in China nowadays. I assume this cost changes over time to reflect the fact that education costs have been increasing in recent years rapidly, which is one of the main reasons that people are reluctant to have more children. Similar to the time cost of raising children, I model $C_t(n)$ as a nonparametric function of $n$, which is directly estimated from the expenditure data.

### 4.4 Pareto Weight Updating in Household Bargaining

The update of the Pareto weight $\theta$ aims to reflect the consequence of the changes in the relative wage and other household outcomes of the couple. Abstracting away from the formal modeling of any outside options, I assume the Pareto weight for the couple is updated in each period in a reduced-form way following the equation below.

$$\pi_t = \gamma_0 + \gamma_1(\widetilde{w}_{f,0} - \widetilde{w}_{m,0}) + \gamma_2[(w_{f,t} - \widetilde{w}_{f,t}) - (w_{m,t} - \widetilde{w}_{m,t})],$$
$$\theta_t = \frac{\exp(\pi_t)}{1 + \exp(\pi_t)},$$

where $\{\widetilde{w}_{f,0}, \widetilde{w}_{m,0}\}$ stands for the expected wage for the wife and the husband at marriage where the expectation is formed at marriage. $\{\widetilde{w}_{f,t}, \widetilde{w}_{m,t}\}$ stands for the expected wage for the wife and the husband at time $t$ where the expectation is formed at marriage. $\{w_{f,t}, w_{m,t}\}$ is the realized wage for the couple at time $t$.

The equation shows that the Pareto weight within the household depends on two parts: (1) expected wage difference at marriage time and (2) difference in realized wage shocks in the current period. The weighted sum of these three terms plus a constant term is then going through a transformation to be located between 0 and 1.

The expected wage difference at marriage time reflects how the initial conditions of the couple would affect the Pareto weight. The higher the wife's expected wage at marriage is compared to the husband's, the higher her Pareto weight would be during the marriage. If this marriage has a full commitment, then the Pareto weight should be fully pinned down by the initial conditions. However, in my model, I assume instead that the household has a limited commitment in the sense that the Pareto weight changes over time after marriage. In each period, the real wage will realize for the couple, and both of them will have a wage shock, which is the difference between the realized wage and the expected wage formed at marriage: $w_{g,t} - \widetilde{w}_{g,t}, g \in \{f, m\}$. The difference in the wage shocks will affect the Pareto weight in each period. For example, if the husband gets a pay raise in period $t$ that was not expected at the time of marriage, $w_{m,t} - \widetilde{w}_{m,t}$ will be a positive number, resulting in a negative number for the third term in the Pareto weight equation, and therefore leading a decrease in the Pareto weight for women.

This structure of the Pareto weight helps to incorporate the impact of the policy change on the Pareto weight. When the unexpected policy change happened in 2016, the difference in realized wage shocks would be negative since the gender wage gap was enlarged by the policy. This will lead to a lower Pareto weight for women, reflecting how a worsened labor market condition for them would result in a lower bargaining power within the household.

## 4.5 The Couple's Problem

Having defined the preliminaries, we can now summarize the couple's problem as solving a dynamic household model under a cooperative framework with limited commitment. In each period, they choose private consumption, working hours, and a binary choice of whether to give birth, subject to budget and time constraints.

The state and choice variables of the couple are:

$$\begin{aligned}\boldsymbol{\Omega_t} &= \{A^f_{jt}, E^f_j, E^m_j, M_j, P_{jt}, S_j, w^f_{jt}, w^m_{jt}, n_{j,t-1}, \widetilde{n}^f_{jt}, \widetilde{n}^m_{jt}, \theta_{jt}\}, \\ q_t &= \{c^f_{jt}, c^m_{jt}, h^f_{jt}, h^m_{jt}, b_{jt}\}.\end{aligned}$$

The state space consists of the age of the wife, the couple's education levels, their marriage year, the strictness of the One-Child policy for the province the couple is in, whether before or after the policy change, wages of the couple, the number of children from last period, the couple's ideal number of children respectively, and the Pareto weight of the wife within the household. The choice variables include the two private consumption levels and working hours, as well as the binary fertility choice.

The household utility is a weighted sum of the individual utility. The joint problem the couple solves is:

$$\begin{aligned}V_t(\boldsymbol{\Omega_t}) = \max_{q_t}\ & \theta_{jt} u(c^f_{jt}, l^f_{jt}, n_{jt}, \widetilde{n}^f_{jt}) + (1-\theta_{jt}) u(c^m_{jt}, l^m_{jt}, n_{jt}, \widetilde{n}^m_{jt}) + \epsilon_{jbt} \\ & + \beta E_t[V_{t+1}(\boldsymbol{\Omega_{t+1}})],\end{aligned}$$

subject to

$$\begin{aligned}c^f_{jt} + c^m_{jt} &= \Big(w^f_{jt} h^f_{jt} + w^m_{jt} h^m_{jt} - C_t(n_{jt})\Big) \cdot e(n) \\ l^g_{jt} + h^g_{jt} &= \bar{h}^g - x^g_{jt}(n_{jt}), \quad g \in \{f, m\}.\end{aligned}$$

As I abstract away from modeling saving decisions, the couple has a period-by-period budget constraint where $C_t(n_{jt})$ refers to the monetary cost of raising children and $e(n)$ is the economies of scale with the presence of $n$ children. The intuition is that when there are children present in the household, the total consumption that the couple can share between the husband and the wife will be less than when there is no child at all. In other words, this term allows for consumption by children, which should be increasing in the number of them.

The husband and the wife also have a time constraint, where the amount of time available for distributing between leisure and work equals the total time $\bar{h}^g$ subtracting the time needed for taking care of the children currently in the household $x^g_{jt}(n_{jt})$. The timeline for making decisions in each period consists of two steps: (1) given the number of children from the last period, the couple first decides on whether to have another child in this period and updates the number of children now accordingly, (2) then conditional on this updated number of children, choose the optimal private consumption and working hours.

# 5 Estimation: Machine Learning and Indirect Inference

This section discusses the details of the estimation of the structural model. The estimation procedure contains three parts: (1) presetting a group of parameters (Section 5.1), (2) pre-estimating the childcare costs and the wage process (Section 5.2), and (3) estimating the remaining structural parameters using the policy gradient methods from the reinforcement learning literature (Section 5.3.1) and indirect inference (Section 5.3.2). Identification arguments are provided in Section 5.3.2. Finally, Section 5.4 presents the estimation results and model fits, followed by Section 5.6 that shows the estimated Pareto weights over time.

## 5.1 Preset Parameters

A group of parameters of the model is set to values consistent with the empirical setting or drawn from the literature. Table 3 presents the preset parameters. The initial age corresponding to the period $t = 1$ is either 20 or the age at marriage, whichever is larger. Because the data is biannual, each period in the model corresponds to two years of life. The age in the last period is 40 since the possibility of having children is low for women after 40 in China. Therefore, a couple can have at most 11 periods, starting from age 20 to age 40 for the wife. For couples that married later than when the wife is 20, the periods are shorter for them. To limit the computation complexity of the problem, I assume the husband is two years older than the wife for all couples [10]. Hence, I only need to track the wife's age for each couple, reducing the dimension of the state space. For the economies of scale for children $e(n)$, following the literature, I assume that $e(n) = (n+1)^{-\frac{1}{2}}$. When there is no child present in the household, $e(n) = 1$. As $n$ increases, the consumption for the couple will be discounted more. Finally, I calibrate the discount factor $\beta$ to be 0.95 following the literature.

## 5.2 Pre-estimated Parameters

**Estimating the childrearing costs** I pre-estimate the childcare time cost function using the time-use module of the data and pre-estimate the childcare monetary costs using the household expenditure data. Since the number of children is discrete, I directly calculate the

[10] The difference between the age of the husband and the age of the wife has both the mean and the median equaling to 2.

Table 3: Preset Parameters of the Model

| Parameters | Value |
|---|---|
| Initial age | max(20, marriage age) |
| Years in each period | 2 |
| Age at last period | 40 |
| Discount factor ($\beta$) | 0.95 |
| Economies of scale for children $e(n)$ | $(n+1)^{-\frac{1}{2}}$ |
| Husband's age $t_m$ | wife's age + 2 |

average hours couples spend on taking care of children, conditional on the different number of children that they have. I use the conditional means as the childcare time needed for raising a certain amount of children. As for the share of the childcare burden borne by women, I calculate the average of the shares in the data and assume this value of the share is the social norm in China that does not vary across couples. Similarly, I take advantage of the household expenditure data where the educational costs of raising children are available, and I calculate the average educational costs conditional on each number of children present in this household. To take into consideration the fact that education costs have been rising dramatically over recent years, I separately calculate the average educational costs before and after the policy change.

Table 4 presents the estimated childcare time and monetary costs. The number of hours needed for childcare increases at a decreasing rate in the number of children. Notice that when there is no child inside the family, the hours needed for childcare is non-zero. This is because I pull together the hours for housework and taking care of the children. Therefore, 22 hours are the hours needed for housework when there are no children [11]. The share of the childcare time burden levied on women is estimated to be 0.73. I assume this number is constant across couples, and people just follow this social norm when distributing the childcare time between the wife and the husband. The second part of the table presents the estimated yearly educational costs for raising children. Similar to the time costs, the monetary costs also increase with the number of children at a decreasing speed. Meanwhile, the after-policy-change values are much larger than the before-policy-change values, consistent with the assumption that educational costs have been rising rapidly over the past ten years.

**Estimating the wage equation** I estimate the parameters of the wage process for the wife and the husband separately using moments of their hourly wage distributions from the

[11] The data pattern suggests that the hours spent on housework also increase with the number of children. Hence, I add together the hours of housework and taking care of children to be the total home production hours and interpret this total number as the time costs for raising children.

Table 4: Estimated Time and Monetary Costs of Raising Children

| | Time Costs for Raising Children (Hours per Week) | |
|---|---|---|
| Number of children | Hours Needed for Care | Share of Burden for the Wife |
| 0 | 22 | |
| 1 | 39.75 | 0.73 |
| 2 | 52.5 | |
| 3 | 51 | |
| | Educational Costs for Raising Children (CNY per Year) | |
| Number of children | Before Policy Change | After Policy Change |
| 0 | 0 | 0 |
| 1 | 4357.15 | 6313.69 |
| 2 | 4463.89 | 7302.47 |
| 3 | 5005.25 | 7523.31 |

CFPS. When estimating women's wage process, I adopt the Heckman two-step procedure to account for the selection of women into the workforce. To identify the parameters of the wage process, I use the stated preference for traditional values as the instrument for the participation equation [12]. The stated preference reflects whether the couple holds the traditional values in China that women should stay at home instead of working outside, which would affect women's decision to work but should be excluded from the wage process [13]. To better satisfy the exclusion restriction, I use the husband's answer instead of the wife's. The coefficient for this variable in the participation equation should be negative as agreeing to traditional values decreased women's probability of working while not affecting the wage they receive.

I use several moment conditions to estimate the parameters of the unexplained part of men's and women's wage process. At least three periods of panel wage data are needed to estimate men's parameters, and at least four periods of panel wage data are needed to estimate women's parameters. The detailed moment conditions used for estimation are listed in Appendix C. Table 5 presents the wage process estimation results. Row (1)-(4) presents the estimated parameters for the unexplained part of the wage process. According to the results, men have a more volatile permanent wage process than women since women's variance of permanent wage shocks is smaller than that of men. They, however, have a similar measurement error distribution, having variances close to each other. For women, the covariance of wage and work shocks is 0.4273, which is quite large. According to the

[12] In the survey, the couple was asked how much they agreed with the statement: "*Men should work outside to support the family while women should stay inside the household to take care of the family.* "

[13] I code "strongly agree" and "agree" as 1 and "neither agree nor disagree", "disagree", and "strongly disagree" as 0.

results, selection into work exists significantly for women.

For the explained part of the wage process, the hourly wage is increasing concavely in age for both men and women. Higher educational levels bring higher wage levels, non-surprisingly. The last three rows report the coefficients for the policy impact on the wage level. Firstly, being in the strict provinces imply a lower wage for both women and men. Secondly, the coefficients for post-policy mainly capture the time trend of the wage level, which are both significantly positive for women and men. Lastly and most importantly, the coefficient for the interaction term between being in strict provinces and post-policy, $\beta_7$, shows the policy impact on the wage level for both women and men. The coefficient is insignificantly positive for women (0.0461) while significantly positive for men (0.0843), confirming the reduced-form finding that the policy change has enlarged the gender wage gap, as the policy has a more positive impact on men compared to that on women.

Table 5: Wage Equation Estimation Results

| Parameters | Notations | Women | Women (s.e.) | Men | Men (s.e.) |
|---|---|---|---|---|---|
| AR(1) coefficient for permanent wage | $\rho$ | 0.7144 | (0.0636) | 0.7416 | (0.0740) |
| Variance of permanent wage shocks | $\sigma_\xi^2$ | 0.0168 | (0.0340) | 0.1070 | (0.0325) |
| Measurement error variance | $\sigma_\epsilon^2$ | 0.1855 | (0.0215) | 0.1858 | (0.0219) |
| Covariance of wage and work shocks | $\sigma_{\xi\eta}$ | 0.4284 | (0.0674) | | |
| | | | | | |
| Age $t$ | $\beta_1$ | 0.1351 | (0.0184) | 0.0896 | (0.0125) |
| Age squared $t^2$ | $\beta_2$ | -0.0099 | (0.0013) | -0.0067 | (0.0009) |
| High School Graduate | $\beta_3$ | 0.2642 | (0.0302) | 0.1552 | (0.0169) |
| College and Higher | $\beta_4$ | 0.5343 | (0.0435) | 0.4397 | (0.0164) |
| | | | | | |
| Strict Province | $\beta_5$ | -0.0628 | (0.0271) | -0.0934 | (0.0204) |
| Post Policy | $\beta_6$ | 0.3671 | (0.0325) | 0.4277 | (0.0182) |
| Strict Province $\times$ Post Policy | $\beta_7$ | 0.0461 | (0.0355) | 0.0843 | (0.0277) |

Note: This table presents the estimated parameters for the wage process of women and men separately. The first section shows the parameters for the unobserved part of the wage process, and the second section shows the parameters for the observed characteristics of the wage process. Finally, the last section presents the parameters for the policy-related covariates of the wage process.

### 5.3 Indirect Inference with Reinforcement Learning

I estimate the remaining nine structural parameters by solving the dynamic model using policy gradient methods for the reinforcement learning literature and obtaining parameter estimates using indirect inference. In this section, I first discuss how I adopt the policy gradient methods in the reinforcement learning literature to obtain the optimal policy function in Section 5.3.1. I then talk about the details of the indirect inference method I use for estimating the structural parameters in 5.3.2.

### 5.3.1 Policy Gradient: Solving the Optimal Policy

To deal with the large state space problem, instead of using the traditional backward induction method to solve for the optimal policy function, I use reinforcement learning to solve for the policy function. Specifically, I adopt policy gradient methods to directly learn the policy function, which allows me to avoid discretizing the state space and calculating the expected value function in each period. Reinforcement learning (RL) addresses the problem of how agents should learn to take actions to maximize cumulative reward through interactions with the environment. Policy gradient methods are among the most successful model-free RL algorithms due to their adaptability and straightforward implementation schemes (Sutton et al., 1999; Kakade, 2001; Silver et al., 2014). This method has been widely applied to various fields with huge success, including robotics and motor skills (Peters and Schaal, 2006; Peters and Schaal, 2008), medical applications (Jin et al., 2023; Hua et al., 2022), and many other empirical practices. However, the application of this method to economic models is still limited.

Let $\boldsymbol{\omega_t}$ denote the vector of the state space in time $t$. The key to policy gradient methods is to parameterize the choice as a probabilistic function of state variables: $\pi_{\boldsymbol{\lambda}}(b_{jt}|\boldsymbol{\omega_t})$, and maximize the lifetime value $V$ by adjusting the parameters $\boldsymbol{\lambda}$ of the policy $\pi_{\boldsymbol{\lambda}}$ by following the gradient $\nabla_{\boldsymbol{\lambda}} V(\pi_{\boldsymbol{\lambda}})$, which is referred to as a policy gradient. It is intractable to directly compute $\nabla_{\boldsymbol{\lambda}} V(\pi_{\boldsymbol{\lambda}})$. Fortunately, we have the policy gradient theorem (Sutton et al., 1999) stating that

$$\nabla_{\boldsymbol{\lambda}} V(\pi_{\boldsymbol{\lambda}}) = E_{\boldsymbol{\omega}}\Big[E_b[Q^{\pi_{\boldsymbol{\lambda}}}(\boldsymbol{\omega}, b)\nabla_{\boldsymbol{\lambda}}\log\pi_{\boldsymbol{\lambda}}(b|\boldsymbol{\omega})]\Big],$$

where $Q^{\pi_{\boldsymbol{\lambda}}}(\boldsymbol{\omega}, b)$ is the state-action value function of the policy $\pi_{\boldsymbol{\lambda}}$. In other words, the gradient of the value function can be transferred to the gradient of the policy function, $\nabla_{\boldsymbol{\lambda}}\log\pi_{\boldsymbol{\lambda}}(b|\boldsymbol{\omega})$, which becomes tractable now. The expectations are taken over the choice probability of $b$ and the transitional probability of state variables. In the algorithm of searching for the optimal $\boldsymbol{\lambda}$, I use Monte Carlo simulation to calculate the expectations.

In my model, I have a binary decision of giving birth $b_t$, and conditional on $n_t$, couples make private consumption and leisure decisions. Since consumption and leisure decisions do not involve inter-temporal aspects and have closed-form solutions, here I only need to model the binary birth decision. I use the logistic regression to model the birth decision conditional on state variables. Let $\varphi_{jt}$ denote the conditional probability of $b_{jt} = 1$. We have:

$$\varphi_{jt} \equiv Pr(b_{jt} = 1|\boldsymbol{\omega}_{jt}; \boldsymbol{\lambda}) = \frac{\exp(\boldsymbol{\omega}_{jt}\boldsymbol{\lambda})}{1 + \exp(\boldsymbol{\omega}_{jt}\boldsymbol{\lambda})},$$

where $\boldsymbol{\omega_t} = (A^f_{jt}, E^f_j, E^m_j, M_j, P_{jt}, S_j, w^f_{jt}, w^m_{jt}, n_{j,t-1}, \widetilde{n}^f_{jt}, \widetilde{n}^m_{jt}, \theta_{jt})$. With this functional form, we can compute the gradient as the following:

$$\nabla_{\boldsymbol{\lambda}} \log \prod_{t>=1} (\varphi_{it}) = \sum_{t=1}^{T} \nabla_{\boldsymbol{\lambda}} \log(\varphi_{it}),$$

Specifically,

$$\nabla_{\boldsymbol{\lambda}} \log(\varphi_{jt}) = \begin{cases} (-\varphi_{jt} + 1)\boldsymbol{\omega_t} & \text{if } b_{jt} = 1 \\ -\varphi_{jt}\boldsymbol{\omega_t} & \text{if } b_{jt} = 0 \end{cases}.$$

After obtaining the functional form for the policy gradient, I can start estimating the policy function parameters $\boldsymbol{\lambda}$ to maximize the lifetime value. I discuss the details and proof of this method in the second chapter of my dissertation. Here, I present the steps of the SGD algorithm in Algorithm 1. At each iteration, we first draw a random batch of X couples from the data. Then, given the current value of $\lambda$, we sample the X individuals' dynamic birth decisions from the decision model and simulate their whole lifetime states $\boldsymbol{\omega_{jt}}$ given the state's transitional model and the chosen birth decisions. Next, we compute the gradient of the log-policy $\nabla_{\lambda} \log(\prod_{t>=1} \pi(b_{it}|\boldsymbol{\omega}_{it}; \boldsymbol{\lambda}))$ and calculate the average $\nabla_{\boldsymbol{\lambda}} V(\boldsymbol{\lambda}_q)$ over the X samples. Lastly, we update $\boldsymbol{\lambda}$ using the calculated gradient of the value function multiplied by a pre-chosen step size. We select the optimal policy parameter $\boldsymbol{\lambda}^*$ to be the one yielding the highest expected reward across all SGD iterations.

**Algorithm 1** Stochastic Gradient Descent for optimizing $\boldsymbol{\lambda}$

**Input:** initial value $\boldsymbol{\lambda}_0$, step size $s_q$, data $D$, deep parameters $\boldsymbol{\gamma}$, transitional model parameter $\boldsymbol{\phi}$, batch size X.
**Initialize:** $\boldsymbol{\lambda}_1 \leftarrow \boldsymbol{\lambda}_0$
**for** $q = 1, \ldots, Q$ **do**
    Sample a batch of X couples from the whole data
    **for** $t = 1, \ldots, T$ **do**
        Sample $\boldsymbol{\omega}_{it}$ and $b_{it}$ conditional on $D$, $\boldsymbol{\phi}$, and $\boldsymbol{\lambda}_q$
    **end for**
    Calculate the lifetime value $V(\boldsymbol{\lambda}_q)$
    $\nabla_{\boldsymbol{\lambda}} V(\boldsymbol{\lambda}_q) \leftarrow V(\boldsymbol{\lambda}_q) \nabla_{\boldsymbol{\lambda}} \log(\prod_{t>=1} \pi(b_{it}|\boldsymbol{\omega}_{it}; \boldsymbol{\lambda}))$
    $\boldsymbol{\lambda}_{q+1} \leftarrow \boldsymbol{\lambda}_q + s_q \nabla_{\boldsymbol{\lambda}} V(\boldsymbol{\lambda}_q)$
**end for**
**Output:** $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}_{q^*}$

I estimate two policy functions, one for the situation where agents behave as if there were no Two-Child Policy carried out in 2016, the other for the after-policy-change situation. Combining the first policy function for the pre-policy-change periods and the second policy function for the after-policy-change periods yields the optimal policy function for an agent's life path. For the hyperparameters, I set the initial value $\lambda_0 = 0$ for all parameters and choose the step size to be $s_q = 1e^{-4}$. Finally, I run the SGD algorithm 10,000 times. Figure D.2 shows the convergence pattern of lifetime value over the updating of $\lambda$ during the iterations.

Table 6 below presents the two sets of results for the optimized policy function parameters $\lambda^*$. The first column shows the estimated policy function parameters for the pre-policy-change periods, while the second column shows the estimated parameters for the post-policy-change periods. The parameters show some patterns for how the birth decisions are related to state variables. For example, $\lambda_1 < 0$ for both policy functions, suggesting that the probability of giving birth decreases with women's age, which makes intuitive sense. $\lambda_7 < 0$ for the pre-policy-change periods, which states that couples in strict provinces have a lower probability of giving birth. $\lambda_7$ almost equals to 0 for the post-policy-change periods, as the policy shock is larger for those in strict provinces, therefore resulting in a higher probability of giving birth after the policy relaxation. Using the estimated policy functions, we can then move to the indirect inference part to estimate the deep structural parameters.

Table 6: Estimated Parameters for the Optimal Policy Function

| Variables | Symbol | Without Policy Change Estimates | After Policy Change Estimates |
|---|---|---|---|
| Intercept | $\lambda_0$ | -2.202 | -1.979 |
| $t$ | $\lambda_1$ | -0.152 | -0.068 |
| $E_1^f$ | $\lambda_2$ | -0.033 | -0.095 |
| $E_2^f$ | $\lambda_3$ | 0.078 | -0.109 |
| $E_1^m$ | $\lambda_4$ | 0.072 | -0.121 |
| $E_2^m$ | $\lambda_5$ | 0.020 | -0.068 |
| $tM$ | $\lambda_6$ | 0.314 | 0.121 |
| $S$ | $\lambda_7$ | -0.145 | -0.003 |
| $w_t^f$ | $\lambda_8$ | -0.142 | 0.008 |
| $w_t^m$ | $\lambda_9$ | -0.117 | -0.059 |
| $N_{t-1}$ | $\lambda_{10}$ | -1.633 | -1.193 |
| $N_1$ | $\lambda_{11}$ | 0.710 | 0.615 |
| $N_2$ | $\lambda_{12}$ | 1.159 | 0.713 |
| Pareto weight$_t$ | $\lambda_{13}$ | -0.053 | 0.084 |

Note: This table shows the estimated optimal policy function parameters $\lambda^*$. There are 14 parameters in total, with 13 for each of the state variables and 1 for the intercept. The first column presents the corresponding state variables. The third column shows the parameters for the case where there is no Two-Child Policy introduced in the agent's lifetime, and the last column shows the parameters for the after-policy change periods.

### 5.3.2 Indirect Inference

After obtaining the approximate optimal policy function, I estimate the remaining 9 structural parameters using indirect inference (Gourieroux et al., 1993). As in all collective models, the key challenge of identification is to identify the Pareto weight parameters. There are two sets of data moments that are useful for this purpose. Firstly, I observe panel data of working hours and private leisure time allocations for the couple over time. The proportion of the wife's working hours out of the couple's total working hours is related to the relative bargaining power of the couple within the household. Together with the wage process of the couple over time, we can therefore identify the Pareto weight parameters. Secondly, I observe the actual number of children the couple has and both the wife's and the husband's ideal number of children. Using the post-policy-change periods' data, the distance between the actual number of children and the couple's ideal number of children helps identify the Pareto weight parameters. Intuitively, if the actual number of children is closer to the husband's ideal number of children than to the wife's, then the husband has a higher Pareto weight for intra-household decisions, other things equal [14]. Exploiting data moments regarding these two aspects enables me to identify the Pareto weight parameters.

To identify the penalty the couple had to pay under the One-Child Policy if they had a second birth, I use the excess fertility rate and the ideal number of children for the couple before the policy relaxation. The intuition is similar to the research design argument in the reduced-form part: conditional on having the same ideal number of children, provinces, where the couple has less excess fertility rate, should have a stricter One-Child Policy (a larger penalty for excess birth) before. Finally, the couple's working hours and fertility choices after the policy change help identify the remaining utility function parameters since now the policy restriction has been removed. The fertility choice, for instance, should only depend on the preferences and Pareto weight within the household.

The detailed steps of indirect inference for the model are listed here. First, I solve the dynamic model for vectors of structural parameters using policy gradient methods, plugging in the pre-estimated parameters. I then simulate the choice path of couples given the introduction of the Two-Child Policy at different time periods of the life cycle. With the simulated choice path for each couple, I compute the corresponding profiles and adopt the Method of Simulated Moments to find an optimal set of parameters $\hat{\Pi}$ to minimize

$$(\hat{\phi}_{data} - \phi_{sim}(\Pi))'\mathcal{F}(\hat{\phi}_{data} - \phi_{sim}(\Pi)),$$

[14] Before the policy relaxation, there are strict binds on the number of children that a couple can have. Therefore, it is tricky to use pre-policy-change fertility choice data to identify the Pareto weight parameters since it would be nontrivial to distinguish between policy restrictions and results from Pareto weights.

where $\mathcal{F}$ is a symmetric and positive semi-definite weighting matrix, which is set to be the inverse of the estimated variance-covariance matrix of the parameters of the auxiliary model. For the optimization process, I combine extensive grid search and Nelder-Mead optimization algorithm to find optimal parameters.

For the auxiliary models used for estimation, I use a total of 58 statistics, including both simple data moments and regression coefficients. Table 7 lists some examples of the selected moments. For time-use data, I include the average working hours for both men and women, as well as the wife's proportion of working hours out of the couple's total working hours to help identify the utility function parameters and Pareto weight parameters. Average excess fertility rates, before and after the policy change and in loose and strict provinces, are included to help identify the penalty parameters for excess birth. I also include the average number of children over the years during the marriage in loose and strict provinces to capture the dynamic birth decisions people make throughout their lifetime, helping to identify the utility parameters and the penalty parameters. Finally, I include the average birth probability conditional on different combinations of the ideal number of children that the wife and the husband have, in strict and loose provinces and before and after the policy change to help identify the penalty parameters and the Pareto weight parameters [15].

Table 7: Part of the Selected Moments

| Moment Description | Data | Identification |
|---|---|---|
| **Time Use** | | Help identify the utility function parameters and Pareto weight parameters |
| Average working hours for women | 33.94 | |
| Average working hours for men | 48.77 | |
| Average women's working-hour proportion | 0.3718 | |
| **Excess Fertility Rate** | | |
| Before policy change, loose province | 0.5396 | Help identify the penalty parameters for excess birth |
| After policy change, loose province | 0.5515 | |
| Before policy change, strict province | 0.2857 | |
| After policy change, strict province | 0.4137 | |
| **Average Number of Children** | | Help identify utility parameters and the penalty parameters |
| Over the years during the marriage | | |
| Before and after policy change | | |
| In loose and strict provinces | | |

The second part of the auxiliary models features the regression coefficients of several

[15] Specifically, the average birth probability conditional on $N_1 = 2$ and $N_2 = 2$ before the policy change helps identify the penalty parameters; the average birth probabilities conditional on $(N_1 = 1, N_2 = 2)$ and $(N_1 = 2, N_2 = 1)$ after the policy relaxation help identify the Pareto weight parameters.

models that I try to match using the real and the simulated data. The first two sets of the regression coefficients are matched to help identify the Pareto weight parameters:

$$\frac{w^f_{jt}}{(w^f_{jt} + w^m_{jt})} = \zeta_0 + \zeta_1 w^d_{j,0} + \zeta_2 w^d_{j,t} + \zeta_3 w^f_{j,t} + \zeta_4 w^m_{j,t} + \zeta_5 n_{j,t} + \nu_{1,j,t}, \tag{5.1}$$

where $w^d_{j,0}$ stands for the initial wage difference at marriage time, and $w^d_{j,t}$ stands for the wage shock difference at time $t$. As mentioned in the identification arguments, the proportion of the wife's working hours out of the couple's total working hours depends on the Pareto weight of the wife within the household. As it also depends on the wage levels of the wife and the husband and the number of children present at home, I include these three terms in the regression. What's more, since the Pareto weight equation involves the initial wage difference and the difference in the realized wage shocks for the couple, I include both of these elements in the regression to identify the parameters in the Pareto weight equation. In the estimation process, I match the coefficients $\{\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5\}$ from the regressions above.

Other than the regression above, I also run a difference-in-differences regression of fertility as the following:

$$\begin{aligned} N_{jt} = \zeta_6 + \zeta_7 \widetilde{n}^f_{j,t} + \zeta_8 \widetilde{n}^m_{j,t} + \zeta_9 age_{j,t} + \zeta_{10} posP_{j,t} + \zeta_{11} strict_{j,t} + \zeta_{12} posP_{j,t} \times strict_{j,t} \\ + \zeta_{13} w^f_{j,t} + \zeta_{14} w^m_{j,t} + \nu_{4,j,t}, \end{aligned} \tag{5.2}$$

where the regression is run only on a subgroup of the population with $t >= 10$ so that they must have finished fertility already. The coefficient of the interaction term indicates the effect of the Two-Child Policy on the number of children people have. Matching the coefficients $\{\zeta_7, \zeta_8, \zeta_9, \zeta_{10}, \zeta_{11}, \zeta_{12}, \zeta_{13}, \zeta_{14}\}$ helps identify both the penalty term for excess birth and the utility parameters.

## 5.4 Estimation Results and Model Fit

Table 8 presents the estimation results for the 9 structural parameters. Standard errors in the table are computed by block bootstrap with 100 replications. The utility function parameters are presented in the first subsection of the table. People derive a relatively larger marginal utility from the log of consumption (7.8209) than from the log of leisure (5.0844). Meanwhile, there will be a significant utility loss from not reaching their ideal number of children (2.8429). The couple, therefore, has the motivation to give birth until they reach their ideal number of children and stop right there. The second set of results for the Pareto weight parameters suggests that the wage shock difference has a larger impact on the Pareto weight (0.1814) than the initial wage difference (0.1239). Finally, the last subsection presents the policy parameters, where the penalty on excess birth in strict provinces (1.3732) is much larger than the penalty on excess birth in loose provinces (0.7413), which is consistent with the story that couples in strict provinces were more stringently regulated before the policy relaxation than those in loose provinces.

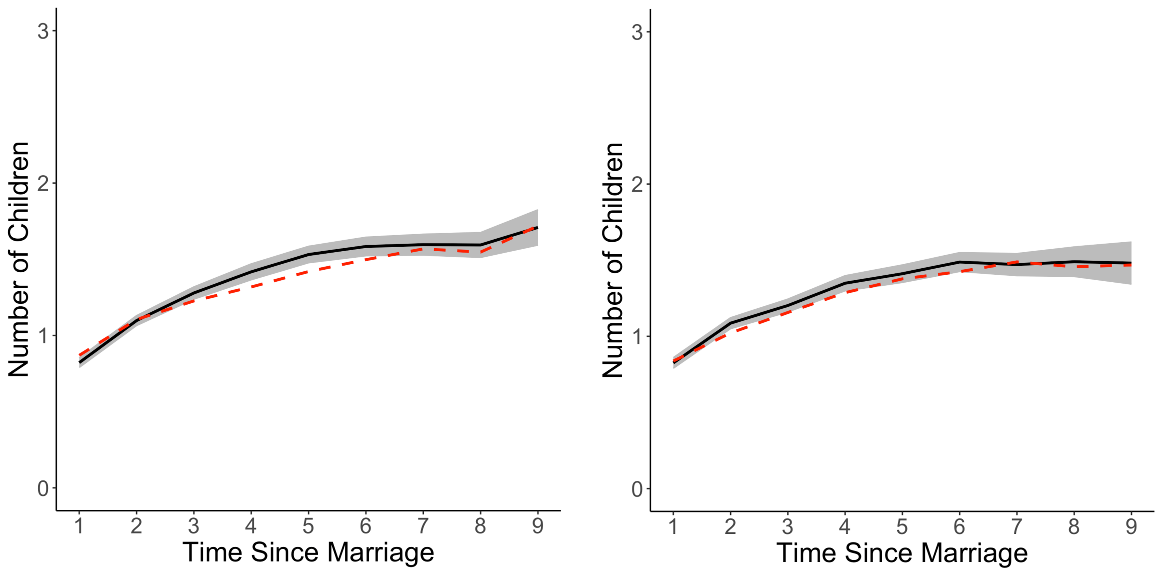
Table 8: Estimation Results for Structural Parameters

| Parameters | Symbol | Estimates | Standard Errors |
|---|---|---|---|
| Utility function parameters | | | |
| Utility from $\ln(c)$ | $\gamma_1$ | 7.8209 | 0.2348 |
| Utility from $\ln(l)$ | $\gamma_2$ | 5.0844 | 0.2582 |
| Dis-utility from not reaching ideal number of children | $\gamma_3$ | 2.8429 | 0.0579 |
| Pareto weight parameters | | | |
| Constant in Pareto weight | $\gamma_4$ | 0.0321 | 0.0239 |
| Initial wage difference | $\gamma_5$ | 0.1239 | 0.0188 |
| Wage shock difference | $\gamma_6$ | 0.1814 | 0.0197 |
| Policy parameters | | | |
| Penalty on excess birth in strict provinces | $p_1$ | 1.3732 | 0.4125 |
| Penalty on excess birth in loose provinces | $p_2$ | 0.7413 | 0.4380 |

Note: This table presents the estimation results for the 8 structural parameters. The first subsection shows the 3 estimated parameters in the utility function. The second subsection shows the 3 estimated parameters in the Pareto weight equation. Finally, the last subsection presents the estimated penalty terms for having a second child before the policy relaxation in both strictly and loosely regulated provinces. Standard errors in the table are obtained by bootstrapping the sample 100 times.

Before going to the counterfactual analysis, It is crucial to validate the estimation results. A starting point is to check the model fit for the targeted moments . Firstly, I check the model fit for people's fertility behavior. Figure 5 presents the model fit for the average number of children over years after getting married, both for couples in loose and strict provinces. These statistics show how well the estimated model is able to capture the dynamics of couples' fertility choices over their life cycles. Each time point in the graph denotes two years starting from their marriage period until age 40. The red dashed line is the average number of children using the simulated data, while the black solid line is the average number of children using the real data, with the gray shaded area being the 95% confidence interval for the data sample uncertainty. The graph shows that the simulated average number of children mostly lies inside the confidence band of the data moments for both of the sub-figures, indicating that the model is good at generating the fertility behavior over time as observed in the data, both in strict and loose provinces.

In addition to the fertility behaviors, I also provide evidence for the fitting of the time choice variables. Table 9 presents the model fit for the targeted moments of working hours. The second column in the table shows the data moments with the bracket under each number




(a) Loose Province

(b) Strict Province

Figure 5: Model Fit for the Average Number of Children (In-Sample)

showing the 95% confidence interval for each data moment, while the third column presents the corresponding simulated moments. The table shows that the simulated average working hours of men and women are quite close to the average working hours in the data. The third row shows the average proportion of the wife's working hours over the sum of the wife's and husband's working hours. Although the simulated average proportion lies outside of the 95% confidence interval of the data moment, it is reasonably close to the upper bound.

The second half of the table presents the fitting results for the regression for the proportion of the wife's working hours out of the couple's total working hours on a series of explanatory variables, as shown in equation 5.1. First of all, we notice that the coefficients generated using the data make sense. The higher the Pareto weight of the wife is, the more leisure she will take, and the less she will work. Therefore, we should expect the coefficients for the initial wage difference and the wage shock difference to be negative, which is indeed the case (-0.014, -0.009). Meanwhile, the wife's proportion of working hours is positively correlated with her own wage (0.016) and negatively correlated with the husband's wage (-0.008). Finally, the wife's proportion of working hours should decrease in the number of children (-0.028), since more childcare burden is levied on the wife, leading to fewer available total hours to be distributed between work and leisure. The third column of the table presents the corresponding regression coefficients using the simulated data set. According to the results, almost all coefficients generated from the data lie within the 95% confidence interval of the coefficients generated by the real data. Combining the results in Table 9 together, we can conclude that the estimated model does a relatively good job of recovering couples' time choices.

Apart from these two sets of model fits, more model-fit results can be found in Appendix D, where Table D.8 presents the real and simulated completed fertility and excess fertility rate, Table D.9 shows the real and simulated coefficients for the Difference-in-Differences regression of fertility, and Figure D.3 (a) presents the real and simulated birth probability conditional on different sets of the ideal number of children. The results show that most of the simulated moments lie within the 95% confidence interval of the data. indicating a good fit in general.

Table 9: Model Fit: Empirical and Simulated Moments for Time Use (In-Sample)

| **Moment Description** | **Data** [95% CI] | **Model** |
|---|---|---|
| **Data moments** | | |
| Average working hours for women | 34.47 [32.97, 34.90] | 34.46 |
| Average working hours for men | 48.89 [47.92, 49.61] | 48.90 |
| Average proportion of working hours for women | 0.3726 [0.3617, 0.3820] | 0.4099 |
| **Working-hour proportion regression** | | |
| Coefficient for $w^d_{j,0}$ | -0.014 [-0.028, 0] | -0.011 |
| Coefficient for $w^d_{j,t}$ | -0.009 [-0.018, 0] | -0.018 |
| Coefficient for $w^f_{j,t}$ | 0.016 [0.006, 0.026] | 0.007 |
| Coefficient for $w^m_{j,t}$ | -0.008 [-0.017, 0] | -0.006 |
| Coefficient for $N_{j,t}$ | -0.028 [-0.042, 0.013] | -0.043 |

## 5.5 Validation of the Estimation Results

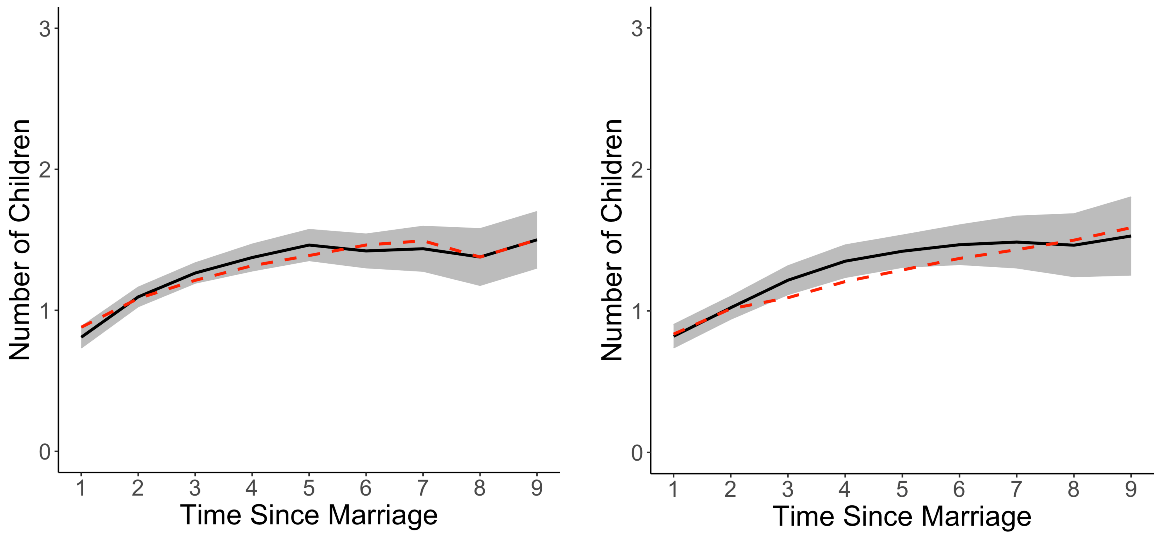
This section provides the out-of-sample validation for the estimation results [16]. Using the estimated parameters, I sample the life cycles for the couples in the validation set and generate the same moments and statistics as in the main sample. Figure 6 shows the model fit for the average number of children over time, grouped by strict and loose provinces. The two sub-figures show that the simulated number of children over time is reasonably close to the real values, suggesting the estimated model behaves well for the validation sample. Table 10 shows the empirical and simulated moments for working hours. Similar to the results for the main sample, the simulated average working hours of the wife and the husband are close to the true values, while the simulated working-hour ratio for the wife is a little bit higher than the confidence interval but by a reasonable amount. For the working-hour proportion regression, the signs of the simulated coefficients are all correct, with the values being a little bit off. Out of all the coefficients, the coefficient for the number of children is within the 95% confidence interval of the true coefficient.

Apart from these two sets of model fits, Table D.10–D.11 in Appendix D provides more out-of-sample model fits for fertility and time choices, where the former presents the real and simulated completed fertility and excess fertility rate and the latter shows the real and simulated coefficients for the Difference-in-Differences regression of fertility. Figure D.3 (b) presents the real and simulated birth probability conditional on different sets of the ideal number of children. The three sets of results show that most of the simulated moments lie within the 95% confidence interval of the real values, suggesting a good fit for the validation sample in general.

## 5.6 Estimated Pareto Weights

Now that we have estimated the parameters in the Pareto weight equation, we can calculate the average Pareto weights over time implied by the model. Figure 7 shows the estimated average Pareto weights from the year 2010 to the year 2020 for couples in strict and loose provinces. As shown in the figure, the Pareto weights started at around 0.4 for the wife in both groups of provinces, with women in strict provinces having a slightly higher Pareto weight than those in loose provinces before the policy change, as the initial gender wage gap was smaller in strict provinces than in loose provinces. The Pareto weight decreased sharply after the policy change in 2016 for couples in strict provinces. This change reflects the fact that the policy has a negative impact on women's Pareto weight. The decrease in

[16] When estimating the model, I only use 80% of the sample and leave the rest 20% to conduct validation tests.




(a) Loose Province

(b) Strict Province

Figure 6: Model Fit for the Average Number of Children (Out-of-Sample)

the Pareto weight was larger for couples in strict provinces since they went through a larger policy shock in 2016. After the policy change, the Pareto weight continued going down and reached below 0.3 in the year 2020 for the wife in strict provinces. This result hints at women's worse intrahousehold allocations after the policy change due to the significant decrease in their bargaining power within the households.

# 6 Counterfactual Analysis

Given the structural estimates, in this section, I start with analyzing how the Two-Child Policy affected people's choice behaviors (Section 6.1). I then conduct welfare and counterfactual analysis to (1) quantify the welfare impact of the Two-Child Policy on both genders and the degree to which it affected gender inequality (Section 6.2) and (2) show the alternative welfare impact of the policy if the effect through the wage channel is shut down (Section 6.3) and (3) evaluate various public policies to improve couples' welfare in the post-policy era and boost fertility at the same time (Section 6.4).

## 6.1 The Impact of the Two-Child Policy on Fertility and Hours

The starting point of the counterfactual analysis is to look at how the Two-Child Policy altered people's behaviors and choices. In particular, I compare couples' choices of fertility and time use with the existence of the Two-Child Policy and the counterfactual scenario where the policy was removed. Firstly, Figure 8 presents the average number of children

Table 10: Model Fit: Empirical and Simulated Moments for Time Use (Out-of-Sample)

| Moment Description | Data [95% CI] | Model |
|---|---|---|
| **Data moments** | | |
| Average working hours for women | 35.60 [33.61, 37.58] | 34.31 |
| Average working hours for men | 49.59 [47.63, 51.56] | 48.55 |
| Average proportion of working hours for women | 0.3753 [0.3564, 0.3941] | 0.4134 |
| **Working-hour proportion regression** | | |
| Coefficient for $w^d_{j,0}$ | -0.1073 [-0.1302, -0.0844] | -0.0381 |
| Coefficient for $w^d_{j,t}$ | -0.1076 [-0.1253, -0.0898] | -0.0446 |
| Coefficient for $w^f_{j,t}$ | 0.1076 [0.0875, 0.1277] | 0.0294 |
| Coefficient for $w^m_{j,t}$ | -0.1047 [-0.1232, -0.0863] | -0.0315 |
| Coefficient for $N_{j,t}$ | -0.0251 [-0.0583, 0.0080] | -0.0535 |

across the lifetime with and without the Two-Child Policy. The graph shows that the Two-Child Policy did have a positive impact on the number of children couples have, and it is unsurprising that the policy effect is larger for couples in strict provinces than those in loose provinces. We define complete fertility as the number of children couples have at the age of 38 and older, then the complete fertility for couples in strict provinces is 1.5650 under Two-Child Policy and 1.3043 under One-Child Policy. Therefore, the policy effect is 0.2607 for couples in strict provinces. Couples in loose provinces, on the other hand, have 1.6722 as the complete fertility under Two-Child Policy and 1.5477 as the complete fertility under One-Child Policy. Therefore, on average, couples in loose provinces will have 0.1246 more children over their lifetime under Two-Child Policy than if there was no policy relaxation. We should notice that, however, even under the Two-Child Policy, the lifetime fertility of couples is still far below two, which is the maximum number of children allowed by the

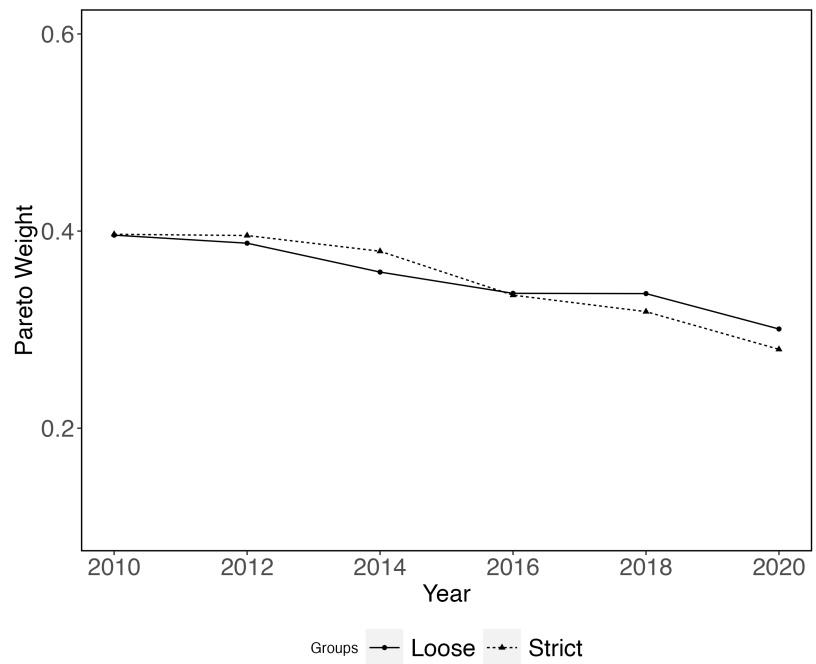
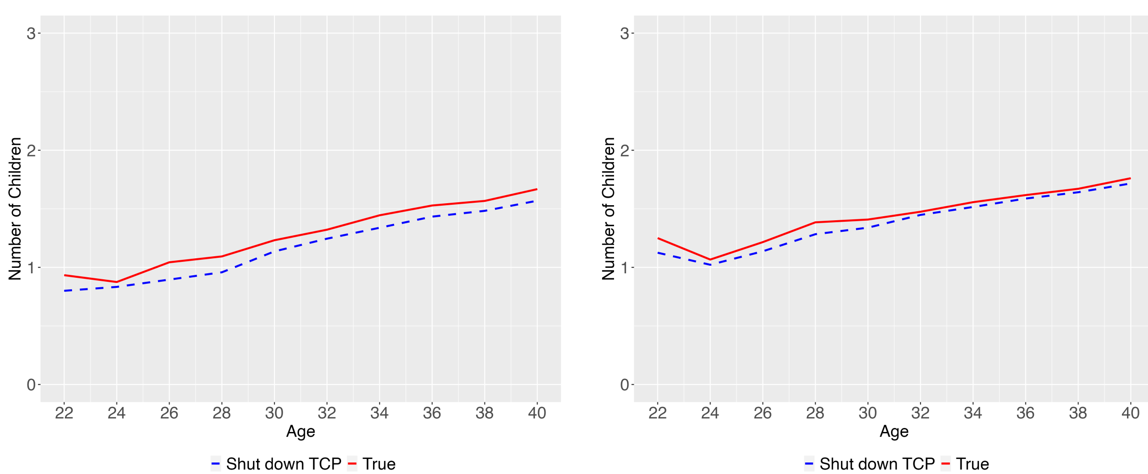



Figure 7: Estimated Pareto Weights Over Time in Strict and Loose Provinces

Two-Child Policy. Relaxing the rigid regulation on couples' fertility behavior is only the first step to encouraging more births. There are other reasons, economically, culturally, and socially, that are preventing couples from having more children.




(a) Strict Provinces

(b) Loose Provinces

Figure 8: Average Number of Children With and Without TCP

Next, I look at how the policy reform affected couples' choices of hours and resources. Table 11 presents the results for couples' choice of working hours, leisure hours, and women's ratio of leisure hours and consumption. It is noticeable that the policy reform decreased women's leisure hours while increased their working hours, suggesting that women's "double burden" problem becomes more severe after the Two-Child Policy became effective. They spend more time on childrearing than before due to the social norm in China, thus resulting in less leisure consumed. For men, on the other hand, they work less and consume more leisure with the Two-Child Policy. They do not need to bear the majority of the childcare burden, and because of the increase in their Pareto weights, they can consume more leisure now. This pattern can also been seen in the last subsection of the table. Women's ratio of leisure hours out of the total leisure hours decreased from 0.431 to 0.424, while their consumption ratio decreased from 0.387 to 0.380. These results are a direct consequence of women's lower Pareto weights after the policy change.

Table 11: Average Value of Time Use: With and Without TCP

| Variables | Two-Child Policy | No Policy Change |
|---|---|---|
| | Women | |
| Leisure hours | 20.52 | 20.96 |
| Working hours | 33.58 | 33.43 |
| | Men | |
| Leisure hours | 27.04 | 26.86 |
| Working hours | 49.90 | 50.18 |
| | Women's Ratio of Hours/Resources | |
| Leisure ratio | 0.424 | 0.431 |
| Consumption ratio | 0.380 | 0.387 |

Note: This table shows the mean values of couples' simulated time choices and allocation of resources. The columns show the scenarios when the TCP is present and when the TCP is shut down. The last subsection presents women's ratio of leisure hours (consumption) divided by the total leisure hours (consumption) consumed by the couple.

## 6.2 The Impact of the Two-Child Policy on Gender Inequality

To study how women's and men's welfare were affected by the Two-Child Policy, we first calculate their lifetime utility separately in the real scenario where there is Two-Child Policy, and then we calculate the counterfactual lifetime utility for them had there been no policy change in 2016. By comparing the lifetime utility with and without the policy change, we can quantify the welfare impact of the policy change on both genders. When conducting the welfare analysis, in addition to looking at the population as a whole, we also look at people with different demographic characteristics to study the potential heterogeneous welfare impact of the policy change on people. Figure 9 presents the real and counterfactual lifetime utility for various groups of people. Figures 9 (a) and (b) show that men's welfare is higher under the Two-Child Policy in both strict and loose provinces, with the improvement in welfare larger in strict provinces than that in loose provinces. This is due to the fact that men can more easily reach their ideal number of children after the policy relaxation and, at the same time, enjoy a boost in their bargaining power within the household brought about by the enlarger gender wage gap. Women, on the other hand, had a welfare loss after the policy change in strict provinces. Although they can also more easily reach their ideal number of children with the Two-Child Policy, they suffer from a worsened intrahousehold allocation due to the lower Pareto weight after the policy change. Therefore, the overall net welfare impact of the new policy is negative for them in strict provinces.

Figures 9 (c) and (d) compare the lifetime utility change across people with different characteristics in strict provinces. For women with high educational levels, the policy is especially bad for them since they typically do not want a second child. Therefore, there is not much gain from relaxing the birth constraint for them, and they will suffer from an enlarged wage gap after the policy change. The overall welfare impact is hence significantly negative. For women with low educational levels, although the lifetime utility change is still negative, the loss in welfare is smaller compared to those with high educational levels. This is because the gain from being able to reach the ideal number of children partly offsets the welfare loss from the lower Pareto weights led by an enlarged gender wage gap. For men, on the other hand, both of the educational groups have a positive welfare gain with the policy change, where men having high educational levels gain a little more than that having low educational levels.

Similar arguments go for people with different ideal number of children as shown in Figure 9 (d). It is intuitive that people who ideally want to have two children would benefit more from the Two-Child Policy than people that are satisfied with only one child. This is indeed the case as shown in the figure. For both men and women, people with the ideal number of children equaling to two has a more positive lifetime utility change with the policy compared to those that have the ideal number of children equal to one. There are some differences across the genders though. For men, even those people whose ideal number of children is one still benefit from the new policy since they would enjoy a higher Pareto weight leading to a better position within the household after the policy change. Therefore, even when they do not directly benefit from the fertility channel, they still obtain a higher welfare under the new policy. Women, in contrary, are having a lower welfare level regardless of the ideal number of children due to the wage and Pareto weight channels, with those whose ideal number of children being two losing less since they can at least benefit from the fertility channel .

Finally, Figures 9 (e) and (f) compare the lifetime utility change across people with different characteristics in loose provinces. For different educational levels, it is again intuitive that both men and women with low educational levels benefit more than those with high educational levels since the opportunity cost of their work time is lower than that of people with high educational levels and they tend to have a higher ideal number of children. For people with different ideal numbers of children, it is consistent with the intuition that both women and men with an ideal number of children being one suffer a loss from the Two-Child Policy, while those wanting two children gain from the new policy. The analysis for different subgroups of people shows the heterogeneous impact of the Two-Child Policy on couples with various characteristics. In summary, men, people with low educational levels, and people wanting two children benefit more from the new policy.

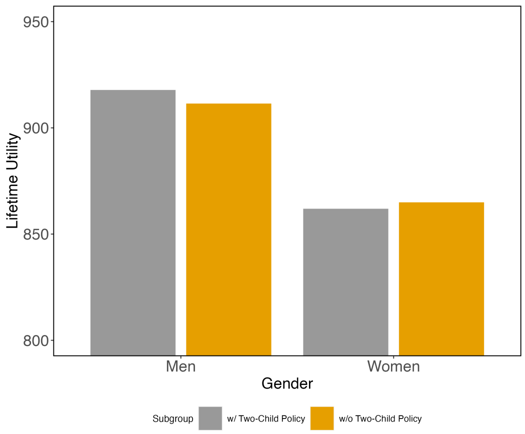
To convert the welfare cost and benefit of introducing the new policy to a dollar-based value, I then calculate the proportion of consumption an individual is willing to pay (receive) to be indifferent between with and without the policy:

$$E_0[U(p)]|_{\widetilde{\pi}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T}\left(u((1-\widetilde{\pi})\cdot c_{it}^g, l_{it}^g, n_{it}, \widetilde{n}_{it}^g)\right),$$

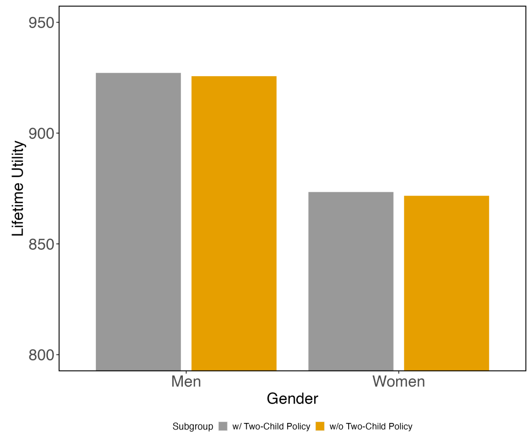
where I solve for $\pi$ such that

$$E_0[U(OneChildPolicy)]|_{\widetilde{\pi}} = E_0[U(TwoChildPolicy)]|_0.$$

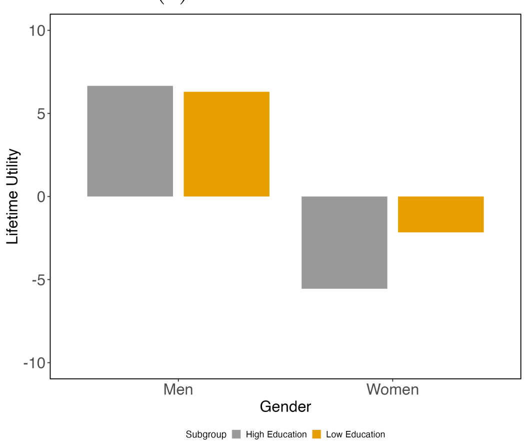
$\widetilde{\pi}$ is the proportion of consumption an individual is willing to pay (receive) to be indifferent between with and without the policy. Table 12 presents the resulting $\pi$ for both men and women across different categories. Only the results for people in strict provinces are provided since they went through a large policy shock and are of the study interest. From the table, we can see that, on average, women are willing to give up 6.00% of their lifetime consumption to avoid the introduction of the Two-Child Policy, which is a nontrivial amount. Consistent with the lifetime utility change we show in the figure above, women with a high educational level or with an ideal number of children equal to one are especially reluctant to have the new policy, willing to give up 8.20% and 10.00% respectively of their lifetime consumption to avoid the policy. Even women with a low educational degree or wanting two children still would like to pay around 4%–5% of their lifetime consumption to be indifferent between the One-Child Policy and the Two-Child Policy. Men, on average, need to receive 7.23% of the lifetime consumption to be indifferent between the two policies. Men with high educational levels are especially welcoming the new policy (8.41%). Another observation is that, among all groups, the gender inequality between men and women with an ideal number of children equal to one was mostly enlarged after the policy since the welfare gap was enlarged most



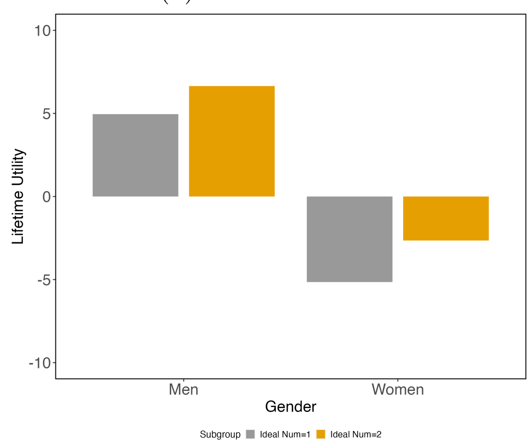

(a) Strict Province



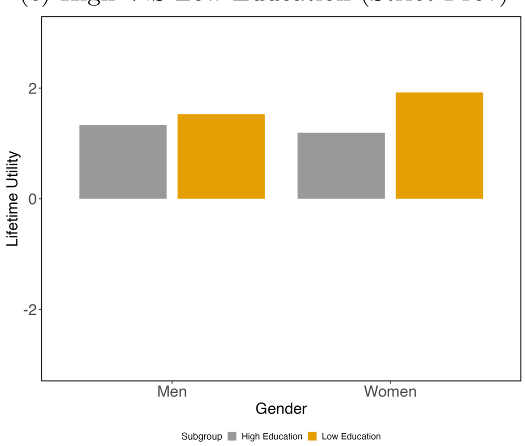

(b) Loose Province



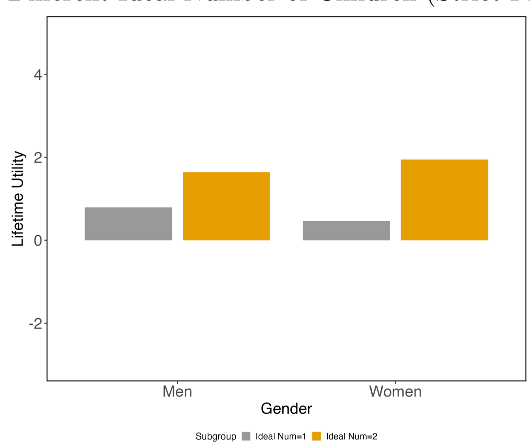

(c) High V.S Low Education (Strict Prov)




(d) Different Ideal Number of Children (Strict Prov)




(e) High V.S Low Education (Loose Prov)




(f) Different Ideal Number of Children (Loose Prov)

Figure 9: Lifetime Utility Change for Couples

significantly by the policy for them.

Table 12: Proportion of Consumption to Generate Equivalence Between Policies

| Gender | Women | Men |
|---|---|---|
| Strict provinces | 6.00% | -7.23% |
| Strict provinces, high educ | 8.20% | -8.41% |
| Strict provinces, low educ | 3.94% | -6.10% |
| Strict provinces, idealNum=1 | 10.00% | -7.50% |
| Strict provinces, idealNum=2 | 4.80% | -6.71% |

Note: This table shows the the proportion of lifetime consumption that the wife and the husband is willing to give up (or receive) to avoid the introduction of the Two-Child Policy. This results contain couples in strict provinces only. Different rows present the results for couples with different characteristics.

### 6.3 Decomposition: Shutting Down the Negative Wage Shock

After gaining a big picture of the welfare impact of the Two-Child Policy on genders, we can now decompose the welfare loss for women into different channels. Since the changes in the Pareto weight due to the changes in the gender wage gap have played an important role in deciding the welfare impact of the new policy, in this subsection, we shut down the negative wage shock for women and study the alternative welfare implication of the Two-Child Policy in the lack of the wage channel. To do this, when simulating the wage process for the couples, I remove the coefficient for the policy impact on the wage gap and set it to equal zero. Other parts remain the same, where the penalty for having excess birth is still removed after the policy change. I then re-generate the choice profiles under this counterfactual scenario and calculate the lifetime utility for couples. Figure 10 compares the results between the real case and several different hypothetical cases. The red bar shows the lifetime utility for men and women had there been no Two-Child Policy. The green bar shows the case in reality where there is Two-Child Policy. Finally, the blue bar shows the case where there is Two-Child Policy, but there is no negative wage shock to women compared to men.

Firstly, men benefit from having Two-Child Policy, as shown in the graphs and tables before. They still benefit from having Two-Child Policy even when there is no wage channel since the blue bar is taller than the red bar, but not as much as when there is a wage channel. Since men typically want two children, they would have a welfare gain through the fertility channel only. They would be even better with the wage channel, leading them to a more superior allocation within the household. That is why the blue bar is a little shorter than the green bar. Secondly, women actually benefit from having Two-Child Policy after shutting down the wage channel. Without the worsened intrahousehold allocations resulting from the lower Pareto weight that was brought about by the negative wage shock, they, on average, enjoy the new policy since they have more freedom to reach their desired number of children.

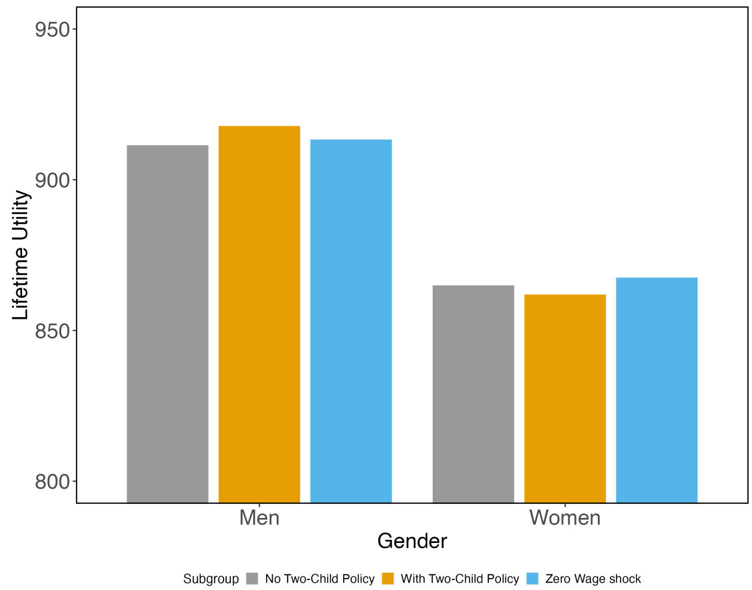



Figure 10: Lifetime Utility Under Two-Child Policy with No Wage Shock

## 6.4 Conducting Policy Experiments

It is widely acknowledged that boosting fertility is difficult in modern countries. This is also the case in China. Although the Two-Child Policy universally removed the fertility bind from all couples, there is still a large proportion of them that did not have a second child even though their ideal number of children is two . Therefore, it is of great importance how various policies could be used to motivate couples to have more children, reach their own ideal number of children and, at the same time, boost fertility in society as a whole. This subsection looks at different related policies, including reducing childrearing costs, providing public childcare services, changing the childcare share burden, and changing the Pareto weight to be more equal between the husband and the wife.

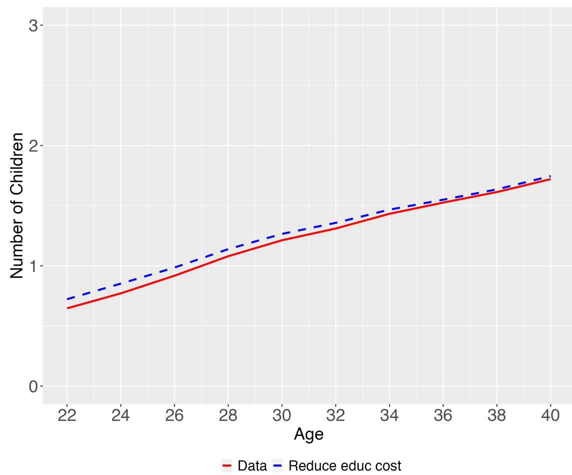
Figure 11 shows the results for the counterfactual average number of children under these different policies. Figure 11 (a) shows the average number of children people would have over their lifetime if they were subsidized 4000 (CNY) per year for the educational costs of their children. Figure 11 (b) shows the number of children couples would have when they were provided 20 hours of childcare service each week for free. Comparing these two graphs, we can find that although both policies lead to an increase in the number of children, the increase in fertility with subsidized educational costs is small compared to that with free childcare services. Therefore, we can draw the conclusion that providing childcare services is more effective in generating more births than reducing educational costs.

I then look at the counterfactual cases of more equally distributed childcare burden and more equal Pareto weight within the household. Firstly, when I change the share of the childcare burden borne by women from the true value of 0.73 to 0.5, couples will be having slightly more children, with the change being quite small, though. This is because although now women are more willing to have more children, it is more costly for the husband to have children. Therefore, the total impact on the number of children not significant. Secondly, when I adjust the Pareto weight to 0.5 from this lower real value from the estimation, the number of children couples going to have decreased dramatically. This is because women tend to have a lower ideal number of children than men. If they are now granted higher bargaining power within the household, then they would make the final birth decisions closer to their preferred locations, which means having fewer kids than in the real case. In summary, the policies in Figure 11 (c) and (d) would improve women's welfare by giving them a more equally distributed childcare burden and a higher Pareto weight. However, these policies actually resulted in a small change or even a negative change in the fertility behavior, given the heterogeneous fertility preferences held by men and women.

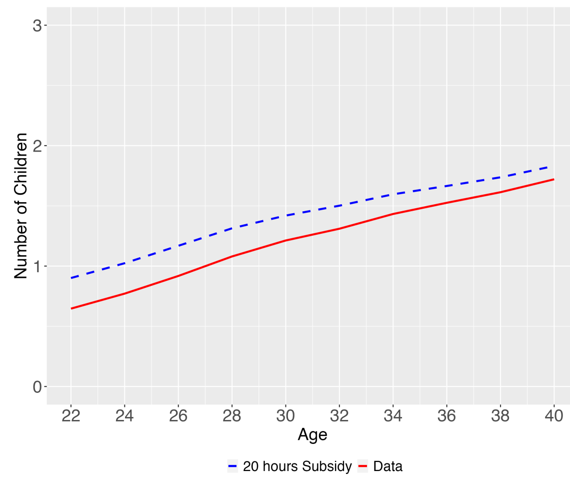
In addition to looking at how different policies affect the number of children people have, we also study how couples' welfare will be affected by these policies. Table 13 below lists the lifetime utility of women and men under the same set of policy experiments as we have discussed before. For this analysis, we only focus on couples in strict provinces since they went through a larger policy shock than those in loose provinces. First of all, we notice that, as we have discussed in previous subsections, women have a higher lifetime utility under without Two-child policy (864.94) than with Two-Child Policy (861.93), while men are better off under the new policy (917.84) than without the policy (911.44). Looking at the results of policy experiments, we find that both reducing educational costs and providing public childcare services boost women's and men's lifetime utility. This is especially true for providing public childcare services, where women's welfare becomes higher than that without the Two-Child Policy. This result suggests the possibility of leveraging fertility-related policies to mitigate the problem that women suffer a loss from having the new policy.

# 7 Conclusion

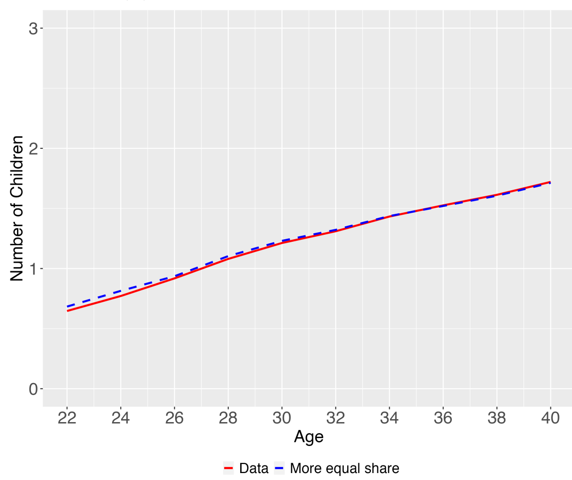
This paper evaluates the welfare impact of the Two-Child Policy on both genders through the interplay between the impact on public and private spheres. I show that the policy negatively affected women's hourly wages, resulting in an enlarged gender wage gap in the labor market.



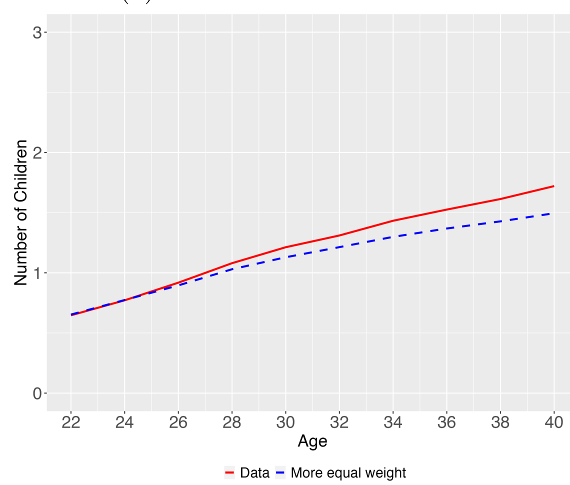

(a) Reduce Educational Costs




(b) Provide Childcare Services




(c) Equal Childcare Share Burden




(d) Equal Pareto Weight

Figure 11: Policy Experiments for Boosting Number of Children

This worsened outside option for women is then transmitted to a lower bargaining power within the household, and I show that, indeed, women's intrahousehold allocation was also negatively affected by the policy change. Motivated by the different channels through which the new policy could affect people's welfare, I built a life-cycle collective household model to structurally estimate couples' preferences and Pareto weights. The model captures the fact that the fertility restriction was removed after the policy change while the gender wage gap was enlarged after the policy as well. I adopt policy gradient methods from reinforcement learning to solve for the policy function and apply indirect inference to estimate the structural parameters.

The results suggest that gender inequality was enlarged after the policy change since the policy had a negative impact on women's welfare and a positive impact on men's welfare.

Table 13: Couple's Lifetime Utility Under Various Policy Experiments

| Lifetime Utility | Two-Child Policy | Policy Off |
|---|---|---|
| Women | 861.93 | 864.94 |
| Men | 917.84 | 911.44 |
| | **Policy Experiments** | |
| | Reduced Educational Costs | Public Childcare Service |
| Women | 863.98 | 873.63 |
| Men | 920.00 | 929.89 |

Note: This table shows couples' lifetime utility under the TCP and without the TCP. It also provides the simulated lifetime utility under various policies experiments. The samples contain couples in strict provinces only.

The welfare cost of the Two-Child Policy for women is equivalent to 6.00% of lifetime consumption, while the welfare benefit of the policy for men is equivalent to 7.23% of lifetime consumption. The estimated Pareto weight shows that women's intra-household bargaining power has been significantly negatively affected by the policy change, from around 0.4 to below 0.3. In addition, this impact is heterogeneous across people, where women with high educational levels and those who have a low fertility preference suffered the most from the policy change. Policy experiments suggest that shutting down the negative wage shock for women significantly improves their welfare while providing public childcare subsidies is most effective in stimulating fertility in the post-policy era.

As a crucial shift in the fertility policy happened in recent years, the Two-Child Policy will have profound and far-reaching effects on various aspects of Chinese society. Further research is essential to fully understand the wide-ranging effects of the Two-Child Policy on different socioeconomic outcomes. For instance, in this paper, I did not consider how the Two-Child Policy would affect marriage formation as it changes the value of marriage differently for men and women. I also take it as given how the policy affected the labor market outcomes for women without fully modeling the employer's side for change of fertility expectation. Future research studying the impact of this massive policy change on gender roles, family structures, and social dynamics will provide valuable insights for informed decision-making and policy optimization.

# References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2017. "The Career Costs of Children." *Journal of Political Economy,* 125(2): 293–337.

Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household Bargaining and Excess Fertility: An Experimental Study in Zambia." *American Economic Review,* 104(7): 2210–37.

Baudin, Thomas, David de la Croix, and Paula E. Gobbi. 2015. "Fertility and Childlessness in the United States." *American Economic Review,* 105(6): 1852–82.

Calvi, Rossella. 2020. "Why Are Older Women Missing in India? The Age Profile of Bargaining Power and Poverty." *Journal of Political Economy,* 49.

Charpentier, Arthur, Romuald Élie, and Carl Remlinger. 2023. "Reinforcement Learning in Economics and Finance." *Computational Economics,* 62:425–462.

Chen, Yi, Hong Zhang, and Li-An Zhou. 2020. "Motherhood and Gender Wage Differentials within a Chinese Firm." *Economic Development and Cultural Change,* 711652.

Chen, Yi, and Hanming Fang. 2021. "The Long-Term Consequences of China's 'Later, Longer, Fewer' Campaign in Old Age." *Journal of Development Economics,* 151: 102664.

Chen, Yi, and Yingfei Huang. 2020. "The Power of the Government: China's Family Planning Leading Group and the Fertility Decline since 1970s." *Demographic Research,* 42 (35): 986-1038.

Doepke, Matthias, and Fabian Kindermann. 2019. "Bargaining over Babies: Theory, Evidence, and Policy Implications." *American Economic Review,* 109(9): 3264–3306.

Ebenstein, Avraham. 2010. "The "Missing Girls" of China and the Unintended Consequences of the One Child Policy." *Journal of Human Resources,* 45 (1): 87-115.

Francesconi, Marco. 2002. "A Joint Dynamic Model of Fertility and Work of Married Women." *Journal of Labor Economics,* vol. 20, no. 2, pt. 1.

Gayle, George-Levi, Andres Hincapie, and Robert A. Miller. 2021. "Life-Cycle Fertility and Human Capital Accumulation." *Working Paper.*

Gourieroux, Christian, Alain Monfort, and Eric Renault. 1993. "Indirect Inference." *Journal of Applied Econometrics,* 8 (Supplement): S85-S118.

He, Haoran, Sherry Xin Li, and Yuling Han. 2022. "Labor Market Discrimination against Family Responsibilities: A Correspondence Study with Policy Change in China." *Journal of Labor Economics,* forthcoming.

Heath, Rachel, and Xu Tan. 2020. "Intrahousehold Bargaining, Female Autonomy, and Labor Supply: Theory and Evidence from India." *Journal of the European Economic Association,* 18(4): 1928–1968.

Hua, William, Hongyuan Mei, Sarah Zohar, Magali Giral, and Yanxun Xu. 2022. "Personalized Dynamic Treatment Regimes in Continuous Time: A Bayesian Approach for Optimizing Clinical Decisions with Timing." *Bayesian Analysis,* 17(3): 849-878.

Huang, Yuting, Juan Pantano, Han Ye, and Junjian Yi. 2023. "Property Division upon Divorce and Household Decisions." *Journal of Human Resources,* 58 (2) 532-560.

Hughes, Neal. 2014. "Applying reinforcement learning to economic problems." Technical report, Australian National University.

Jin, Wei, Yang Ni, Jane O'Halloran, Amanda B. Spence, Leah H. Rubin, and Yanxun Xu. 2023. "A Bayesian Decision Framework for Optimizing Sequential Combination Antiretroviral Therapy in People with HIV." *Annals of Applied Statistics,* forthcoming.

Kaji, Tetsuya, Elena Manresa, Guillaume Pouliot. 2022. "An adversarial Approach to Structural Estimation." *Working Paper.*

Kakade, Sham. 2001. "A natural policy gradient." *Advances in neural information processing systems,* 14.

Keane, Michael P., Kenneth I. Wolpin. 2010. "The Role of Labor and Marriage Markets, Preference Heterogeneity, and the Welfare System in the Life Cycle Decisions of Balck, Hispanic, and White Women." *International Economic Review,* 51(3): 851-892.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics,* 11(4): 181–209.

Lise, Jeremy, and Ken Yamada. 2019. "Household Sharing and Commitment: Evidence from Panel Data on Individual Expenditures and Time Use." *Review of Economic Studies,* 86, 2184–2219.

Li, Xiongfei. 2022. "The Unintended Impacts of the One-child Policy Relaxation in China on Female Labor Market Outcomes." University of Southern California Working Paper.

Low, Hamish, Costa Meghir, Luigi Pistaferri, and Alessandra Voena. 2020. “Marriage, Labor Supply and the Dynamics of the Social Safety Net.” *Working Paper.*

Myong, Sunha, JungJae Park, and Junjian Yi. 2021. “Social Norms and Fertility.” *Journal of the European Economic Association,* 19(5): 2429-2466.

Qian, Yue, and Yonggai Jin. 2018. “Women’s Fertility Autonomy in Urban China: The Role of Couple Dynamics Under the Universal Two-Child Policy.” *Chinese Sociological Review,* 50(3): 275-309.

Rasul, Imran. 2008. “Household Bargaining over Fertility: Theory and Evidence from Malaysia.” *Journal of Development Economics,* 86: 215-241.

Peters, Jan and Stefan Schaal. 2006. “Policy gradient methods for robotics.” In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems,* pages 2219–2225. IEEE.

Peters, Jan and Stefan Schaal. 2008. “Reinforcement learning of motor skills with policy gradients.” *Neural networks,* 21(4):682–697.

Silver, D., Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. 2014. “Deterministic policy gradient algorithms.” In *International conference on machine learning,* pages 387–395. Pmlr.

Sutton, Richard S and Andrew G Barto. 2018. “Reinforcement Learning: An Introduction.” *MIT Press.*

Sutton, Richard S, David McAllester, Satinder Singh, and Yishay Mansour. 1999. “Policy gradient methods for reinforcement learning with function approximation.” *Advances in neural information processing systems,* 12.

Voena, Alessandra. 2015. “Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples?” *American Economic Review,* 105(8): 2295–2332.

Wang, Fei, Liqiu Zhao, and Zhong Zhao. 2017. “China’s Family Planning Policies and Their Labor Market Consequences.” *Journal of Population Economics,* 30(1): 31–68.

Yu, Yewen, Yi Fan, and Junjian Yi. 2020. “The One-Child Policy Amplifies Economic Inequality Across Generations in China.” IZA Discussion Paper No. 13617. Available at https://ssrn.com/abstract=3679023.

Zhang, Junsen. 2017. “The Evolution of China’s One-Child Policy and Its Effects on Family Outcomes.” *Journal of Economic Perspectives,* 31(1): 141-160.

# A Details for the Policy Impact on Various Outcomes

## A.1 Effect of the Policy on the Birth Rate

I start with estimating the effect of the Two-Child Policy on the fertility level. Since the CFPS data set contains a panel structure for couples, it will be difficult to disentangle the age effect from the policy effect on people's birth decisions. Hence, for this analysis, I use the data from Chinese Statistical Yearbook, where I construct a year-province level of birth rates. The birth rates in this data set are calculated using the total new births divided by the total population in the sampled set from each province annually. Using this data set, I run the following Difference-in-Differences regression:

$$\begin{aligned} Birth_{kt} = \alpha_0 + \alpha_1 Post_{2016} + \alpha_2 PolicyShock_k + \alpha_3 Post_{2016} \times PolicyShock_k \\ + f_t + f_k + \epsilon_{kt}, \end{aligned} \tag{A.1}$$

where $Birth_{kt}$ is province $k$'s annual birth rate in year $t$ and $PolicyShock_k$ is the $0-1$ indicator of treatment and control status that depends on the excess fertility rate residual $\varepsilon(k)$ we obtained from last step. The estimated coefficient $\widehat{\alpha}_3$ captures the policy effect on the average birth rate in each province and thus on the fertility level of people. The identifying assumption is that in the lack of the Two-Child Policy, the birth rates should trend in the same way in the treatment and control groups. To make sure that there is no pre-trend before the policy change, I run an event study analysis for the birth rates. The graph is shown in Figure 3 (a). The event study graph confirms that the parallel trend assumption holds for the provincial birth rates. Table A.1 presents the results for the regression in equation A.1. When running the difference-in-differences regression, I control for year fixed effects and province fixed effects for both specifications, and add province-specific time trends in the second specification. Both columns (1) and (2) in the table show a significant, positive coefficient for the interacted term, suggesting that the policy had a significantly positive impact on the provincial birth rates. Compared to men, the policy led to a lower wage for women for those in the treatment provinces compared to their counterparts in the control provinces. The result confirms that there was a significant and positive fertility response to the policy relaxation happened in 2016.

## A.2 Effect of the Policy on the Gender Wage Gap

In this section, I quantify the effect of the Two-Child Policy on the gender wage gap in the labor market. With the definition of treatment and control groups, I run the following

Table A.1: Difference-in-Differences Results for Birth Rates

| VARIABLES | (1) Birth Rate | (2) Birth Rate |
|---|---|---|
| $\text{Post}_{2016} \times$ Policy Shock | 0.6985*** | 0.6309* |
| | (0.2022) | (0.3439) |
| Sample mean outcome | 11.34 | 11.34 |
| Observations | 250 | 250 |
| R-squared | 0.9258 | 0.9540 |
| Year FE | Y | Y |
| Province FE | Y | Y |
| Province-specific Linear Trend | N | Y |

Note: This table presents the results for the difference-in-differences estimation for the impact of the Two-Child Policy on the birth rate. The regression is on the province level, where there are 25 provinces and 10 years, resulting in 250 observations in total. The dependent variable is the average birth rate (‰) for each province. The coefficient in the first row is for the variable that intersects being in post-policy time and being in the treatment provinces (provinces with large policy shocks), which shows the policy effect on fertility. I control for province and year fixed effects in both columns and add province-specific linear trend in the second column.

triple-difference regression analysis for the hourly wage in the labor market:

$$\begin{aligned}Wage_{ikt} = \beta_0 &+ \beta_1 Post_{2016} + \beta_2 Female_i + \beta_3 PolicyShock_k + \beta_4 Post_{2016} \times Female_i \\ &+ \beta_5 Post_{2016} \times PolicyShock_k + \beta_6 Female_i \times PolicyShock_k \\ &+ \beta_7 Post_{2016} \times Female_i \times PolicyShock_k + \boldsymbol{X}_{ikt}\boldsymbol{\alpha} + f_t + f_k + f_i + \epsilon_{ikt},\end{aligned} \tag{A.2}$$

where $Wage_{ikt}$ is individual $i$'s hourly wage rate in province $k$ at time t and $PolicyShock_k$ is the $0-1$ indicator of treatment and control status that depends on the excess fertility rate residual $\varepsilon(k)$ we obtained from last step. The coefficient $\beta_7$ in front of the triple-interacted term is the policy effect on the gender wage difference (female's hourly wage − male's hourly wage). The estimated $\hat{\beta}_7$ is the triple-difference estimator for the policy effect on the gender wage gap. The identifying assumption is that, in the lack of the Two-Child Policy, the difference between the female wage and the male wage should trend in the same way in the treatment and control provinces. I test this assumption by drawing the trend graph for the gender wage gap over time and running an event study analysis for the gender wage gap to validate my results. The trend graph can be found in Figure D.1 in Appendix D and the event study graph can be found in Figure 3 (b). The event study dynamics confirm that there was no pre-trend for the gender wage difference before the policy reform in 2016.

Table A.2 presents the results for the regression in equation A.2. When running the triple-difference regression, I control for various individual characteristics, including age, education, number of children, and so on. I also include year fixed effects, province fixed effects, and province-specific time trends when running the baseline regression. Column (1) in the table shows a significant, negative coefficient for the triple-interacted term, suggesting that the policy had a significantly negative impact on the gender wage difference. Compared to men, the policy led to a lower wage for women for those in the treatment provinces compared to their counterparts in the control provinces. For columns (2)-(3) in the table, I gradually add industry dummies and individual fixed effects, and the results did not change meaningfully.

The results suggest that the policy effect on the gender wage gap is rather exogenous, possibly coming from the employer's side due to discrimination motives. With the new policy, employers now expect couples to have more children. Since China's traditional social norms require women to bear a higher share of the childcare burden than men, they will have to spend more time within the family. Therefore, when it is easy to find an equally qualified male worker, employers will choose to hire him instead. Even if the women do not plan to have a second birth, it is difficult for them to faithfully convince employers about that. In summary, the expectation of a potentially higher fertility level for female employees, no matter whether the expectation is fulfilled at last, will lead to employers' discrimination against female workers. In my study, the discrimination is reflected in a lower wage for women compared to men, all else equal.

To further validate the results for the impact of the policy on the gender wage gap, I conduct two sets of robustness checks. The first analysis is to run the same triple difference regression for couples in which the wife's age is greater than 40. The argument is that, for women out of the fertility age range, the Two-Child Policy should have no effect on employers' expectations of their fertility, and therefore should not affect the gender wage gap. Table D.6 in Appendix D presents the results for the triple difference regressions. The coefficients are insignificant across all the specifications, confirming the argument that the policy has no effect on couples out of the fertility age range. This robustness check defends the main result that the Two-Child Policy enlarged the gender wage gap due to employers' expectations of potentially higher fertility levels of women.

The second set of robustness checks is to take into consideration that women might select into work. I use the classic Heckman's two-step methods to correct for the bias for women's selection into work and re-run the same regression as in the main analysis. Table D.7 in Appendix D shows the results. After controlling for the inverse Mills ratio, the coefficients remain significant in the first and second specification. This suggests that considering women's selection into work, the policy still has a significant negative impact on

Table A.2: Triple Difference Results for Hourly Wage Rate of Women

| VARIABLES | (1) Hourly Wage | (2) Hourly Wage | (3) Hourly Wage |
|---|---|---|---|
| $\text{Post}_{2016} \times$ Female $\times$ Policy Shock | -2.357*** | -2.307** | -1.715* |
| | (0.876) | (0.944) | (0.923) |
| Observations | 16,082 | 13,373 | 12,729 |
| R-squared | 0.198 | 0.233 | 0.233 |
| Individual Characteristics | Y | Y | Y |
| Year FE | Y | Y | Y |
| Province FE | Y | Y | Y |
| Province-specific Time Trend | Y | Y | Y |
| Industry Dummies | N | Y | Y |
| Individual FE | N | N | Y |

Note: This table presents the results for the triple-difference estimation results for the impact of the Two-Child Policy on the wage of women. The dependent variable is the hourly wage rate of women (CNY). The coefficient in the first row is for the variable that intersects being in post-policy time, being female, and being in the treatment provinces (provinces with large policy shock), which shows the policy effect on the gender wage gap. Across the columns, I gradually add more control variables, including industry and occupation dummies and individual fixed effects. In the last column, I only look at couples whose number of children is constant over time.

the gender wage gap, thus confirming our main results.

## A.3 Effect of the Policy on the Intrahousehold Bargaining Power

After obtaining the results for the effect of the policy on the gender wage gap in the labor market, I now turn to the private sphere and study how the policy change affected the intrahousehold bargaining power of women. In the literature on family economics, there are many ways to measure the bargaining power of women within the household. One is to use a direct measure of bargaining power[17]. Another direction is to use assignable private consumption to back out the resource-sharing rule. For instance, Calvi (2020) uses clothing for men and women separately to identify resource-sharing rules [18]. Following Huang et al. (2023), I use the alcohol consumption as a proxy for individual's bargaining power within

[17] For example, questions on which one of the spouses makes decisions on important family objects can shed light on the bargaining power of the wife. Heath and Tan (2020) use these questions to measure the effect of the Hindu Succession Act on Indian women's autonomy in the household. In CFPS, only the wave in 2014 has questions on which one of the spouses makes decisions on several objects. No after-policy measure is available in the data. Therefore, I cannot reply on the direct measure of bargaining power as in this stream of literature.

[18] Ideally, one would want to observe the private consumption of the husband and wife since the resource-sharing rule is directly related to the Pareto weight. However, private consumption is not available in the CFPS data. I do observe alcohol consumption in the data, which is assignable to the individual.

the household. In the survey, both the husband and the wife were asked about the frequency of drinking alcohol each week. In other words, I can observe the private consumption of alcohol for both spouses. Since alcohol is traditionally a male-favored good, if the husband's bargaining power increased relatively to the wife's, we should expect an increase in his consumption of alcohol. Following the same research design as in studying the policy impact on the gender wage gap, I use the following Difference-in-Differences regression analysis for cosmetic expenditure share:

$$\begin{aligned} Alco_{ikt} = \gamma_0 + \gamma_1 Post_{2016} + \gamma_2 PolicyShock_k + \gamma_3 Post_{2016} \times PolicyShock_k \\ + \boldsymbol{X}_{ikt}\boldsymbol{\alpha} + f_t + f_k + f_j + \epsilon_{ikt}, \end{aligned} \tag{A.3}$$

where $Alco_{ikt}$ is the binary indicator of having consumed alcohol for at least three times last week for individual $i$ in province $k$ at time $t$. The estimated coefficient $\widehat{\gamma}_3$ captures the policy effect on the private alcohol consumption and thus on the intrahousehold bargaining power of men and women. I run the regression for both men and women. Since alcohol is mainly consumed by men, I would expect an increase in men's alcohol consumption followed by a higher bargaining power for them than before, while no significant effects on women's alcohol consumption since their demand is inelastic. Table A.3 presents the results of the regression analysis (equation A.3). When running the Difference-in-Differences regression, I control for individual and household characteristics, including age, educational levels, and the number of children in the household. I also include province fixed effects and year fixed effects in my basic specification. For the second specification, I further add individual fixed effects. Columns (1)–(2) show the results for husbands, which confirm that the policy had a significantly positive impact on their alcohol consumption ($0.0435^{***}$), suggesting that the policy positively affected men's bargaining power within the household [19]. Columns (3)–(4) show the results for wives. As expected, there is no significant impact on wives' alcohol consumption, since women traditionally do not drink alcohol in Chinese society, therefore leading to a low demand elasticity for female alcohol consumption. The identifying assumption is that in the lack of the Two-Child Policy, the alcohol consumption should trend in the same way in the treatment and control groups for both husbands and wives. To make sure that there is no pre-trend before the policy change, I run an event study analysis for the alcohol consumption frequency for husbands. The graph is shown in Figure 3 (c). It can be confirmed that there was no pre-trend existing for the alcohol consumption for husbands, proving the validity of the difference-in-differences results.

The result that the Two-Child Policy increased men's bargaining power while decreased women's bargaining power stems from the fact that women's labor market outcomes worsened

[19] Column (2) presents the results for the regression analysis adding individual fixed effects.

after the policy change. As shown in the first part of the empirical results, due to labor market discrimination, women's wages became lower compared to men's, resulting in an enlarged gender wage difference. If we consider the outside option of a cooperative marriage as having a divorce, a lower wage and a worse labor market environment compared to the husband would lead to decreased outside option values for women than before. As a consequence, this change in outside options is transmitted into the household, leading to a higher bargaining power for husbands and a lower bargaining power for wives.

Table A.3: Difference-in-Differences Results for Alcohol Consumption

| VARIABLES | (1) Husband Alcohol Consumption | (2) | (3) Wife Alcohol Consumption | (3) |
|---|---|---|---|---|
| $\text{Post}_{2016} \times$ Policy Shock | 0.0435*** | 0.0397** | 0.0056 | 0.0028 |
| | (0.0167) | (0.0163) | (0.0049) | (0.0052) |
| Observations | 11,144 | 11,144 | 13,506 | 13,506 |
| R-squared | 0.0518 | 0.0510 | 0.0125 | 0.0122 |
| Year FE | Y | Y | Y | Y |
| Province FE | Y | Y | Y | Y |
| Individual Characteristics | Y | Y | Y | Y |
| Individual Fixed Effects | N | Y | N | Y |

Note: This table presents the results for the difference-in-differences estimation of the impact of the Two-Child Policy on the alcohol consumption frequency. The dependent variable is a binary indicator of whether consuming alcohol for at least three times last week. The first two columns contain results for husbands, and the last two columns contain results for the wives. The coefficients in the first row represent the policy effect. Across the columns, I add individual fixed effects.

# B Leisure Hours Imputation

There is a data issue with CFPS that, in the six waves, only the year 2010 has a detailed time-use module, which contains hours spent on leisure for survey respondents. All the other waves only have weekly watching-TV time available. Since hours on watching TV should be highly correlated with the total leisure hours, I could leverage this information to impute the missing leisure hours for later waves. The detailed steps are as follows. First, I gather leisure hours, watching TV hours, and other characteristics of individuals in the year 2010 to train a Random-forest model to predict leisure hours in later years:

$$Y_i = f(T_i, \overrightarrow{X}_i | \theta, \varepsilon),$$

where $Y_i$ stands for weekly leisure hours, $T_i$ is weekly hours on watching TV, and $\overrightarrow{X}_i$ are individual characteristics including age, education, gender, number of children, employment status, province dummy, and whether having young children at home. After obtaining the estimated parameters $\hat{\theta}$, I then plug those values into the model to predict leisure hours in later waves.

# C Moment Conditions for Wage Process

The moment conditions for men are

$$\begin{aligned} \frac{E[u_{it}u_{it+2}]}{E[u_{it}u_{it+1}]} &= \rho \\ E[u_{it}^2] - \frac{(E[u_{it}u_{it+1}])^2}{E[u_{it}u_{it+2}]} &= \sigma_\epsilon^2 \\ \frac{(E[u_{it}u_{it+1}])^2}{E[u_{it}u_{it+2}]} - E[u_{it}u_{it+2}] &= \sigma_\xi^2. \end{aligned}$$

The moment conditions for women are

$$\begin{aligned} \frac{E[u_{it}u_{it+2}|L_{it}, L_{it+2}=1]}{E[u_{it}u_{it+1}|L_{it}, L_{it+1}=1]} &= \rho \\ E[\Delta u_{it}|L_{it}, L_{it-1}=1] &= \sigma_{\xi\eta}\frac{\phi(\alpha_{it})}{1-\Phi(\alpha_{it})} \\ &\quad + (\rho-1)\sigma_{\xi\eta}\frac{\phi(\alpha_{i,t-1})}{1-\Phi(\alpha_{i,t-1})} \\ E[u_{it}^2|L_{it}=1] - \frac{(E[u_{it}u_{it+1}|L_{it}, L_{it+1}=1])^2}{E[u_{it}u_{it+2}|L_{it}, L_{it+2}=1]} &= \sigma_\epsilon^2 \\ E[(\Delta u_{it+1})^2|L_{it}, L_{it+1}=1] - \frac{(\rho-1)^2}{\rho}E[u_{it}u_{it+1}|L_{it}, L_{it+1}=1] &= \sigma_\xi^2 + 2\sigma_\epsilon + \sigma_{\xi\eta}^2\alpha_{it}\frac{\phi(\alpha_{it})}{1-\Phi(\alpha_{it})}. \end{aligned}$$

# D Additional Tables and Graphs

Table D.4: Regression Result for the Fertility Model

| | (1) |
|---|---|
| VARIABLES | Excess Fertility Rate |
| High School above | 0.0396 |
| | (0.0579) |
| Agriculture | 0.0540 |
| | (0.0321) |
| Ideal Num. of Child | 0.0988*** |
| | (0.0206) |
| Age | -0.00592 |
| | (0.00835) |
| Log(House price) | 0.00159 |
| | (0.0198) |
| Log(GDP) | -0.0235 |
| | (0.0152) |
| Constant | 0.252 |
| | (0.361) |
| Observations | 25 |
| R-squared | 0.707 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D.5: Correlation Between Different Policy Strictness Measures

| | EFRR | Policy Fertility | Fine (Avg) | Fine (Min) |
|---|---|---|---|---|
| EFRR | | | | |
| Policy Fertility | 0.05 | | | |
| Fine (Avg) | -0.28 | -0.03 | | |
| Fine (Min) | -0.18 | -0.05 | 0.83*** | |
| Fine (Max) | -0.29 | -0.02 | 0.96*** | 0.65*** |

Table D.6: Triple Difference Results for Hourly Wage (Couples Out of Fertility Age)

| VARIABLES | (1) Hourly Wage | (2) Hourly Wage | (3) Hourly Wage |
|---|---|---|---|
| $Post_{2016} \times$ Female $\times$ Policy Shock | -1.476 | -3.215 | -2.127 |
| | (2.219) | (2.324) | (1.985) |
| Observations | 3,453 | 2,620 | 2,503 |
| R-squared | 0.193 | 0.258 | 0.252 |
| Individual Characteristics | Y | Y | Y |
| Year FE | Y | Y | Y |
| Province FE | Y | Y | Y |
| Province-specific Time Trend | Y | Y | Y |
| Industry Dummies | N | Y | Y |
| Individual FE | N | N | Y |

Note: This table presents the results for the triple-difference estimation results for the impact of the Two-Child Policy on the wage of women for couples out of fertility age. The estimation sample only includes couples where the wife's age is greater than 40. The dependent variable is the hourly wage rate of women (CNY). The coefficient in the first row is for the variable that intersects being in post-policy time, being female, and being in the treatment provinces (provinces with large policy shock), which shows the policy effect on the gender wage gap. Across the columns, I gradually add more control variables, including industry dummies and individual fixed effects.

Table D.7: Triple Difference Results for Hourly Wage (Consider Selection Into Work)

| VARIABLES | (1) Hourly Wage | (2) Hourly Wage | (3) Hourly Wage |
|---|---|---|---|
| $Post_{2016} \times$ Female $\times$ Policy Shock | -2.208** | -2.059** | -1.338 |
| | (0.907) | (0.974) | (0.925) |
| Observations | 15,106 | 12,565 | 12,413 |
| R-squared | 0.196 | 0.231 | 0.230 |
| Individual Characteristics | Y | Y | Y |
| Year FE | Y | Y | Y |
| Province FE | Y | Y | Y |
| Industry Dummies | N | Y | Y |
| Individual FE | N | N | Y |

Note: This table presents the results for the triple-difference estimation results for the impact of the Two-Child Policy on the wage of women, taking into consideration women's selection into work. The classic Heckman's two-step method is applied to correct for the selection bias. The dependent variable is the hourly wage rate of women (CNY). The coefficient in the first row is for the variable that intersects being in post-policy time, being female, and being in the treatment provinces (provinces with large policy shock), which shows the policy effect on the gender wage gap. Across the columns, I gradually add more control variables, including industry dummies and individual fixed effects.

Table D.8: Model Fit: Empirical and Simulated Moments: Completed Fertility and Excess Fertility Rate (In-Sample)

| **Moment Description** | **Data**<br>[95% CI] | **Model** |
|---|---|---|
| **Completed fertility** | | |
| Before policy change, strict province | 1.412<br>[1.273, 1.553] | 1.211 |
| After policy change, strict province | 1.500<br>[1.419, 1.581] | 1.545 |
| Before policy change, loose province | 1.656<br>[1.505, 1.807] | 1.584 |
| After policy change, loose province | 1.624<br>[1.550, 1.697] | 1.603 |
| **Excess Fertility Rate** | | |
| Before policy change, strict province | 0.376<br>[0.251, 0.501] | 0.367 |
| After policy change, strict province | 0.428<br>[0.365, 0.491] | 0.543 |
| Before policy change, loose province | 0.576<br>[0.457, 0.695] | 0.552 |
| After policy change, loose province | 0.577<br>[0.514, 0.639] | 0.575 |

Table D.9: Model Fit: Empirical and Simulated Moments for Difference-in-Differences Regression of Fertility (In-Sample)

| **Moment Description** | **Data** [95% CI] | **Model** |
|---|---|---|
| **Diff-in-diff fertility regression** | | |
| Coefficient for $\tilde{n}^f_{j,t}$ | 0.470 [0.383, 0.557] | 0.639 |
| Coefficient for $\tilde{n}^m_{j,t}$ | 0.373 [0.281, 0.466] | 0.807 |
| Coefficient for $age_{j,t}$ | 0.015 [-0.024, 0.055] | -0.022 |
| Coefficient for $posP_{j,t}$ | 0.141 [0.010, 0.272] | 0.198 |
| Coefficient for $strict_{j,t}$ | -0.165 [-0.333, 0.004] | -0.216 |
| Coefficient for $posP_{j,t} \times strict_{j,t}$ | 0.060 [-0.088, 0.208] | 0.213 |
| Coefficient for $w^f_{j,t}$ | -0.023 [-0.034, -0.012] | -0.012 |
| Coefficient for $w^m_{j,t}$ | -0.002 [-0.008, 0.003] | -0.006 |

Table D.10: Model Fit: Empirical and Simulated Moments: Completed Fertility and Excess Fertility Rate (Out-of-Sample)

| **Moment Description** | **Data** [95% CI] | **Model** |
|---|---|---|
| **Completed fertility** | | |
| Before policy change, strict province | 1.552 [1.412, 1.692] | 1.345 |
| After policy change, strict province | 1.456 [1.375, 1.537] | 1.526 |
| Before policy change, loose province | 1.324 [1.172, 1.475] | 1.176 |
| After policy change, loose province | 1.463 [1.390, 1.537] | 1.537 |
| **Excess Fertility Rate** | | |
| Before policy change, strict province | 0.483 [0.358, 0.608] | 0.448 |
| After policy change, strict province | 0.474 [0.411, 0.537] | 0.544 |
| Before policy change, loose province | 0.353 [0.234, 0.472] | 0.412 |
| After policy change, loose province | 0.415 [0.352, 0.477] | 0.520 |

Table D.11: Model Fit: Empirical and Simulated Moments for Difference-in-Differences Regression of Fertility (Out-of-Sample)

| **Moment Description** | **Data** [95% CI] | **Model** |
|---|---|---|
| **Diff-in-diff fertility regression** | | |
| Coefficient for $\tilde{n}^f_{j,t}$ | 0.197 [0.110, 0.284] | 0.644 |
| Coefficient for $\tilde{n}^m_{j,t}$ | 0.571 [0.479, 0.664] | 0.737 |
| Coefficient for $age_{j,t}$ | 0.049 [0.010, 0.088] | -0.013 |
| Coefficient for $posP_{j,t}$ | 0.290 [0.159, 0.421] | 0.447 |
| Coefficient for $strict_{j,t}$ | 0.171 [0.002, 0.340] | 0.144 |
| Coefficient for $posP_{j,t} \times strict_{j,t}$ | -0.166 [-0.314, -0.017] | -0.091 |
| Coefficient for $w^f_{j,t}$ | -0.033 [-0.044, -0.023] | 0.014 |
| Coefficient for $w^m_{j,t}$ | 0.004 [-0.001, 0.010] | -0.011 |

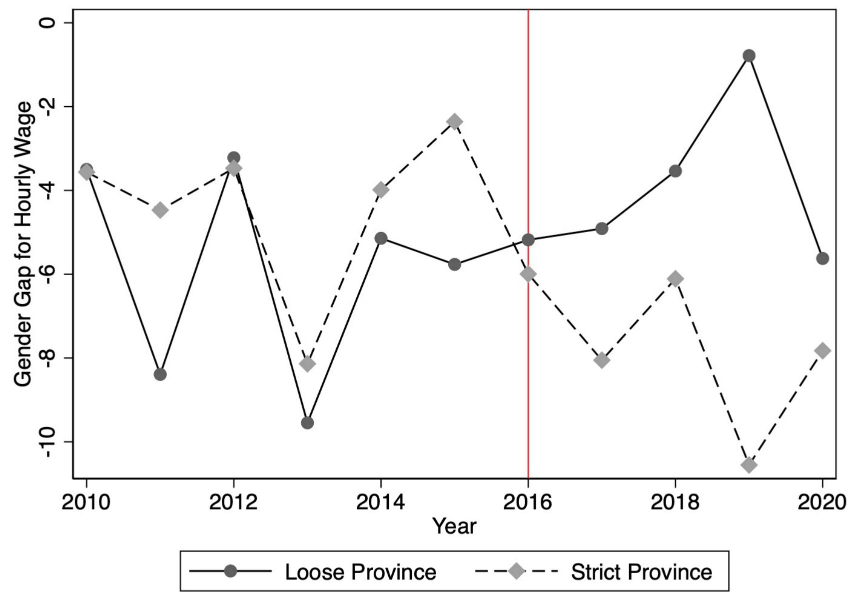
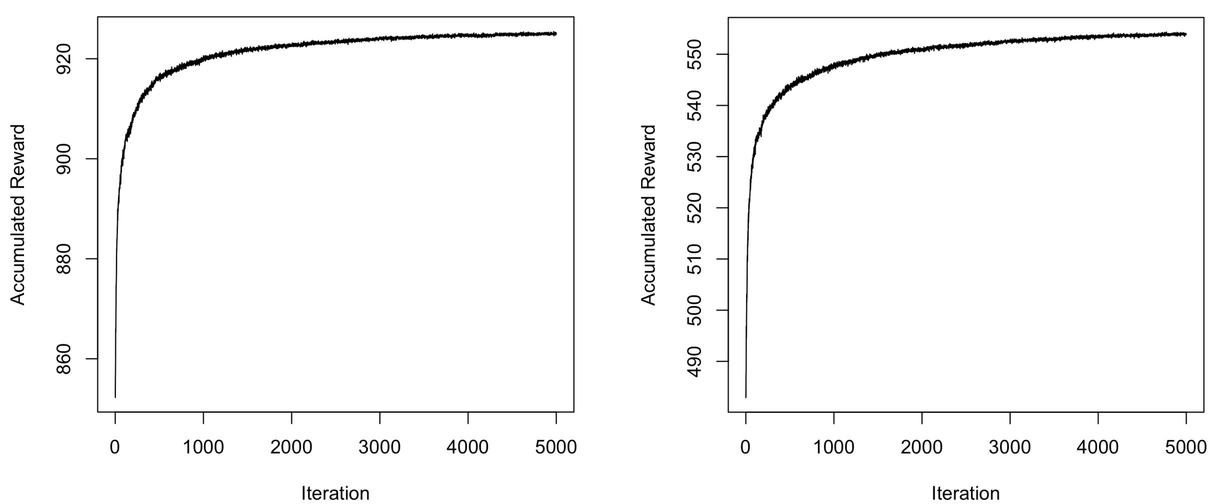


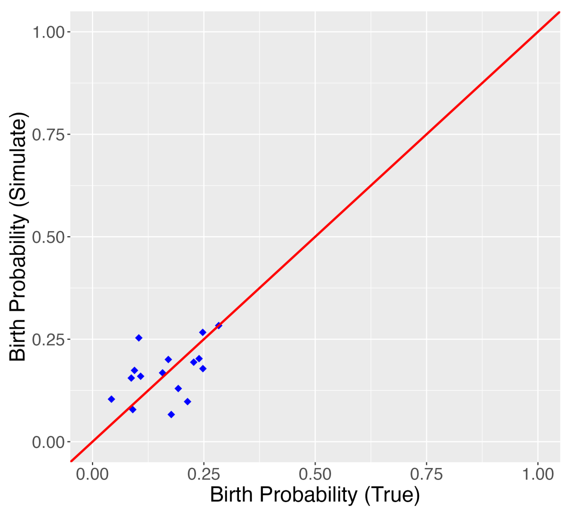

Figure D.1: Trend Graph for the Gender Wage Gap Over Time



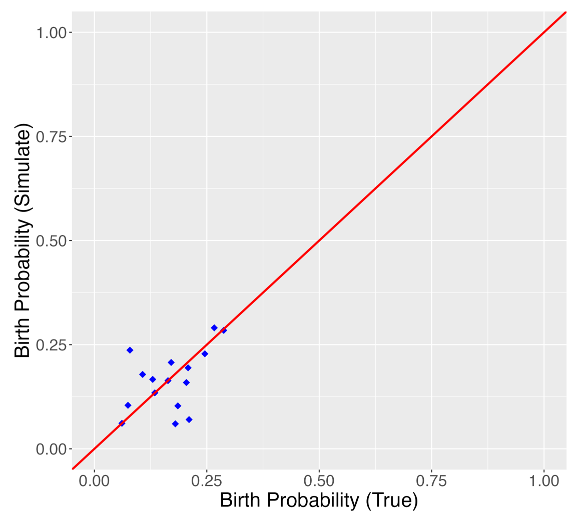

(a) w/o Policy Change

(b) w/ Policy Change

Figure D.2: Lifetime Value over Updating Process of $\lambda$




(a) In-Sample Prediction




(b) Out-of-Sample Prediction

Figure D.3: Model Fit for the Birth Probability Conditional on Different Ideal Number of Children