Exploring Gender Gaps: The Roles of Institutional Policy, Workplace Flexibility, and Intra Household Dynamics

Jingyi Li

A thesis submitted for the degree of

Doctor of Philosophy

Institute for Social and Economic Research

University of Essex

February 2026

AUTHOR'S DECLARATION

I, Jingyi Li, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

I declare that the work in this thesis was carried out in accordance with the requirements of the University's Regulations and that it has not been submitted for any other academic award.

Chapters 1 and 2 are the result of a collaborative effort with Dr. Guohua He, Department of Economics, University of Essex. Chapters 3 of this thesis are authored solely by me. I am responsible for any errors.

Signed by: Jingyi Li

Date: February 2025

Summary

Gender disparities in labor markets are shaped by institutional policies, firm practices, and household dynamics. This thesis examines underexplored aspects of these inequalities across macro, meso, and micro levels.

The first chapter investigates China's two-child policy as a natural experiment using CFPS data and a triple-differences model. Results show that one-child mothers faced higher wage penalties after the policy through statistical discrimination, while two-child mothers experienced lower wage penalties, reflecting a taste-based discrimination.

The second chapter analyzes flexible working arrangements in the UK. Using longitudinal data, it finds that men transition more often into high-skill flexi-time roles with small penalties, while women disproportionately enter low-skill term-time roles with large penalties, reinforcing wage inequality.

The third chapter studies fertility decisions in Australia through intra-household bargaining. Using HILDA data, it shows that both tangible and intangible bargaining power shape fertility timing, though deviations occur among career-oriented and norm-conforming women.

Introduction

Gender disparities in the labor market have been a longstanding issue, deeply rooted in historical, cultural, and economic structures that persist in influencing women's experiences in the workforce around the world. Understanding these persistent inequalities has long been a central focus of family and labor economics. These disparities are the result of complex dynamics spanning macro-level institutional policies, meso-level firm-specific work arrangements, and micro-level intra-household bargaining dynamics. While much of the existing literature has focused on isolated factors contributing to gender disparities, this thesis contributes to the evolving body of knowledge by exploring a selection of less well-researched aspects of these disparities. More specifically, this thesis offers critical insights into the mechanisms underlying gender disparities and their consequences on the gender gap, particularly concerning labor market outcomes (e.g., labor force participation, working time patterns, wage outcomes, and career trajectories) and family outcomes (e.g., household labor division, fertility decisions, and childcare responsibilities) across diverse national contexts.

This thesis consists of three chapters, each focusing on one of these dimensions of gender inequality, from macro-level institutional policies to meso-level firm-specific practices and micro-level family dynamics. It investigates some less well-understood factors at each level, contributing to the broader literature on gender disparities. At the macro-level, while much research has examined family policies on motherhood wage penalty, such as parental leave, this thesis focuses on the relatively underexplored area of fertility policy, particularly the relaxation of fertility restrictions such as China's two-child policy. Specifically, the policy shift has reshaped employers' perceptions of female employees' future fertility, with profound consequences for wage outcomes and career advancement. On the firm level, the adoption of flexible working arrangements (FWAs) has been proposed as a strategy to narrow the gender wage gap in the UK. However, gender disparities persist in the types of FWAs men and women access. This gendered pattern of FWA adoption not only reflects structural inequalities within firms but also contributes to the persistence of the gender wage gap. At the micro level, intra-household bargaining power plays a crucial role in shaping fertility decisions and family outcomes in Australia. In particular, variations in bargaining power and fertility preferences between partners can lead to divergent fertility behavior, perpetuating gender disparities in both the home and the labor market.

To ensure robust causal inference, this thesis employs a comprehensive set of advanced econometric techniques, including difference-in-differences (DID), triple-differences (Triple-DID), Heckman selection models, propensity score matching (PSM), factor analysis (FA), event-study DID approaches, discrete-time survival models, shift-share instrument variables (SSIV), and mediation models, along with universal treatment tests. These methodologies are applied across the three chapters to rigorously examine the effects of policy changes, intra-household dynamics, and organizational practices on gender inequalities.

The first chapter, titled "*How Does Fertility Relaxation Policy Affect the Motherhood Wage Penalty*" investigates the impact of fertility-related policies on the motherhood wage penalty, utilizing China's two-child policy as a natural experiment. This chapter draws on data from the China Family Panel Studies (CFPS) and employs a triple difference (Triple-DID) model to identify the causal effects of the policy change. This chapter highlights how policy changes can alter employer perceptions and discrimination dynamics, highlighting heterogeneity in wage penalties among different groups of mothers. For one-child mothers, the policy exacerbates statistical discrimination, leading to a significant increase in the wage penalty post-policy. Conversely, mothers who previously violated the policy by having a second child experience a lower wage penalty post-policy. These outcomes are primarily attributable to changes in job discrimination rather than shifts in human capital.

Building on the insights from the first chapter, the second chapter, titled "*Gender Disparities in Flexible Working Arrangements and Wage Inequality*", extends the analysis to the workplace level, examining how flexible working arrangements (FWAs) affect the gender wage gap. Using longitudinal survey data from the UK and employing causal inference methods, this chapter demonstrates that FWAs, instead of narrowing gender disparities, may in fact exacerbate them. The gender asymmetry and gender structure frameworks provide a lens through which to understand how gendered patterns of FWA adoption shape these outcomes. Specifically, the study reveals significant gender disparities in transitions to different types of FWAs and their subsequent effects on wage inequality. Men are more likely to transition from traditional work patterns to flexi-time working, while women are more likely to switch to term-time work. These two types of FWAs have very different wage consequences: the wage penalty for women transitioning to term-time work is over six times larger than the penalty for men moving to flexi-time work. Furthermore, occupational segregation exacerbates this gap, as men are more likely to move into high-skill flexi-time work roles, while women predominantly transition into low-skill term-time work positions.

The third chapter, titled "*Understanding the Fertility Puzzle: How Bargaining Power Deviates from Traditional Models*", shifts focus from macro-level factors to the household level, examining fertility decisions within households and highlighting the limitations of traditional intra-household bargaining models in predicting fertility behaviors. Specifically, this chapter solves some of the puzzle of traditional intra-household bargaining models failing to accurately predict fertility behavior. To better understand this puzzle, we firstly identify specific proxies of bargaining power that influence fertility decisions, and then we highlight which groups deviate from the model's predictions. Utilizing a discrete-time survival model and data from HILDA (i.e., Household, Income and Labor Dynamics in Australia Survey), we find that both tangible bargaining power (measured by the shared wage) and intangible bargaining power (measured by financial control) significantly affect birth intervals, with their combined presence strengthening this effect. Greater female bargaining power tends to align fertility timing with the woman's preferences when there is a disagreement between partners. However, deviations from the model's predictions occur: career-oriented women may delay childbearing despite having both higher bargaining power and fertility preferences, whereas norm-conforming women may have children earlier than expected even when they possess greater bargaining power but lower fertility preferences.

To summarize, this thesis offers a comprehensive analysis of gender inequalities across macro, meso, and micro levels, with a focus on different national contexts, including China, the UK, and Australia. By examining institutional policies, organizational practices, and intra-household dynamics, it contributes to a better understanding of the mechanisms that perpetuate gender disparities in both labor and family outcomes.

For my family, with endless gratitude for their love, patience, and encouragement.

Acknowledgements

I would like to express my deepest gratitude to my principal supervisor, Dr. Silvia Avram, for her continuous support, insightful guidance, and unwavering encouragement throughout my PhD journey. Her academic expertise has been instrumental in shaping the direction and depth of this research. In particular, her rigorous approach to empirical methods greatly enhanced the development of the econometric models used in this thesis. More than academic support, she has also helped shape the way I think and encouraged me to believe in the value of my work. Her trust in me and her passion for research have deeply influenced me, and I am especially grateful for her encouragement to pursue meaningful contributions to the field of gender inequality.

I would also like to express my sincere appreciation to my second supervisor, Dr. Karon Gush, for her invaluable guidance and support throughout the course of this research. She upheld high standards for the overall structure and presentation of the thesis, and consistently encouraged me to refine my academic voice. I am especially grateful for her insightful reflections on the potential social contributions of this work, which challenged me to consider its broader relevance and societal impact. Her thoughtful advice and warm encouragement, both academically and personally, provided reassurance and strength during moments of uncertainty.

My thanks also extend to Dr. Daria Popova and Dr. Laura Fumagalli, who served as Chairs of my supervisory board. In addition to providing thoughtful feedback during annual reviews, they offered valuable advice on journal submissions, conference participation, and future career development. Their guidance throughout the doctoral process has been both generous and deeply appreciated.

Finally, I would like to thank my examiners, Prof. Alita Nandi and Prof. Man-Yee Kan, along with the anonymous reviewers, for their careful reading and constructive comments. Their evaluations have significantly contributed to enhancing the academic

rigor and clarity of this thesis. Moreover, the guidance and encouragement I received throughout this journey, particularly during the review and examination process, have inspired me to continue exploring this field and to pursue future research with renewed passion and purpose.

Contents

List of Tables

List of Figures

Chapter 1

How Does Fertility Relaxation Policy Affect the Motherhood Wage Penalty

Abstract

This paper examines the effect of fertility relaxation on the motherhood wage penalty, utilizing the China's two-child policy as quasi-experiment. Our findings reveal heterogeneity in the policy's effect. For one-child mothers, the policy change signals employers to update their beliefs about potential for future fertility, leading to a significant 9% increase in the wage penalty for them post-policy, compared to the non-mothers. Conversely, mothers who previously violated the policy by having a second child experienced an 8% decrease in their wage penalty as their illegal status was lifted post-policy. These outcomes are primarily attributable to changes in job discrimination rather than shifts in human capital. A further mechanism analysis suggests that the increase in the anticipatory wage penalty for one-child mothers is driven by statistical discrimination, whereas the decrease for two-child mothers is linked to taste-based discrimination post-policy.

1.1 Introduction

In labor economics, the relationship between female labor market participation and fertility choices highlights a significant economic dilemma. As more women enter the workforce, the growing focus on balancing careers and motherhood has gained increasing attention in both developed and developing countries. One significant consequence of this trend is the "motherhood wage penalty", which refers to the sustained wage reduction that women experience after giving birth. This penalty is typically measured longitudinally as the growing earnings gap between mothers and childless women (Correll et al., 2007; Lundborg et al., 2017; Kleven et al., 2019). To mitigate this issue, many countries have enacted policies aimed at reducing the motherhood wage penalty. These policies often include parental leave, maternity leave, and childcare support. However, while aimed at supporting families, these policies increase anticipated costs for employers, such as maternity insurance, maternity leave, and the need for replacement hires, due to the expectation of future fertility plans (Jessen et al., 2019). Thus, our research probes the theory of an "anticipatory motherhood wage penalty". This ex-ante penalty would emerge not from actual motherhood, but from employer discrimination based on the expectation of future fertility, which may be amplified by the policies designed to support families. To the best of our knowledge, the exploration of the anticipatory motherhood wage penalty prior to childbirth has been limited, with the available studies seeming to yield mixed results (Kleven et al., 2020; He et al., 2023; Agarwal et al., 2024).

This hypothesis can be tested within the quasi-experimental framework provided by the reform of China's two-child policy (TCP).[1] Before the TCP, women were restricted to having only one child under the one-child policy (OCP). During the OCP, hiring one-child mothers may have been perceived by employers as a more stable option, with a

[1] The two-child policy was nationwide, and no expert or media outlet had accurately predicted their implementation prior to its announcement (Jia et al., 2021; He et al., 2023; Agarwal et al., 2024).

lower likelihood of having to accommodate costs related to maternity leaves, replacement hires (Jessen et al., 2019). The relaxation of the OCP, which granted women the right to have a second child and was officially enacted on 28th December 2013 (2014, hereafter), serves as a pivotal shift in employers' expectation and belief.[2] This policy change signals to employers to update their beliefs that one-child mothers could potentially have another child at any time. As a result, in pursuit of maximizing profits, employers may engage in statistical discrimination against these mothers by offering them lower wages, anticipating the additional costs this change might entail.[3,4] One the other hand, for mothers who previously violated the OCP by having a second child, the motherhood wage penalty they face may stem from taste-based discrimination.[5] Employers might hold a bias against those mothers who flout the legal norms established by the OCP. However, the introduction of the TCP could mitigate this taste-based discrimination, potentially leading to an increase in wages for these mothers after that the policy change.

To estimate the effect of fertility relaxation policy on the motherhood wage penalty, we use a representative dataset, the China Family Panel Studies (CFPS), and employ a Difference-in-Differences (DiD) regression approach. Our baseline approach consists of comparing mothers of one-child to non-mothers to derive the motherhood child penalty for the first child, and then to test whether this changed with the switch from the OCP to the TCP. We argue that, under the assumption that non-mothers remain unaffected by the TCP, this approach identifies changes in the motherhood penalty that are driven by employers updating their expectation of how many children a one-child

[2] The initiation of the two-child policy, known as the selective two-child policy, rolled out across all provinces in January 2014 and fully enacted by September of the same year, applying to families where either partner is an only child. Subsequently, in 2016, the Chinese government enacted the universal two-child policy, permitting all couples to have two children. For more detailed information, please refer to Section 2.

[3] Some women may bear an additional child as a consequence of the policy, attributing any increase in the wage penalty to this child effect rather than to the policy itself. To isolate the child effect, we only concentrate on a subset of women who did not have an additional child after the two-child policy.

[4] Economists have traditionally modeled statistical discrimination as fully rational based on the signaling model, which also is called "rational stereotyping" (Phelps, 1972; Arrow, 1972; Coate and Loury, 1993).

[5] Taste-based discrimination involves bias in hiring and wage decisions driven by employers' personal prejudices against certain groups, independent of their qualifications (Becker, 1957; Charles and Guryan, 2008; Lang and Lehmann, 2012).

mother is planning to have. However, even though the policy's intended focus was on one-child and two-child mothers, it could also change employers' anticipation regarding the future fertility behavior of non-mothers. Under this scenario we would still identify a valid reform effect on the motherhood penalty, but it would be unclear whether the change in the motherhood penalty would be driven by wage changes of mothers, or of non-mothers. To address this concern, we exploit the fact that in 2014 the two-child policy was initially introduced as a selective policy (applying to families where either partner is an only child), before turning into a universal policy (permitting all couples to have two children) in 2016. This allows us to estimate the effects of the selective two-child policy within the group of only-child one-child mothers.

Our empirical analysis uncovers three primary findings. First, the wage penalty for the one-child mothers increases significantly by 0.09 log points, while for the two-child mothers it reduces significantly by 0.08 log points after the relaxation of OCP. To ascertain that our effects are indeed driven by one-child mothers, we conduct a triple DiD model using a treatment group women from households where either partners is only-child.[6] The triple DiD coefficient indicates a significant increase in the wage penalty for targeted only-child one-child mothers by 0.11 log points during the selective two-child policy period, confirming our conjecture that the effect of the two-child policy on the motherhood penalty is mainly driven by one-child mothers.

Secondly, we examine the mechanisms behind the post-policy increase in the wage penalty for one-child mothers due to statistical discrimination, and the reduction in the wage penalty for two-child mothers resulting from taste-based discrimination. For instant, if the policy's effect is rooted in statistical discrimination, younger mothers might face more pronounced effects. Our findings indicate that post-policy, younger one-child mothers face a wage reduction of 0.08 log points, whereas the wages of older one-child mothers show no significant change. Among these younger one-child mothers, those with children aged 4-7 years face the highest statistical discrimination, with their

[6] Under the selective two-child policy, the treatment group comprises one-child mothers from households in which either partner was born after 1982.

wage penalty increasing by 0.11 log points post-policy.[7] Furthermore, we extend our investigation through a comparative analysis of the public and private sectors, informed by OCP which historically enforced more stricter regulations on women in the public sector. This scenario suggests public sector employers had higher expectations of women having a second child post-policy. Our findings corroborate this, showing a post-policy increase in the wage penalty for public sector one-child mothers by 0.12 log points, whereas those one-child mothers in private sector see no such increase.

Third, to explore how taste-based discrimination affects the wage penalty for two-child mothers, we examine the illegal stigma effect's persistence. Specifically, we compare mothers who had a second child illegally under OCP with those who had a second child legally under TCP. Our analysis supports the hypothesis, revealing that during the TCP period, mothers who had their second child illegally under the OCP experience a lower log hourly wage by 0.11 log points compared to those who had their second child legally under the TCP. Nonetheless, we do not observe any significant changes in education levels among one-child or two-child mothers post-policy. Interestingly, post-policy, the working hours for one-child mothers even have significantly increased by 0.03 log points compared to non-mothers. This suggests that the human capital channel is unlikely to be a primary driver behind the increased motherhood wage penalty according to the policy.

This study considerably expands three strands of literature. Firstly, it contributes to research on fertility policy on the labor market outcomes. Most of the existing studies focus on the impact of fertility policies on the female labor market outcomes, such as maternity leave and parental leave (Havnes and Mogstad 2011; Dahl et al., 2016; Kleven et al., 2024), but they found the mixed effect. Moreover, no study has examined the "anticipation effect" of the fertility policy on the motherhood wage penalty, rather than on the broader gender wage gap (Kleven et al., 2024; He et al., 2023; Agarwal et al., 2024).

[7] This is attributed to the historical pattern where the typical gap between the first and second child falls within 4-7 years.

Secondly, although studies by Budig and England (2001; 2012), Anderson et al. (2002), and Killewald and Bearak (2014) have explored aspects of the motherhood wage penalty, none have specifically investigated how fertility policy intersects with the motherhood wage penalty, particularly in terms of the anticipatory effects of the motherhood wage penalty. Our research addresses this void by examining the mechanisms behind the anticipatory changes in the motherhood wage penalty, focusing on shifts in statistical and tasted-base discrimination following fertility policy changes.

Thirdly, our study contributes to the literature on fertility policy in China and other developing countries. Under the OCP, there was a significant reduction in the total fertility rate (TFR), and post-TCP, the TFR remained relatively stable (Zhang 2017; He et al., 2023; Agarwal et al., 2024). The fertility relaxation policy may lead to changes of anticipatory motherhood wage penalty in the labor market through the job discrimination channel, which in turn may have feedback effects on fertility behavior (Catalina and Jean, 2005; Schoen et al., 1999). Consequently, following the relaxation of fertility restrictions, individuals without a strong inclination towards parenthood may decide against having children, contributing to a decrease in lower-order births in the short term and a dramatic long-term decrease in the TFR.

The structure of this paper is organized as follows: The next section delves into the policy background. Section 3 introduces the dataset, while Section 4 outlines the empirical strategy employed. Section 5 presents the empirical analyses conducted. Following this, Section 6 explores the underlying mechanisms driving the observed effects. Section 7 is dedicated to a heterogeneity analysis. Finally, Section 8 concludes with a discussion.

1.2 Policy Background

The voluntary family-planning initiative launched in 1971 evolved into a more structured policy by 1979. By December 1982, the principles of this policy were embedded in the Constitution. It is important to clarify that this family-planning program is often associated with limiting couples to a single child, hence commonly known as the "one-child policy". Couples who surpassed the policy's birth limitations encountered strict penalties, such as the inability to secure local household registration for their offspring, facing heavy fines, or the risk of losing their employment.

After the implementation of the OCP in 1982, the total fertility rate (TFR) in China fell from 5.1 to 2.23, with a net decrease in population of 11.58 million. Since then, China has entered a new period with low fertility rates coexisting with an ageing population. In January 2014, the Chinese government enacted the selective two-child policy, which allows a couple to have two children if either partner is an "only-child".[8,9] However, as reported by the NHFPC[10], the TFR only increased from 1.60 in 2011 to 1.65 in 2014, and then it decreased to 1.60 in 2015 (see Appendix B).[11,12] In other words, the selective two-child policy appears not to have achieved its purpose of actively promoting fertility. The Chinese government decided to implement the universal two-child policy[13,14] in 2016. Compared to the selective two-child policy, this new policy allowed any couple to have two children, irrespective of their own sibling statuses. The aim of both policies was to relax the restriction on fertility for households and was expected to result in a substantial increase in births. However, as reported by NHFPC, the number of births in China did not increase significantly, and the TFR even decreased by 0.12 in two years following the introduction of universal two-child policy.

Nonetheless, the effect of the adjustment should be evaluated primarily by changes

---

[8] Notably, the "only-child" must have no siblings or half-siblings.

[9] Prior to the enactment of the selective two-child policy, which permitted couples to have two children if either partner is an "only-child", the Chinese government had introduced a different version of a selective two-child policy. This earlier policy allowed couples to have two children only if both partners were "only child". Implementation of this earlier policy varied regionally between 2000 and 2012; for instance, Guangdong province adopted it in 2002, while Hunan province followed suit in 2011. However, the target demographic of this early selective two-child policy accounted for only approximately 4% of China's total population, rendering its overall impact relatively limited (see Appendix C).

[10] National Health and Family Planning Commission of the People's Republic of China.

[11] The National Health and Family Planning Commission of the People's Republic of China (NHFPC) and many demographers have adjusted the statistics and concluded that the total fertility rate in the new century is around 1.6 (see Appendix B).

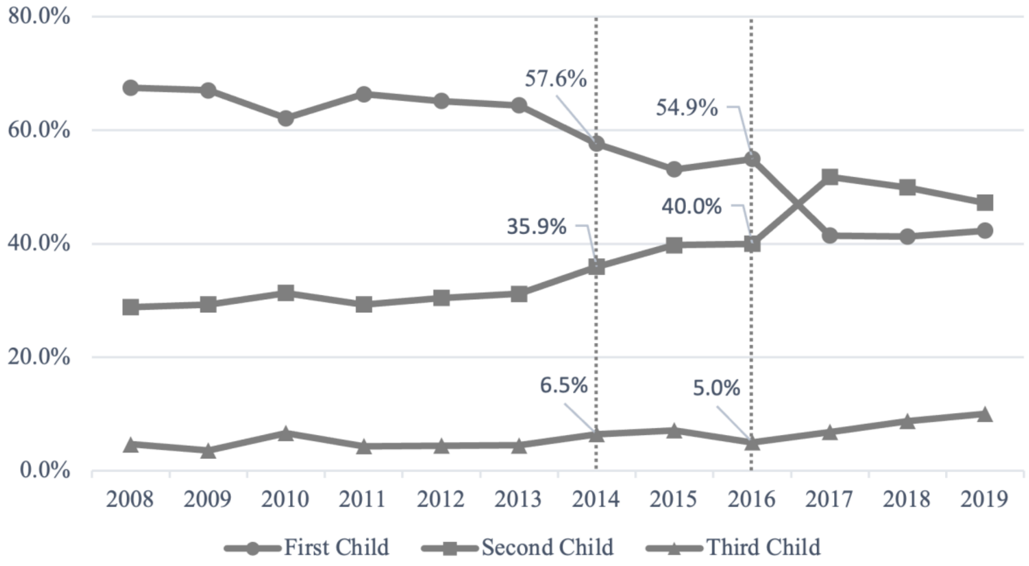
[12] Notably, the Chinese Population census does not count the total fertility rate after 2015.

[13] The universal two-child policy liberalizes restrictions on whether a couple is an "only-child", allowing every couple to have two children.

[14] In the Appendix A, further details about the policy are provided.

in second births, rather than broad shifts in overall fertility levels. A closer analysis of the fertility data reveals that second births have increased after the policy adjustment. Between 2013 and 2017, second child births rose from 5.11 million to 8.83 million, whereas first child births declined from 10.56 million to 7.24 million (see Appendix C). The proportion of second children in total births also rose from 31.2% to 51.8% (China Statistical Yearbook, 2020). In contrast, the proportion of first children in total births decreased from 65.6% to 41.4% after the two-child policy was implemented (see Figure 1.1). Thus, the insignificant improvement of the overall fertility level may be due to the decline of first births.

Figure 1. 1: Order-specific Fertility Rate



Source: China Statistical Yearbook, 2020

1.3 Data

The data for this study comes from the China Family Panel Studies (CFPS) follow-up surveys conducted in 2010, 2012, 2014, 2016, and 2018 (Xie et al., 2020). The CFPS is a nationally representative, large-scale household survey conducted by the China Social Science Survey Centre of Peking University. The survey was conducted in 162 counties across 25 provinces/municipalities/autonomous regions using the stratified multi-stage sampling method. The study covers a variety of topics including the economy, education, family relations, family dynamics, and population movement, with a target sample of 16,000 households. Based on the research objectives, the raw sample was modified in the following ways. Firstly, women between 16- and 49-year-old are kept. The minimum legal working age in China is 16, and the fertile period of women is considered to be between 14 and 49. Secondly, employed individuals with a labor income lower than the threshold set in the Minimum Labor Income Act are dropped.[15, 16] Thirdly, we exclude women with more than two children from our analysis as they represent a small fraction of our sample (only around 5.6%), and the majority of them are unemployed.[17] After these adjustments, the effective sample contains 15,405 women aged from 16 to 49, 61.3% of whom are employed.

The hourly wage is measured from the main job, including salary, bonuses, cash perks, and in-kind allowances, after taxes have been deducted. As Table 1.1 shows, the average hourly wage in the full sample is 14.52 yuan, with 15.08 yuan for the non-mothers and 14.14 for the mothers. Interestingly, mothers exhibit working hours that are similar to those of non-mothers.[18] Only about 61.3% of women are employed in the labor market. Among them, 73.5% women without children are employed and 63.5% of women with one child are employed, while only 43.9% of women with two children are employed. The primary explanatory variable in our study is the total number of children a woman has.[19] The mean number of children per working women in the full sample is 0.855. Additionally, 48.3% are the one-child mothers, and 18.6% women have

---

[15] In China, the Minimum Labour Income Act is about 460, 530, 700, 1000, 1280 yuan in 2010, 2012, 2014, 2016, 2018, respectively. Some observations below the Act are dropped because of misreported or underreporting.

[16] We are unable to exclude self-employed workers because the variable identifying self-employment is only available from 2014 onwards. Excluding self-employed workers would therefore compromise consistency across time periods, given that the DiD model compares individuals before and after 2014.

[17] This is likely due to the one-child policy, which made it rare for families to have three or more children.

[18] This pattern may be explained by longer working hours among mothers in both the private and public sectors. On the one hand, mothers are relatively more concentrated in the public sector, where jobs often provide greater stability and stronger support for work–family balance. On the other hand, mothers in the private sector tend to work longer hours due to greater financial pressures, including the need to support their children. However, this study does not explore these further.

[19] This approach is inspired by Waldfogel (1997), who posited that the presence of children can impact a mother's salary throughout her lifetime. This is because changes in human capital and the job discrimination, resulting from motherhood, are likely to influence a woman's entire working life. In the Chinese context, this is particularly relevant as legislation regarding birth policies focuses on the total number of children a woman has, rather than the ages of the children.

two children, reflecting the impact of the "one-child" policy previously imposed by the Chinese government. For more detailed information on individual, occupational, and family characteristics, refer to Table 1.1.

Table 1. 1: Summary Statistics

	All		Non-mothers		Mothers		
Variables	Mean	SD	Mean	SD	Mean	SD	Diff
Dependent							
Hourly Wage	14.52	33.03	15.08	23.16	14.14	38.11	0.945
Independent							
Num. child	.855	.704	-	-	1.382	0.486	-
One child	.483	.5	-	-	0.618	0.486	-
Two children	.186	.389	-	-	0.382	0.486	-
Personal Chara.							
Age (year)	33	8.8	24	4.7	37	6.7	12.72***
Education (1-9)	2.04	.425	2.15	.427	2.07	.392	0.081***
Work Chara.							
Working hours	219	71	214	66	220	68	-5.827***
Public (private)	.266	.442	.245	.43	.324	.468	-0.07***
Firm size	570	4734	623	5656	547	4276	75
Family Chara.							
Savings (k)	7.199	4.63	7.053	4.602	7.534	4.639	-0.48***
Housework hours	59	28	50	29	62	26	12.45***
Parents at home	.264	.441	.67	.47	.085	.28	0.58***
Parents education	2.491	1.105	2.544	.941	2.273	.907	0.27***
Heckman							
Employed rate	.613	.487	.735	.441	.556	.482	0.10***
Married Status	.773	.419	.225	.418	.946	.225	-0.72***

NOTE.—Data is sourced from the CFPS spanning 2010 to 2018. 'Wage' denotes hourly income from the primary job. 'Working hours' and 'Housework hours' are represented monthly. 'Parents at home' signifies the presence of the respondent's both parents in their household. 'Parents education' refers to the average educational attainment of the respondent's parents and is measured on the same 1–9 scale as the respondent's own education. *** p < 0.01.** p < 0.05.* p < 0.1.

1.4 Empirical strategies

To estimate the causal effect of the policy on the motherhood wage penalty, we employ a difference-in-differences (DiD) model. The selective two-child policy, rolled out across all provinces in 2014, applied to families where either partner was an only child.[20] This expanded eligibility encompassed around 18% of the population (refer to Appendix D for more details). With the universal two-child policy in January 2016, the policy's application became nationwide. Importantly, the application of both the selective and universal two-child policies was uniform across the country, eliminating the feasibility of distinguishing treatment and control groups based on geographical differences.[21]

Building on the above policies, we initially assume that non-mothers are unaffected by the two-child policy, an assumption grounded in the policy's aim towards mothers with one child and two children. Additionally, women who are currently non-mothers may be perceived as not having concrete short-term plans for childbirth. Hence, our "control group" consists of non-mothers, whereas the "treatment groups" include one-child and two-child mothers in the baseline model. However, concerns remain that employers may still expect some non-mothers, especially younger ones, might choose to have two children in the long-term post-policy change. Such anticipatory actions by employers could introduce bias into the DiD analysis, as both control and treatment group affected by the policy.

To address this concern, we conduct two robustness checks by capitalizing on the distinctions in family types targeted by the various iterations of the two-child policy. Specifically, the selective two-child policy of 2014 targeted families where "either partner is an only child", while the universal two-child policy of 2016 applied to all families. Therefore, in the 2014 analysis, women in couples where at least one partner is an only child serve as the "treatment group", as they are eligible under the policy.[22]

[20] The restricted version of the selective two-child policy, applicable only to households where both parents were only children, was introduced in 1984. Consequently, this target demographic constitutes a relatively minor segment of the population, accounting for approximately 3% of the total (see appendix c). Furthermore, the implementation of the selective two-child policy varied regionally between 1984 and 2012; for example, Guangdong province adopted this policy in 2002, whereas Hunan province did so in 2011. Given its limited scope, our analysis shifts focus to the subsequent selective two-child policy.

[21] The selective two-child policy was progressively introduced across various provinces from January 2014 and was comprehensively enforced nationwide by June of the same year. However, the CFPS predominantly collected its data in July and August, with over 90% of the data gathered during these months. This timing significantly constrains the ability to delineate treatment and control groups based on geographical variations.

[22] Despite the brief two-year interval between the selective and universal two-child policies, it is important to

Addressing the challenge of accurately identifying women belonging to the specified "treatment group" involves utilizing the one-child policy initiated in 1982 as a reference point.[23] We operate under the assumption that individuals born after 1982 are likely to be an "only child", while those born before or in 1982 are not (additional details can be found in appendix D). This presumption allows us to classify "only child" non-mothers as the treatment group and non "only child" non-mothers as the control group. Should the non-mothers truly remain unaffected by the policy, we anticipate the treatment effect to be insignificant post-policy.

Except for the non-mothers, we apply a similar identification strategy for a robust test focusing on one-child mothers. The control group consists of one-child mothers from households where both partners were born before or in 1982, indicating they were not directly targeted by the selective two-child policy. The treatment group comprises women from households where either partner was born after 1982.[24,25] Additionally, we incorporate this treatment dummy in the analysis of the 2016 universal two-child policy to perform a universal treatment test. Given that the selective two-child policy of 2014 was only applicable when either partner is an only child, we anticipate a significant treatment effect in this context. In contrast, given that the universal two-child policy of 2016 was applicable to all households, we anticipate an insignificant treatment effect in this scenario (more details see Section 1.5.C).

Additionally, heterogeneity might affect the wage regression, particularly in terms of wage levels varying across years, provinces, and industry. To eliminate the time-

contextualize this within the more extensive timeline of the one-child policy, which spanned over 32 years. Employers in 2014, accustomed to the longstanding one-child policy framework, would likely not have anticipated the rapid introduction of the "universal two-child" policy just two years later, in 2016.

[23] In the CFPS, identification of an "only child" can be directly through the question, "*How many brothers/sisters do you have?*" However, this question was only posed in the 2010 survey. Given the CFPS's low average number of repetitions (1.6 across five waves), this creates a substantial challenge in utilizing this method for identifying "only children" in subsequent periods.

[24] To process it, we match the household information to the women. In the CFPS, accurately identifying a woman's husband presents a challenge due to the data structure, where individuals residing together are assigned a shared family ID. Consequently, we match women to other individuals using the household ID, resulting in multiple records that could potentially represent a woman's husband, father, or son. To refine our matching process and more accurately identify spousal pairs, we exclude records where the absolute age difference between the woman and the matched individual is greater than 25 years.

[25] In Section 1.5C, we describe in detail how to construct the triple-DID to more precisely capture the policy's treatment effect on one-child mothers from households where at least one partner is an only child.

invariant unobserved factor that may affect the wages, we employ the fixed effect DiD model. We do not use individual fixed effects because the average number of observations per respondent across the five collections is only 1.6, indicating that the data are closer to cross-sectional (although the CFPS is designed as a longitudinal). Overall, the fixed effect DiD model can be expressed by:

$$lwage_{it} = \sum_{k\in\{1,2\}} \left(\beta_k D_{it}^{(k)} + \gamma_k D_{it}^{(k)} P_t\right) + X_{it}\eta + \varphi_t + \gamma_o + \mu_p + \varepsilon_{it} \tag{1}$$

where $lwage_{i,t}$ represents the log of hourly wage of female respondent i at time t; $D_{it}^{(k)}$ denotes a group of dummy variables that indicate whether the woman has one child and two children ($k \in \{1,2\}$), in contrast to control group of non-mothers. P_t is a dummy variable signifying the two-child policy, assigned a value of 1 for year in and after 2014 and 0 for all other years. X_{it} is a vector of control variables, including personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size).[26] Additionally, we include year fixed effects φ_t, industry fixed effect γ_o, province fixed effect μ_p.

Nonetheless, there are five principal concerns regarding the identification of the causal impact of the fertility relaxation policy on the motherhood wage penalty. Firstly, it is necessary to consider that some women may bear an additional child as a consequence of the policy, attributing any increase in the wage penalty to this child effect rather than to the policy itself. To isolate the child effect, we only concentrate on a subset of women who did not have an additional child after the selective two-child policy. By doing so, we can mitigate the impact on the motherhood wage penalty that arises from human capital impacts due to having an additional child. Secondly, in the

[26] In general, controlling for these characteristics addresses compositional changes in these variables and helps to ensure that any such changes are not driving the results, since the policy effect is identified from time variation.

DiD framework, ensuring unbiased causal inference hinges on the establishment of equivalent groups for analysis. To achieve a balanced and precise comparison, we use Propensity Score Matching (PSM) with nearest neighbor matching based on the observed personal controls. This methodological approach leads to the exclusion of 114 observations, enhancing the comparability of treatment and control groups.

Thirdly, the DiD methodology could be influenced by confounding events. According to our assessment, no event in 2014 and 2016 seems to have selectively affected the motherhood wage penalty. However, to rigorously validate our analysis, we conduct a permutation test as a robustness measure. In this test, instead of employing the actual data on mothers and non-mothers, they are randomly allocated while ensuring a consistent ratio. Following this, we implement the model (1) to evaluate the policy's effect on this placebo cohort and document the resulting DiD estimate. By repeating this process 500 times, we are able to create a density plot of the placebo DiD estimates.

Fourthly, the validity of causal inference in the DiD approach also relies on the assumption of parallel trends. This assumption posits that, absent any policy alterations, the average outcomes for the control and treatment groups would have followed identical trends over time. To substantiate this critical assumption, we employ the event study approach. A fifth issue arises from self-selection into employment, which presents two concerns. Firstly, prior to the policy, approximately 67% of women were employed. If mothers opt out of employment upon having a child, then the wage regression analysis may be biased due to observing only a truncated sample. Secondly, the policy itself could influence employment patterns. Should women either choose to leave employment or face discrimination leading to unemployment post-policy, excluding these individuals from the analysis could lead to an underestimation of the policy's effect. To rectify this self-selection, we use a Heckman correction, with "marital status" as the exclusion restriction based on its impact on labor force participation but not on wages for women (Heckman, 1997). There are some papers where Heckman himself has used marital status in the first stage of his selection model (Heckman, 1974;

Heckman and MaCurdy, 1980). However, later research has generally indicated that marriage leads to some growth in wages for women, calling into question the validity of marital status as an exclusion restriction (Hill, 1979; Krashinsky, 2004). However, Meng et al. (1997) found no evidence of marital status affecting female workers' wages once accounting for education, occupation and other factors in China. Thus, in the context of our study, we assume that marital status remains a valid exclusion restriction. We also conduct a series of tests to demonstrate that "marital status" serves as a valid instrument variable in the subsequent analyses (see Table 1.3). Applying the Heckman two-step model typically involves the following stages. In the first stage, a Probit model predicts the likelihood of employment ($employ_{it}$) for individual i at time t, formulated as:

$$Probit(employ_{it}) = \zeta marr_{it} + \sum_{k\in\{1,2\}} \left(\beta_k D_{it}^{(k)} + \gamma_k D_{it}^{(k)} P_t\right) + X_{it}\eta + \varphi_t + \gamma_o + \mu_p + \varepsilon_{it} \quad (3)$$

where $marr_{it}$ indicates marriage status. Following this, we calculate the inverse Mills ratios imr_{it} to address selection bias. The second stage involves estimating the primary regression model for respondents who are employed, incorporating the imr_{it} to adjust for selection bias:

$$lwage_{it} = \sum_{k\in\{1,2\}} \left(\beta_k D_{it}^{(k)} + \gamma_k D_{it}^{(k)} P_t\right) + \tau imr_{it} + X_{it}\eta + \varphi_t + \gamma_o + \mu_p + \varepsilon_{it} \quad (4)$$

1.5 Empirical Results

This section discusses the effect of fertility relaxation policy on motherhood wage penalty. The analysis proceeds in four steps. Initially, a naïve DiD model is utilized to examine how the policy changes the motherhood wage penalty.[27,28] Subsequently, a series of robustness checks, including Heckman and triple-DiD models, are conducted to further validate the policy's effect on the motherhood wage penalty. The third step involves examining potential mechanisms behind these observed effects. The analysis concludes with a heterogeneity analysis to uncover varying impacts across different groups.

A. Main results

Table 1.2 shows the effect of fertility relaxation policy on motherhood wage penalty by number of children. Recognizing age as a crucial determinant of both fertility and wages, we initially adjust for age and its square in column 1 to account for the age effect comprehensively. Beyond capturing the age effect, we also include controls for education level and urban residency, acknowledging these as significant factors influencing both fertility rates and wage levels in China, as evidenced by previous studies (Jia et al., 2013; Yu and Xie, 2014). The results show that the hourly wage gap between one-child mother and non-mothers is about -0.02 log points prior to the policy implementation, a finding that lacks statistical significance. In contrast, for the two-child mothers, the wage gap is significantly larger at -0.24 log points, with statistical significance at the 1% level. These wage gaps also could stem from other various factors, with the primary consideration being the control of differences in observable,

---

27 We also document how the "two-child" policy changes fertility behavior for each birth order by using sequential Logit model in Appendix E.

28 As displayed in Figure 1.1 and Appendix E, the "two-child" policy appeared to influence the fertility decisions of women in two significant ways: fewer non-mothers transitioned into one-child mothers, and a greater number of one-child mothers transitioned into two-child mothers.

among mothers and non-mothers in the workforce and family characteristics.[29] Taking these factors into account and controlling for firm and family characteristics (see column 2), the wage gap for one-child and two-child mothers remains similar compared to model (1).

It is important to note that model (2) does not incorporate spatially fixed effects, predicated on the anticipation that the magnitude of the motherhood wage penalty varies spatially. But as we discussed before, the one-child and two-child policy is enforced differently in different provinces and industry. When province and industry fixed effects are accounted for, the observed motherhood wage penalties for one-child mothers remain insignificant, while for two-child mothers, it reduces by 0.18 log points ($p<0.01$), prior to the policy implementation, as shown in column 3. This finding diverges from Lalive and Zweimüller (2009), who observed that mother's wage penalties still persist approximately 84 months after childbirth. Our results imply that one-child mothers were somewhat insulated from wage penalties due to the one-child policy. This insulation could be attributed to employers having lower expectations for one-child mothers to have more children in the future during the one-child policy period. In contrast, the high wage penalty for the two-child mothers can be primarily attributed to the prohibition of second births under one-child policy, alongside to the adverse effects on human capital associated with raising multiple children.

After the implementation of the two-child policy, there is a notable divergence in the motherhood wage penalty: it increases for one-child mothers while it decreases for two-child mothers, as detailed in column 3. Specifically, the motherhood wage penalty for one-child mothers significantly increases by approximately 0.09 log points ($p<0.01$) following the two-child policy. This trend corroborates our hypothesis that the policy possibly sends a signal that one-child mothers may have a second child anytime.

[29] Prior research emphasizes the significant role of both firm and family characteristics in influencing the motherhood wage penalty. For instance, Duvivier and Narcy (2015) note that larger corporations often implement more "family-friendly" policies to alleviate the motherhood wage penalty. Additionally, wealthier and larger families often mitigate the motherhood wage penalty by engaging childcare support, either through assistance from the women's parents or by employing childcare providers (Ruhm, 2004).

Consequently, employers might anticipate additional costs related to maternity leave and reduced productivity following childbirth. As these anticipated costs mount, profit-motivated employers make rational predictions and exhibit statistical discrimination against one-child mothers, leading to lower wages for these women. This shifting labor market landscape may inadvertently deter non-mothers from undertaking the transition to motherhood, thereby contributing to a decline in the first-child fertility (see Figure 1.1 and Appendix E).

In contrast, the wage penalty for two-child mothers significantly reduces by about 0.09 log points ($p<0.01$) after the two-child policy. Since it was the first time China recognized the legality of having two children, employers became less likely to taste-based discriminate against two-child mothers who were considered rule breakers during the one-child policy period, leading to a reduction in the wage penalty. As a result, one-child mothers with a strong inclination toward larger families are more likely to have a second child after the policy (see Figure 1.1 and Appendix E). Nonetheless, the surge in the number of two-child mothers can be attributed to a dual-factor incentive: firstly, the amplified wage penalty for one-child mothers acts as a deterrent against remaining with a single child, and secondly, the diminished wage penalty for two-child mothers encourages one-child mothers to expand their families further.

Table 1. 2: The effect of FR Policy on the Motherhood Wage Penalty

Log Hourly Wage	Person	W & F	Fixed	PSM	Exclude
	(1)	(2)	(3)	(4)	(5)
	Baseline motherhood wage penalty (during the one-child policy), conditional on number of children				
One child	-0.02	-0.03	-0.02	-0.02	-0.02
	(0.027)	(0.028)	(0.026)	(0.027)	(0.027)
Two children	-0.24***	-0.23***	-0.18***	-0.18***	-0.18***
	(0.035)	(0.035)	(0.034)	(0.034)	(0.034)
	Interaction of motherhood wage penalty with two-child policy				
One child × Policy	-0.08***	-0.10***	-0.09***	-0.08***	-0.09***
	(0.029)	(0.029)	(0.027)	(0.027)	(0.028)
Two children × Policy	0.10***	0.08**	0.09***	0.10***	0.08**
	(0.036)	(0.036)	(0.034)	(0.034)	(0.034)

	Pre-trend test (2010 vs. 2012)				
One child × 2010	-0.07	-0.02	-0.04	-0.03	-0.03
	(0.046)	(0.046)	(0.043)	(0.043)	(0.043)
Two children × 2010	-0.02	0.05	0.01	0.01	0.01
	(0.060)	(0.059)	(0.056)	(0.056)	(0.056)
N	10,399	10,399	10,399	10,285	9,860
R^2	0.3069	0.3198	0.4053	0.4038	0.4020
Year FE	✓	✓	✓	✓	✓
Age & Edu	✓	✓	✓	✓	✓
Work & Family	-	✓	✓	✓	✓
Province & Industry FE	-	-	✓	✓	✓
Matched	-	-	-	99%	99%

NOTE.—Estimates from regressions on the log of the hourly wage for the working women. In DiD framework, we use non-mothers as the control group. The DiD effect on the hourly wage is captured by the interaction terms. *Policy* represents the two-child policy in 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in column (5). To examine pre-policy trends, we employ an DiD event study approach, omitting the year 2012 to serve as the reference. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

As we discussed before, there are five main considerations. Firstly, within the DiD analytical framework, the accuracy of causal inference critically depends on creating equivalent comparison groups. To facilitate a balanced and precise analysis, we employ PSM with nearest neighbor matching. As demonstrated in column 4, the coefficients of motherhood wage penalty for one-child and two-child mothers, both before and after the policy's implementation, align closely with those found in model (3). This analysis confirms that the wage penalties for one-child and two-child mothers, both before and after the policy's implementation, are robust. Secondly, some women may bear an additional child during the two-child policy period, attributing any increase in the wage penalty to this child effect. To isolate the child effect, we only concentrate on a subset group of women who did not have an additional child during the two-child policy period. Specifically, we observe the wage penalty coefficient of one-child and two-child mothers increase by 0.09 log points ($p<0.01$) and reduce by 0.08 log points ($p<0.05$) after the two-child policy, respectively, in this subset group. Given that these coefficients mirror those found in models (3) and (4), we reaffirm the robustness of our findings on the wage penalties for one-child and two-child mothers.

Thirdly, the validity of causal inference in the DiD approach also relies on the assumption of parallel trends. Panel of pre-trend test in Table 1.2 presents a parallel trend test to evaluate the above assumption. Specifically, for the period leading up to the policy implementation, the year 2010, compared to the reference year of 2012, the analysis indicates a wage penalty of -0.03 log points for one-child mothers and 0.01 log points for two-child mothers in both including and excluding groups (see columns 4 and 5). Given the standard errors, 0.043 for one-child mothers with and 0.056 for two-child mothers, these variations are not statistically significant. These patterns, particularly the pre-policy observations for both groups, support the parallel trends assumption.

Fourthly, the consideration that non-mothers might be impacted by the policy through employers' anticipatory actions could potentially bias the DiD analysis. To strengthen the robustness of our findings, we delineate "only child" non-mothers and non "only child" non-mothers as the treatment and control groups, respectively. Our analysis is concentrated on comparing the periods immediately before and after the policy implementation (i.e., 2012 and 2014). According to the results detailed in Appendix F, the coefficient of the treatment effect is 0.027 log points, and with a standard error of 0.173, this effect is statistically insignificant. This outcome suggests that the policy does not significantly impact non-mothers, reaffirming that the observed changes in our baseline model are primarily attributable to the effects on mothers. In general, this result reinforces the robustness of the initial assumption that non-mothers are not impacted by the two-child policy.

Fifthly, the DiD model may be affected by confounding events. To address this concern, we perform a permutation test. As illustrated in Appendix G, the placebo DiD estimates cluster around zero and follow a normal distribution. Crucially, our baseline DiD estimates shown in column 3 of Table 1.2 (represented by the vertical dashed line in Panels (a) and (b) of Figure G.1.1) significantly deviate from these placebo estimates, which lends additional support to the robustness of our findings.

B. The effect of policy on Employment Probability and Self-selection

Furthermore, one of the issues arises from self-selection into employment (for more detail, refers to Section 5). To rectify this self-selection, we employ the Heckman correction, with "marital status" as the exclusion restriction. As seen in column 1 in Table 1.3, before the policy, the likelihood of employment among one-child mothers and two-child mothers is lower by about 0.03 ($p>0.1$) and 0.47 ($p<0.01$) on the Probit scale than that of non-mothers, respectively. Following the two-child policy, the employment probability for one-child mothers significantly declines by approximately 0.14 on the Probit scale ($p<0.05$). Conversely, for two-child mothers, the policy brings about a significant reduction in the employment penalty, by about 0.19 on the Probit scale ($p<0.01$). Interestingly, employment and wage patterns tend to move in tandem; a policy influencing wages directly impacts employment patterns, as wages and the decision to engage in the labor market are intrinsically linked (Goldin and Lawrence, 2008).

Given the significant impact of the policy on employment patterns, overlooking employment self-selection could introduce potential bias into the wage regression. The outcomes presented in column 2 reveal that, after correcting for self-selection, the motherhood wage penalty for one-child and two-child mothers, both before and after the policy, remains largely unchanged compared to the baseline model in column 5 of Table 1.2. Crucially, the insignificance of the inverse Mills ratio suggests the absence of a self-selection problem in our wage analysis. The robustness of our results hinges on the validity of our instrument variable, "marital status". To substantiate its validity, we carry out tests aimed at satisfying two principal assumptions: Relevance and Exogeneity. The Relevance assumption is confirmed through a strong negative correlation between marital status and labor market participation, evidenced by a coefficient of -0.70 on the Probit scale (p<0.01), as shown in column 1. Although empirically validating Exogeneity is challenging, our reduced form in column 3 indicates that marital status is not significantly correlated with log hourly wage, thereby lending some credence to the instrument's validity. This analysis further solidifies the conclusion that our wage regression is not affected by self-selection issues.

Table 1. 3: The effect of FR Policy on the labor market outcomes

	Heckman DiD		Reduce Form
	Employment (F)	Wage (S)	Wage
	(1)	(2)	(3)
	Baseline motherhood penalty (during the one-child policy), conditional on number of children		
One Child	-0.03	-0.02	-0.02
	(0.053)	(0.028)	(0.029)
Two Children	-0.47^{***}	-0.16^{***}	-0.17^{***}
	(0.060)	(0.041)	(0.036)
	Interaction of motherhood penalty with two-child policy		
One Child × Policy	-0.14^{**}	-0.08^{***}	-0.09^{***}
	(0.057)	(0.028)	(0.028)
Two Children × Policy	0.19^{***}	0.07^{**}	0.08^{**}
	(0.062)	(0.036)	(0.034)
	Relevance and Exogeneity Test		
Married	-0.70^{***}	-	-0.01
	(0.044)	-	(0.021)
imr	-	-0.04	-
	-	(0.052)	-
N	16,196		9,860
R^2	0.4069		0.4068
Year & Prov & Ind FE	✓		✓
Person & Work & Fam	✓		✓

NOTE.—Estimates from regressions on employment status for all women and the log of the hourly wage for the working women. Employment status is modelled using a Probit model. In DiD framework, the control group is non-mothers. *Policy* represents the two-child policy in 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in all columns. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

C. Robustness check: Birth cohort

The above results rely on the assumption that non-mothers are unaffected by the two-child policy, but the employer expectations that non-mothers might bear a second child in the future can introduce bias into the DiD analysis. To address this concern, we further identify the "control group" of women from households where both partners were born before or in 1982 as they are not the primary targets of the selective two-child policy of 2014. Additionally, we incorporate this treatment dummy in the analysis of the 2016 universal two-child policy to perform a universal treatment test (see Section 1.4):

$$\begin{aligned} lwage_{it} = {} & \gamma_1 D_{it}^{(1)} + \gamma_2 tr_{it} + \gamma_3 S_t tr_{it} + \gamma_4 U_t tr_{it} + \gamma_5 D_{it}^{(1)} S_t + \gamma_6 D_{it}^{(1)} U_t \\ & + \gamma_7 D_{it}^{(1)} tr_{it} + \boldsymbol{\gamma_8 D_{it}^{(1)} tr_{it} S_t} + \boldsymbol{\gamma_9 D_{it}^{(1)} tr_{it} U_t} + X_{it}\eta + \varphi_t \\ & + \gamma_o + \mu_p + \varepsilon_{it} \end{aligned} \tag{5}$$

The results, as illustrated in column 1 in Table 1.4, the treatment effect we examine through the triple DiD model exhibits that the selective two-child policy's influence on wage penalty faced by one-child mothers where either partner is an only-child. Specifically, the triple DiD coefficient of $One \times Tr \times S$ indicates a significant increase in the wage penalty by 0.11 log points (p<0.01). This significant outcome highlights the distinct influence of the selective two-child policy on increasing the wage penalty for one-child mothers among the targeted group. Remarkably, this coefficient magnitude is similar to the DiD coefficient that reported in column 5 of Table 1.2, reinforcing the baseline results that one-child mothers are impacted by the two-child policy is robust. Moreover, we employ another triple DiD coefficient, $One \times Tr \times U$, as a universal treatment test. Given that the "universal two-child" policy potentially affects all cohorts, irrespective of whether one-child mothers come from only-child or non-only-child families, we anticipate this coefficient to be close to zero and statistically insignificant. The findings in column 1 bolster our hypothesis: the treatment effect on the wage penalty for one-child mothers across both cohorts following the universal two-child policy is minimal, at 0.02 log points, with a standard error of (0.029), thereby indicating no significant difference. This outcome lends further credence to the robustness of our chosen treatment and control groups in examining the motherhood wage penalty under the two-child policy.

Table 1. 4: The triple DiD effect of FR Policy on the Anticipatory Motherhood Wage Penalty

	Triple DiD	PSM
	(1)	(2)
	Interaction of motherhood wage penalty with policy of selective (S) and universal (U) two-child policy	
One Child × Tr × S	-0.11***	-0.11***
	(0.000)	(0.001)
One Child × Tr × U	0.02	0.02
	(0.029)	(0.029)
N	6,247	6,228
R^2	0.3858	0.3857
Year & Prov & Ind FE	✓	✓
Person & Work & Fam	✓	✓
Matched		99%

NOTE.—Estimates from regressions on the log of the hourly wage for the working women. The policy effect on the hourly wage is captured by the triple interaction terms. *S* represents the "selective two-child" policy in 2014, and *U* represents the "universal two-child" policy in 2016. The term *Tr* refers to whether women from households where both partners were born before or in 1982. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in all columns. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

1.6 Mechanism

The earlier sections confirmed that the two-child policy can escalate the wage penalty for one-child mothers, while mitigating the wage penalty for two-child mothers. This wage penalty fluctuation may be traced back to two principal channels: human capital (Becker, 1985; Mincer, 1989) and job discrimination (Budig and England, 2001; Gough and Noonan, 2013).[30]

A. Statistical discrimination

Job discrimination can be divided into statistical and tasted-based discrimination. In this section, we delve into the hypothesis that the post-policy increase in the wage penalty for one-child mothers is driven by statistical discrimination. To confirm this argument, we analyze heterogeneous effects by age and sector. On the one hand, if the policy's effect is rooted in statistical discrimination, its impact might differ based on the ages of both the mother and her child. This notion relies on the belief that older mothers are less inclined to have another child, implying that younger mothers might face more pronounced effects of statistical discrimination. To explore this hypothesis, identifying an age threshold for mothers becomes essential. Below this threshold, employers may statistically discriminate against younger mothers, assuming a higher likelihood of them having a second child following the policy. Beyond this threshold, however, the likelihood of engaging in statistical discrimination diminishes, as employers may perceive these mothers as less likely to have additional childbirth post-policy. Upon determining the threshold for mothers' ages, the next step involves examining the child's age to gauge the extent of statistical discrimination effects. For this purpose, we categorize children's ages into three groups reflective of stages in the

[30] The two-child policy has limited effect on the motherhood wage penalty through compensating differential channel. Women find it challenging to switch between sectors in China, and within each sector, family-friendly policies tend to be consistent (Zhou and Xie, 2019). Also, the Bivariate Probit regression test results indicate the p-value of correlation coefficient is 0.1875, suggesting no sector selection problem in our findings.

Chinese educational system: 0-3 years (before kindergarten), 4-7 years (kindergarten), and 8+ years (primary school). If employers exhibit statistical discrimination against one-child mothers, we anticipate the 4-7 age group to experience the highest wage penalty post-policy, as the typical age gap between the first and second child in China falls within this range. On the other hand, we conduct another analysis through a comparative analysis of the public and private sectors, informed by China's one-child policy which historically enforced more stricter regulations on women in the public sector. Consequently, our result reveals that women employed in the public sector were half as likely to have a second child compared to those in the private sector, as detailed in Appendix H. This discrepancy underscores a higher expectation among employers in the public sector regarding the likelihood of women having two children after the policy, indicating that one-child mothers in the public sector may face heightened statistical discrimination post-policy.[31]

The age heterogeneity findings presented in Table 1.5, consistent with the baseline results, indicate that during the one-child policy period, the motherhood wage penalty was not evident for one-child mothers. However, after the two-child policy, younger one-child mothers (aged 36 and below) experience more pronounced statistical discrimination, leading to a wage reduction of approximately 0.08 log points ($p<0.01$). In contrast, older one-child mothers (aged over 36) do not face any statistical discrimination, even their wages increase by about 0.03 log points although it is insignificant (see columns 1 and 5). As previously outlined, if statistical discrimination is present, it would be most noticeable within the child's age group of 4-7 for younger one-child mothers. Our analysis supports this hypothesis, revealing that within the 4-7 child's age group, younger one-child mothers face the highest level of statistical discrimination, with their wage penalty increasing by 0.11 log points ($p<0.01$) post-policy. For children aged 0-3, the discrimination is less severe, with a decrease in wage

[31] Moreover, due to the policy may affect the women wage in the different sectors, women may self-select to work in the different sector to mitigate wage penalty. But unlike the European public sector, it is difficult for workers to freely move between the two sectors in China (Zhou and Xie, 2019). Also, the Bivariate Probit regression test results indicate the p-value of correlation coefficient is 0.1875, suggesting no sector selection problem in our findings.

penalty of 0.08 log points (p<0.1) post-policy, as employers likely do not anticipate mothers to have a second child in the immediate future. Finally, for children aged above 8 years, the wage penalty change post-policy is minimal and statistically insignificant, decreasing by 0.03 log points (p>0.1), due to the rare occurrence of having a second child with such an age gap in China prior to the policy.

Table 1. 5: The effect of FR Policy on the Anticipatory Motherhood Wage Penalty through Statistical Discrimination

	Women's age: Age <= 36				Age > 36
Child's age	Full	0-3	4-7	8+	Full
	(1)	(2)	(3)	(4)	(5)
	Baseline motherhood wage penalty (during the one-child policy)				
One Child	0.02	-0.03	0.05	-0.04	-0.03
	(0.026)	(0.037)	(0.034)	(0.038)	(0.093)
	Interaction of motherhood wage penalty with two-child policy				
One Child × Policy	-0.08***	-0.08*	-0.11***	-0.03	0.03
	(0.032)	(0.048)	(0.043)	(0.048)	(0.130)
N	5,496	3,790	3,987	3,807	2,709
R^2	0.4052	0.4097	0.3997	0.4021	0.4474
Year & Prov & Ind FE	✓	✓	✓	✓	✓
Person & Work & Fam	✓	✓	✓	✓	✓

NOTE.—Estimates from regressions on the log of the hourly wage for the working women. In DiD framework, the control group is non-mothers. *Policy* represents the two-child policy in 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women who have two children in all columns. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

The findings on sector heterogeneity, detailed in Table 1.6, reveal that during the one-child policy period, the wage penalty was not significant for one-child mothers within the public sector; a slight wage premium was observed, although it was not statistically significant, with a coefficient of 0.06 log points (p>0.1). This suggests that one-child mothers in the public sector were somewhat shielded from the wage penalties due to the stringent enforcement of the one-child policy. In contrast, in the private sector, where adherence to the one-child policy was less stringent and employers expect a higher likelihood of one-child mothers having a second child, lower wages were offered, evidenced by a coefficient of -0.08 log points ($p<0.05$). However, after the two-child policy, the protective effect of the one-child policy on wage penalties for one-child mothers in the public sector has diminished. As a result, the wage penalty for these mothers in the public sector saw a significant rise, increasing by 0.12 log points ($p<0.05$) post-policy. In the private sector, where there was already an anticipation of mothers having a second child during the one-child policy, the change in the wage penalty for one-child mothers post-two-child policy was not significant, with a decrease of -0.04 log points ($p>0.1$).

Table 1. 6: The effect of FR Policy on the Anticipatory Motherhood Wage Penalty by Sector

	Private	Public	Diff.
	(1)	(2)	(3)
	Baseline motherhood wage penalty (during the one-child policy)		
One child	-0.08^{**}	0.06	-0.14^{***}
	(0.032)	(0.051)	-
	Interaction of motherhood wage penalty with two-child policy		
One child × Policy	-0.04	-0.12^{**}	0.08^{***}
	(0.032)	(0.052)	-
N	5,601	2,319	-
R^2	0.4182	0.3665	-
Year & Prov & Ind FE	✓	✓	-
Person & Work & Fam	✓	✓	-

NOTE.—Estimates from regressions on the log of the hourly wage for the working women. In DiD framework, the control group is non-mothers. *Policy* represents the two-child policy in 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women who have two children in all columns. Column 3 reports the coefficient differences between the private and public sectors by employing Bootstrap and Permutation tests to assess the differences in coefficients between the two groups, conducted 500 times. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

B. Taste-based discrimination

To investigate the reduction in wage penalty for two-child mothers as a result of taste-based discrimination, we consider the persistence of a stigma effect. Specifically, we compare two groups of two-child mothers: those who had their second child before the two-child policy (referred to as the 'illegal' group) and those who had their second child after the policy (the 'legal' group). During the two-child policy period, we expect the wage penalty for the illegal group to decline due to reduced stigma, while the legal group is not subject to stigma-related wage penalties in the first place. Moreover, to focus on the effect of discrimination, this analysis only considers children aged 2 and older. Additionally, considering the different age distributions of children born within and outside the two-child policy period, controlling for the child's age becomes essential to our analysis. Furthermore, we include one-child mothers as part of our robustness check. Given that having one child was allowed both before and after the two-child policy, we do not expect to observe significant differences between these groups.

The results displayed in Table 1.7 corroborate our hypothesis regarding the persistence of a stigma effect among two-child mothers, reflecting taste-based discrimination. As illustrated in column 3, the log hourly wage for two-child mothers who had their second child illegally under the one-child policy is lower by 0.11 log points ($p<0.05$) compared to those who had their second child legally after the two-child policy was enacted. This finding remains robust when employing PSM with nearest neighbor matching. Additionally, our analysis does not reveal any significant taste-based discrimination against one-child mothers, as the wage comparison between these groups shows no significant difference, support our argument the wage penalty for one-child mothers is mainly driven by statistical discrimination.

Table 1. 7: The effect of FR Policy on the Anticipatory Motherhood Wage Penalty through tasted-based discrimination

	One-child mother		Two-child mother	
	(1)	(2)	(3)	(4)
one- vs. two-child policy	0.03	0.03	-0.11**	-0.13**
(illegal vs. legal)	(0.046)	(0.049)	(0.048)	(0.052)
N	1,458	1,309	953	866
R^2	0.3421	0.3423	0.2575	0.2760
Year & Prov & Ind FE	✓	✓	✓	✓
Person & Work & Fam	✓	✓	✓	✓
Child's age	✓	✓	✓	✓

NOTE.—Estimates from regressions on the log of the hourly wage for the working women. The term "*One- vs. two-child policy*" refers to the dummy variable whether the child born under the one-child policy or two-child policy period. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). Columns 2 and 4 include the PSM method with the nearest matching, given the person controls. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

C. Human capital

From the perspective of human capital theory (Becker, 1985; Mincer, 1989), the motherhood wage penalty could experience indirect shifts after relaxation of OCP. This can be attributed to the fact that some one-child mothers who desire to have a second child might reduce their working hours or educational level to prepare for the future child after the two-child policy. Such a reduction in work engagement and educational activities could slow down their rate of human capital accumulation, which could, in turn, lead to a decrease in wages (Goldin and Lawrence, 2008).

The hypothesis is that if the wage variations were genuinely rooted in shifts through human capital channel, then we would expect to observe significant changes in these two variables among one-child and two-child mothers after the policy. The CFPS employs a scale ranging from 1 to 8 to assess educational attainment. However, our analysis reveals that the bulk of the data is skewed, with a concentration in the 1-3 range;

only 0.75% of the respondents fall within the 7-8 range. To mitigate these limitations, we create a binary variable, classifying individuals into categories of either 'high' or 'low' educational attainment. We use high school as the cutoff, as it follows the 9-year free and compulsory education policy in China, which targets children up to age 15.[32,33] To isolate the causal impact of the policy on human capital variables, it is crucial to control for log monthly wage, as the change of wage could directly influence education level and working hours.[34]

The findings, as detailed in column 1 of Table 1.8, reveal that the two-child policy does not significantly affect the education levels of one-child and two-child mothers. Interestingly, after the policy, the working hours for one-child mothers significantly increase by 0.04 log points ($p<0.05$) in comparison to non-mothers, as shown in column 2. However, as evidenced by our previous findings in Table 1.2 and Table I.1.1 in Appendix I, both log hourly and monthly wages decrease for one-child mothers after the policy compared to the non-mothers. Logically, an increase in working hours post-policy would suggest a corresponding rise in monthly wages. However, the observed decrease in monthly wages suggests that one-child mothers may be increasing their working hours strategically, perhaps as a signal to employers that they do not intend to have additional children in the future. This could be an attempt to mitigate the negative wage effects they face due to employer perceptions of their potential fertility. In general, we can conclude that the change in the wage penalty of one-child and two-child mothers post-policy is unlikely to be primarily driven by human capital channels. Were the changes driven by human capital channels, we would expect to see a reduction in both educational levels and working hours.

[32] In Appendix J, we test the robustness of our findings by using different education levels as the cut-off. The results almost remain consistent with our initial observations.
[33] Our analysis reveals that approximately 38.2% of employed women in our sample fall into the 'high education level' category.
[34] Since the hourly wage is calculated by dividing the monthly wage by the number of working hours, including work hour variable as a control variable introduces an endogeneity issue due to its direct calculation from the dependent variable.

Table 1. 8: The effect of FR Policy on human capital variables

	Education	Work Hours
	(1)	(2)
One child × Policy	0.00	0.04**
	(0.018)	(0.018)
Two children × Policy	-0.01	-0.01
	(0.023)	(0.022)
N	10,399	10,399
R^2	0.4030	0.1228
Year & Prov & Ind FE	✓	✓
Person & Work & Fam	✓	✓

NOTE.—Estimates from regressions on working hours and education level by using the linear model. Working hours are represented using the logarithmic scale. Education level is divided into two categories: low and high education. In DiD framework, the control group is non-mothers. *Policy* represents the "two-child" policy after 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

1.7 Heterogeneity analysis

We have shown that the motherhood wage penalty for the one-child mothers increases, while the wage penalty for the two-child mothers is reduced after the two-child policy. An important question is whether these effects were heterogeneous across subpopulations. That is, the effects of the two-child policy on the wage penalty may differ by wage level (high vs. low-middle) and employment status (part-time vs. full-time).[35] Table 1.9 reveals that the effects of the two-child policy on the motherhood wage penalty do indeed differ by wage level and employment status. For mothers at high wage levels, the two-child policy does not lead to significant changes in the wage penalty for either one-child or two-child mothers, as shown by coefficients of 0.03 and

[35] Since CFPS does not directly collect information on part-time and full-time employment, we construct a dummy variable based on working hours, classifying individual as part-time if they work below 8 hours per day. Our findings indicate that roughly 26.3% of employed women with children are engaged in part-time jobs. This is consistent with the report on Statista "*Employment status of women who have children in China in 2018 and 2020*" indicating that around 24% of employed Chinese women with children were in part-time positions in both 2018 and 2020.

0.04 log points (p>0.1), respectively. Conversely, for mothers at low and middle wage levels, there is a significant increase in the wage penalty for one-child mothers, with a coefficient of -0.08 log points (p<0.01) post-policy, and a reduction in the wage penalty for two-child mothers, indicated by a coefficient of 0.05 log points (p<0.05) post-policy. These results corroborate our hypothesis that the post-policy variation in wage penalties is primarily through the discrimination channel, with workers at higher wage levels being less susceptible to discrimination in the labor market (Arulampalam et al., 2007).

Nonetheless, for one-child mothers in part-time employment, the implementation of the two-child policy does not result in a significant change in the wage penalty, as shown by a coefficient of 0.03 log points with a standard error of 0.071. In contrast, there is a significant increase in the wage penalty for one-child mothers in full-time employment, indicated by a coefficient of -0.10 log points (p<0.01) post-policy. For two-child mothers in part-time employment, the two-child policy significantly reduces the wage penalty, with a coefficient of 0.38 log points (p<0.01). While for those in full-time employment, it leads to a slight increase but not significant, evidenced by a coefficient of 0.05 log points (p>0.1). In general, this heterogeneity analysis suggests that taste-based discrimination is more prevalent among part-time two-child mothers, whereas statistical discrimination is more likely to occur in full-time one-child mothers post-policy.

Table 1. 9: The effect of FR policy on the Anticipatory Motherhood Wage Penalty (Heterogeneity analysis)

	Hourly Wage		Employment Status	
	Low & Mid	High	Part-time	Full-time
	(1)	(2)	(3)	(4)
One child × Policy	-0.08***	0.03	0.03	-0.10***
	(0.022)	(0.047)	(0.071)	(0.026)
Two children × Policy	0.05**	0.04	0.38***	0.05
	(0.026)	(0.081)	(0.098)	(0.031)
N	6,954	2,904	2,540	7,319
R^2	0.3665	0.0625	0.1664	0.4578
Year & Prov & Ind FE	✓	✓	✓	✓

Person & Work & Fam	✓	✓	✓	✓

NOTE.—Estimates from regressions on the log of the hourly wage for the working women. In DiD framework, the control group is non-mothers. *Policy* represents the "two-child" policy in 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in all columns. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

1.8 Conclusions

The focus of this paper is the exploration of the impact of fertility relaxation policy on the motherhood wage penalty. By drawing from China's policy transition from a one-child to a two-child norm, we find that after the fertility relaxation policy, the wage penalty for the one-child mothers increases significantly by 0.09 log points, while for the two-child mothers, it decreases significantly by 0.08 log points resulting. This change in the motherhood wage penalty post-policy is primarily driven by job discrimination rather than changes in human capital. Furthermore, our analysis delves into the mechanisms contributing to the post-policy rise in wage penalty for one-child mothers through statistical discrimination and the decrease in wage penalty for two-child mothers due to taste-based discrimination.

A direct theoretical implication of our study is the expansion of the motherhood wage penalty literature to include channels that have not been fully explored. Despite controlling for observable factors such as labor market experience and tenure, a substantial motherhood wage penalty persists, and while admittedly smaller, a wage gap remains. We find that the wage gap originates from anticipatory effects that occur before childbirth and is driven by statistical discrimination. A key policy insight from our study is that fertility relaxation policies can influence fertility rates in divergent ways. For one-child mothers with a strong desire to expand their families, such policies facilitate the possibility of having a second child. Conversely, these policies may increase the anticipatory motherhood wage penalty in the labor market via the statistical discrimination channel, creating potential feedback effects on fertility. This, in turn, could deter non-mothers from becoming mothers after the policy change.

There are some limitations of this paper. First, while the DiD approach is well-suited to evaluating policy effects, its validity relies on the parallel trends assumption, meaning that treated and control groups would have followed similar trends in the absence of treatment. This assumption may be challenged in the Chinese context, given that the policy was introduced amidst broader socioeconomic shifts, including changing gender norms, rising living costs, and labor market restructuring. These factors could confound the estimated treatment effects. Second, except for the job discrimination, there are some alternative mechanisms may also explain the observed patterns. For instance, the policy change may influence women's fertility planning and labor market strategies, prompting some to alter their career trajectories in anticipation of future childbearing. Third, our analysis is limited to data up to 2018. This cutoff was chosen to avoid confounding effects from the introduction of the COVID-19 in 2020 and three-child policy in 2021, thereby isolating the impact of the two-child policy.

In light of these limitations, future research could explore the following two avenues. First, using alternative identification strategies, such as natural experiments or randomized controlled trials, could help validate the findings and reduce concerns about confounding factors. Second, future studies could examine the role of evolving gender norms, fertility planning, and labor market restructuring in shaping women's career trajectories, offering deeper insights into the mechanisms behind the observed gender wage gap.

Overall, the findings contribute to the broader literature by demonstrating how fertility policy reforms can exacerbate the gender wage gap, particularly through the motherhood wage penalty, driven by both statistical discrimination and taste-based discrimination.

1.9 References

Agarwal, Sumit & Li, Keyang and Qin, Yu and Wu, Jing (2024), The Impact of Fertility Relaxation on the Gender Wage Gap.

Anderson, D. J., Binder, M., & Krause, K. (2002). The Motherhood Wage Penalty: Which Mothers Pay It and Why? *The American Economic Review*, *92*(2), 354–358.

Arrow, Kenneth J. (1972). Models of Job Discrimination. In *Racial Discrimination in Economic Life*, edited by Anthony H. Pascal, 83–102. Lexington, MA: D.C. Heath

Arulampalam, W., Booth, A. L., & Bryan, M. L. (2007). Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *ILR Review*, 60(2), 163-186.

Becker, G.S., (1985). Human capital, effort, and the sexual division of labor. *Journal of labor economics*, *3*(1, Part 2), pp. S33-S58.

Becker, Gary S. (1957). *The Economics of Discrimination*. 2nd ed. Chicago: Chicago University Press.

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American sociological review*, 204-225.

Budig, M. J., Misra, J., & Boeckmann, I. (2012). The motherhood penalty in cross-national perspective: The importance of work–family policies and cultural attitudes. *Social Politics*, *19*(2), 163-193.

Catalina Amuedo-Dorantes & Jean Kimmel (2005). The Motherhood Wage Gap for Women in the United States: The Importance of College and Fertility Delay. *Review of Economics of the Household*

Charles, Kerwin Kofi, & Jonathan Guryan (2008). Prejudice and The Economics of Discrimination. *Journal of Political Economy*, 116(5), 773–809.

Coate, Stephen, & Glenn C. Loury. (1993) Will Affirmative–Action Policies Eliminate Negative Stereo-types? *American Economic Review*, 83(5), 1220–40.

Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty?. *American journal of sociology*, *112*(5), 1297-1338.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, & Kari Vea Salvanes (2016). What Is the Case for Paid Maternity Leave? *Review of Economics and Statistics*, 98(4), 655-670.

Duvivier, C., & Narcy, M. (2015). The motherhood wage penalty and its determinants: A public–private comparison. *Labour*, *29*(4), 415-443.

Goldin, Claudia, & Lawrence F. Katz. (2008). Transitions: Career and Family Life Cycles of the Educational Elite. *American Economic Review*, 98 (2): 363-69**.**

Gough, M., & Noonan, M. (2013). A review of the motherhood wage penalty in the United States. *Sociology Compass*, *7*(4), 328-342.

Havnes, Tarjei, & Magne Mogstad (2011). Money for Nothing? Universal Child Care and Maternal Employment. *Journal of Public Economics*, 95(11-12), 1455-1465

He, H., Li, S. X., & Han, Y. (2023). Labor Market Discrimination against Family Responsibilities: A Correspondence Study with Policy Change in China. *Journal of Labor Economics*, *41*(2), 361–387.

Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica*, *42*(4).

Heckman, J. J., & Macurdy, T. E. (1980). A Life Cycle Model of Female Labour Supply. *The Review of Economic Studies*, *47*(1), 47–74.

Heckman, J. J., Hidehiko, I., & Petra, E. T. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies, 64*(4), 605–654.

Hill, M. S. (1979). The wage effects of marital status and children. *Journal of Human Resources*, 579-594.

Jessen, J., Jessen, R., & Kluve, J. (2019). Punishing potential mothers? Evidence for statistical employer discrimination from a natural experiment. *Labour Economics*, *59*, 164–172.

Jia, N. and Dong, X.Y., (2013). Economic transition and the motherhood wage penalty in urban China: Investigation using panel data. *Cambridge Journal of Economics*, 37(4), pp.819-843.

Jia, N., Zhou, Y., & Yang, T. (2021). “Selective two-child” policy and household resource allocation. *China Economic Review*, 101639.

Killewald, A., & Bearak, J. (2014). Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression. *American Sociological Review*, *79*(2), 350–357.

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

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, & Josef Zweimüller (2020). Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation. *American Economic Journal: Economic Policy* (Forthcoming).

Krashinsky, H. A. (2004). Do Marital Status and Computer Usage Really Change the Wage Structure? *Journal of Human Resources*, *XXXIX*(3), 774–791.

Lalive, R., & Zweimüller, J. (2009). How does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments *. *Quarterly Journal of Economics*, *124*(3), 1363–1402.

Lang, Kevin, and Jee-Yeon K. Lehmann (2012). Racial Discrimination in the Labor Market: Theory and Empirics. *Journal of Economic Literature*. 50(4), 959–1006.

Li S & Zhang Y, Feldman MW. (2010). Birth Registration in China: Practices, Problems and Policies. *Popul Res Policy Rev*, *29*(3), 297-317.

Lundborg, Petter, Erik Plug, & Astrid Würtz Rasmussen. (2017). "Can Women Have Children and a Career? IV Evidence from IVF Treatments." *American Economic Review*, 107 (6), 1611-37**.**

Meng, X., & Kidd, M. P. (1997). Labor Market Reform and the Changing Structure of

Wage Determination in China’s State Sector during the 1980s. *Journal of Comparative Economics*, *25*(3), 403–421.

Mincer, J. (1989). Human capital and the labor market: A review of current research. *Educational researcher*, *18*(4), 27-34.

Phelps, Edmund S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review* 62(4), 659–61.

Ruhm, Christopher J. (2004). “Parental Employment and Child Cognitive Development.” *Journal of Human Resources*, 39(1): 155-192.

Schoen, R., Astone, N. M., Kim, Y. J., Nathanson, C. A., & Fields, J. M. (1999). Do Fertility Intentions Affect Fertility Behavior? *Journal of Marriage and Family*, *61*(3), 790–799.

Waldfogel, J., (1997). The effect of children on women's wages. *American sociological review*, pp.209-217.

Xie, Yu, Zhang, Xiaobo, Tu, Ping, and Ren, Qiang, (2020). China Family Panel Studies (CFPS).

Yu, J. & Xie, Y., (2014). The Effect of Fertility on Women's Wages in China. *Population Research*, *38*(1), pp.18-29.

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

1.10 Appendix

Appendix A: Evolution of China's family planning policy

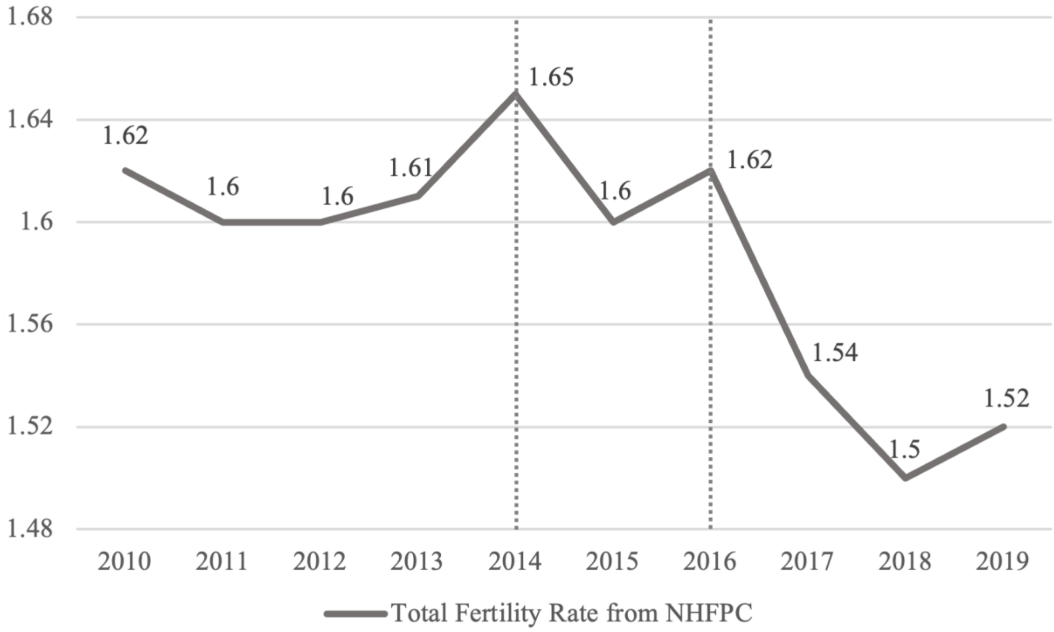
Table A.1. 1: Evolution of China's family planning policy

Time	Issue Unit	Document Name	Content
1980	Party Central Committee	*"An Open Letter to all Communist And CYL Members on The Issue of Controlling China's Population Growth"*	Couples are encouraged to have only one child
1982	The 12th National People's Congress (NPC)	*"Constitution of People's Republic of China(1982)"*	Family planning was written into the Constitution
1991	Central Committee of the Communist Party of China、State Council	*"Decision on Strengthening Family Planning Work and Strictly Controlling Population Growth"*	Implement the current family planning policy and strictly control population growth
2013	The 18th Central Committee of the Communist Party of China	*"Decision of the CPC Central Committee on Major Issues concerning Comprehensively Deepening Reform"*	Couples can have two children if one of them is an only child
2015	The Standing Committee of National People's Congress	*"Amendments to the Population and Family Planning Law"*	Implement the universal two-child policy

CYL: Communist Youth League

Appendix B: Total fertility rate over time

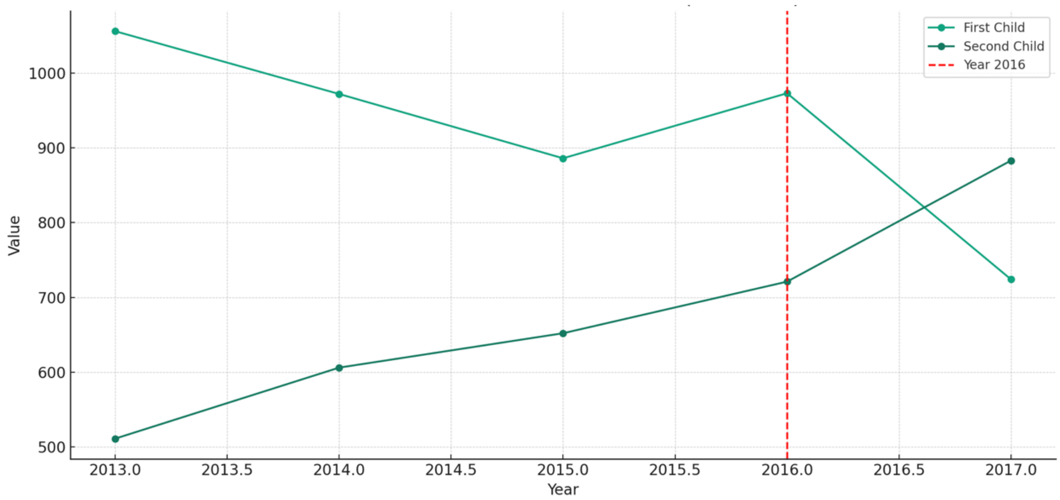
Figure B.1. 1: Total fertility rate over time



Source: National Health and Family Planning Commission of the People's Republic of China (NHFPC)

Appendix C: Trends in first and second births over time

Figure C.1. 1: Trends in first and second births over time (2013-2017)

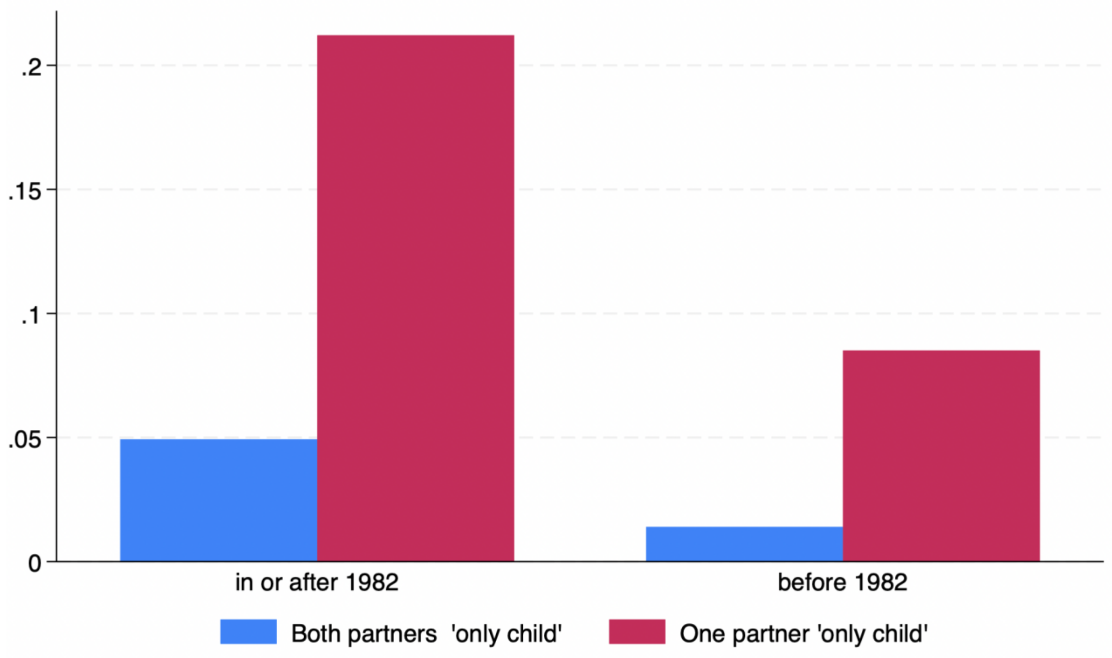


Note: The unit is ten thousand births.

Source: China Statistical Yearbook, 2020

Appendix D: Proportion of Partners Who Are ‘Only Children’ in 2010

Figure D.1. 1: Proportion of Partners Who Are ‘Only Children’ in 2010



Appendix E: The effect of fertility relaxation policy on fertility behavior

Our primary focus is understanding how fertility relaxation policy impacts fertility behavior. As discussed in section 1.2, the policy might spur high-order births while decreasing low-order births. To empirically test this in our dataset, we use sequential Logistic regression to assess the likelihood of women having one or two children under "two-child" policies:

$$Logit\left(\frac{P(N_{it}=j|N_{it-1}\geq j-1)}{P(N_{it-1}=j-1)}\right)=\gamma_j P_t+X_{it}\delta+\varepsilon_{it} \tag{E.1}$$

where N_{it} represents the number of children for woman i at time t (with $j\in\{1,2\}$).

Table E.1.1 illustrates the effect of "two-child" polices on the fertility behavior of women, focusing on the change of proportion of one-child and two-child mothers after policy. Our initial step involves estimating a regression that includes a range of individual characteristics. As evident in column 1, the "two-child" policy appear to influence these proportion, with an observable decrease of about 3 percentage points ($p<0.01$) in the likelihood of non-mother having their first child. Concurrently, there is an increase of about 9 percentage points ($p<0.01$) in the likelihood of one-child mothers deciding to have a second, as shown in column 2. In general, above results suggest that the "two-child" policy appeared to influence the fertility decisions of women in two significant ways: fewer non-mothers transitioned into one-child mothers, and a greater number of one-child mothers transitioned into two-child mothers. This is consistent with the data from the China Statistical Yearbook (2020), which documented a decline in first births and an increase in second births (see Figure 1.1).

Table E.1. 1: The effect of fertility relaxation on the fertility decision

	Zero → One	One → Two
	(1)	(2)
Policy	-0.03***	0.09***
	(0.004)	(0.007)
N	17,361	
Individual controls	✓	✓
Family controls	✓	✓

NOTE.—Estimates are from sequential logistic regressions on probability of women having 0, 1 and 2 children, with coefficients reported in the marginal level. We use the code "seqlogit" in Stata. The complete set of controls described in equation (A.1) is included but not reported. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

Appendix F: The effect of fertility relaxation policy on the anticipatory motherhood wage penalty for non-mothers

Table F.1. 1: The effect of FR policy on the anticipatory motherhood wage penalty for non-mothers

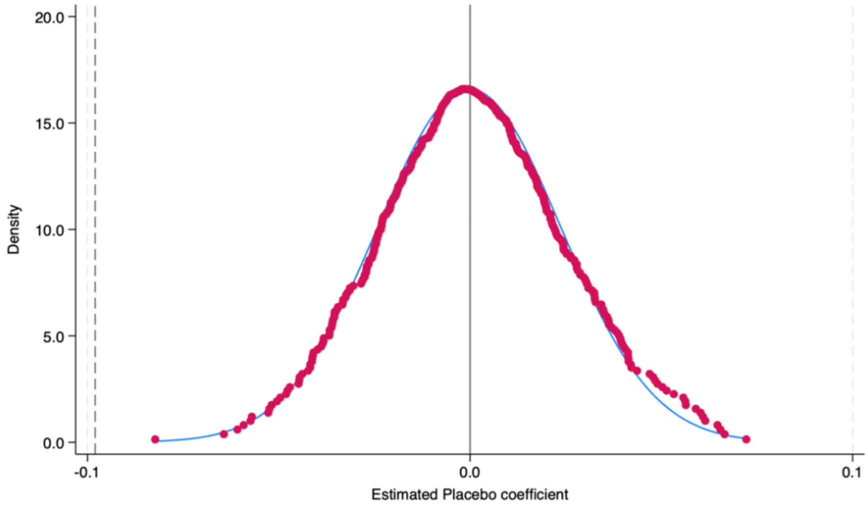
Non-mothers	Fixed	Industry	PSM
	(1)	(2)	(3)
Tr × S	0.035	0.035	0.027
	(0.171)	(0.173)	(0.173)
N	1,234	1,158	1,158
R^2	0.3144	0.3377	0.3134
Prov &Year & Age FE	✓	✓	✓
Person & Work & Fam	✓	✓	✓
Industry FE		✓	✓
Matched			94%

NOTE.—Estimates from regressions on the log of the monthly wage for the working women. In DiD framework, the control group is non-only-child non-mothers. *S* represents the selective two-child policy in 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in all columns. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

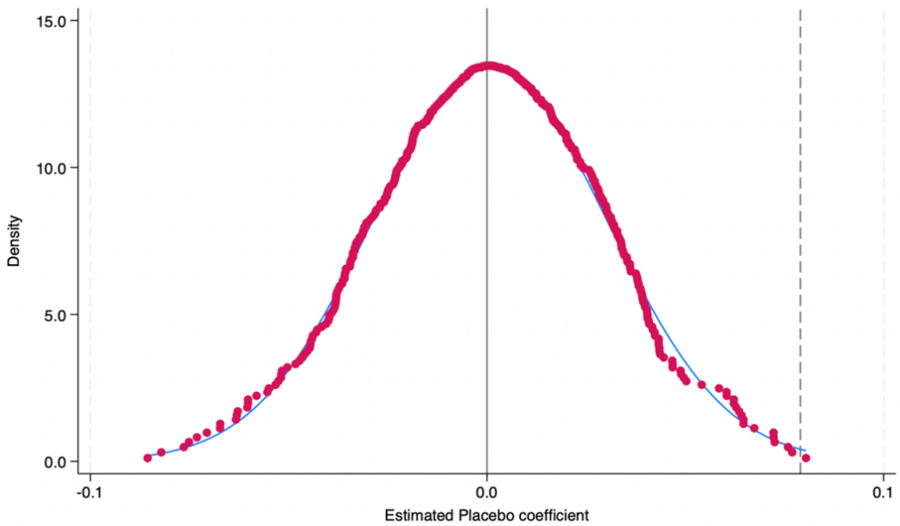
Appendix G: Permutation Test

Figure G.1. 1: Permutation Test

(a)



(b)



Note: These figures present the permutation test by randomly assigning the one-child/two-child mothers and non-mothers and repeating the main analysis 500 times in panel (a)/(b). The sample period is from 2010 to 2018. The outcome is the log of hourly wage. The complete set of controls described in equation (1). This graph plots the distribution of the placebo DiD estimates. The vertical solid line is the true DiD estimate in column 3 of Table 1.2.

Appendix H: Summary Statistics by Sector

Table H.1. 1: Summary Statistics by Sector

	All		Private		Public		
Variables	Mean	SD	Mean	SD	Mean	SD	Diff
Dependent							
Monthly Wage	2393	2170	2319	2112	2598	2310	-279***
Independent							
Num. child	.855	.704	.874	.735	.801	.609	0.073***
One child	.483	.5	.445	.497	.589	.492	-0.145***
Two children	.186	.389	.215	.411	.106	.308	0.109***
Personal Chara.							
Age (year)	33	8.8	33	8.8	35	8.4	-1.817***
Education (1-9)	2.048	.425	1.978	.372	2.243	.494	-0.265***
Urban (Rural)	.686	.464	.641	.48	.811	.391	-0.170***
Working hours	219	71	230	72	187	55	42.855***
Public (private)	.266	.442	-	-	-	-	-
Firm size	541	4409	550	5000	514	2021	36.66
Family Chara.							
Savings (k)	7.199	4.63	7.065	4.593	7.568	4.712	-0.503***
Housework hours	59	28	59	28	58	28	1.004
Parents at home	.264	.441	.263	.44	.267	.442	-0.00400
Parents education	2.491	1.105	2.391	1.041	2.767	1.225	-0.376***

NOTE.—Data is sourced from the CFPS spanning 2010 to 2018. 'Wage' denotes monthly income from the primary job. 'Promotion satisfaction' ranges from 0-5. 'Working hours' and 'Housework hours' are represented monthly. 'Parents at home' signifies the presence of the respondent's both parents in their household. 'Father/mother education' refers to the educational attainment of the respondent's parents. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Appendix I: The effect of FR policy on the motherhood wage penalty (monthly wage)

Table I.1. 1: The effect of FR policy on the anticipatory motherhood wage penalty (monthly wage)

Log Monthly Wage	Person	W & F	Fixed	PSM	PSU Hourly wage
	(1)	(2)	(3)	(4)	(5)
			Main effect		
One	-0.03	-0.05**	-0.04*	-0.04*	-0.02
	(0.022)	(0.022)	(0.021)	(0.021)	(0.032)
Two	-0.18***	-0.19***	-0.14***	-0.14***	-0.18***
	(0.028)	(0.028)	(0.027)	(0.027)	(0.040)
			Policy effect		
One × S	-0.03	-0.05**	-0.05**	-0.04*	-0.09***
	(0.023)	(0.023)	(0.021)	(0.022)	(0.030)
Two × U	0.11***	0.09***	0.08***	0.09***	0.08**
	(0.029)	(0.028)	(0.027)	(0.027)	(0.040)
N	9,860	9,860	9,860	9,731	9862
Year FE	✓	✓	✓	✓	
Age & Edu control	-	✓	✓	✓	
Work & Family control	-	-	✓	✓	
Province & Industry FE	✓	✓	✓	✓	
Matched	-	-	-	98%-	

NOTE.—Estimates from regressions on the log of the monthly wage for the working women (columns 1-4). Column 5 estimate from regression on the log of hourly wage for the working women clustering at primary sampling unit (PSU) level. In DiD framework, the control group is non-mothers. *S* represents the "selective two-child" policy in 2014, and *U* represents the "universal two-child" policy in 2016. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in all columns. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Appendix J: The effect of FR policy on the different educational level categories

Table J.1. 1: The effect of FR policy on the different category educational level categories

	Junior	Undergraduate
	(1)	(2)
One × Policy	0.03**	0.01*
	(0.013)	(0.004)
Two × Policy	-0.01	-0.00
	(0.016)	(0.005)
N	10,399	10,399
R^2	0.2983	0.0463
Year FE	✓	✓
Age & Edu & Urban	✓	✓
Work & Family	✓	✓
Province & Industry FE	✓	✓

NOTE.—Estimates from regressions on the education level for the working women. In DiD framework, the control group is non-mothers. *Policy* represents the "two-child" policy after 2014. The complete set of controls includes personal characteristic (e.g., age, age square and education), family characteristics (e.g., household working hours, family saving, parents' education and parents at home), and working background (e.g., type of sector, and firm's size). We exclude the women opt to additional child after the policy in all columns. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

Chapter 2

Gender Disparities in Flexible Working Arrangements and Wage Inequality

Abstract

Flexible working arrangements (FWAs) have been proposed to narrow the gender wage gap, but their effectiveness remains debated. The gender asymmetry and gender structure frameworks may shape the gendered selection of FWAs and their outcomes, but empirical evidence remains limited. Using UK longitudinal survey data and a shift-share instrumental variable method, this study reveals significant gender disparities in FWA transitions and their consequences for the wage gap. Men are more likely to transition from traditional work to flexi-time work, while women are more likely to switch to term-time work. These two FWA regimes have very different wage consequences. The wage penalty for women transitioning to term-time work is more than six times greater than that for men transitioning to flexi-time work. Additionally, occupational segregation further deepens this gap, as men are more likely to move into higher-skill flexi-time roles, while women transition to lower-skill term-time roles.

2.1 Introduction

The UK government has promoted Flexible Working Arrangements (FWAs) as a social intervention to facilitate women labor participation and reduce the gender wage gap. By strengthening workers' rights to request flexible schedules, FWA policies may support women in maintaining full-time employment, enhancing their earning potential, and advancing their careers at a pace comparable to that of men (Chung, 2018; Chung and Van der Horst, 2018; Fuller and Hirsh, 2019; Piasna and Plagnol, 2017; Van der Lippe et al., 2019). While FWAs are thought to provide avenues for greater gender equality in the workforce, research indicates that their impact may be double-edged, potentially exacerbating the gender wage gap in some instances. For instance, Lott and Chung (2016) observe that men often utilize FWAs to enhance job performance and earn income premiums, whereas women in similar roles frequently face increased family responsibilities (Hilbrecht et al., 2008; Chung, 2022). Additionally, some studies argue that women are more likely to experience stigma when using FWAs for the purpose of taking care of family. This stigma is commonly referred to as 'flexibility stigma' or 'femininity stigma' (Chuang, 2022; Williams et al., 2013; Cech and BlairLoy, 2014). Such gendered motivations of FWAs contribute to perpetuate rather than alleviate gender wage gap (Jennifer, 2004; Jennifer and Mary, 2016). However, to the best of our knowledge, research on gender disparities in FWAs and their impact on the gender wage gap remains limited.

According to Chandola et al. (2019), FWAs include three major types: (1) variable hours (e.g., flexi-time work, compressed working week and annualized hours), (2) reduced hours (e.g., part-time, job-sharing and term-time work) and (3) flexplace (e.g., home working).[36] Although part-time work is classified as reduced-hours FWAs by the government, previous studies often distinguish it from the other FWA types when

[36] Flexplace is outside the scope of our study because it is not a time-flexible arrangement.

analyzing its impacts on the gender wage gap (Chung, 2020). This distinction arises because part-time work often overlaps with other FWA types, such as part-time flexi-time or part-time term-time arrangements, complicating classification and hindering the identification of causal effects.[37,38] Therefore, this paper focuses on gender differences in transitions from traditional full-time work to specific types of full-time FWAs and their contribution to gender wage gaps (hereafter, 'full-time' is implied and not explicitly stated). Unlike previous studies that have often treated FWAs as a homogeneous category (e.g., Alsulami et al., 2022; Chung, 2020; Lu et al., 2023), this paper highlights the heterogeneity between 'reduced-hours' and 'variable-hours' arrangements, which has been largely overlooked. However, as Chandola et al. (2019) note, FWAs differ significantly depending on the nature of their time arrangements, which is critical in uncovering a pronounced gender disparity in FWAs adoption. For example, gender differences in FWAs adoption may be influenced by individual constraints, such as the conflict between work performance and family responsibilities. These constraints often interact with traditional gender norms, which may drive women to choose reduced-hours arrangements to accommodate family responsibilities, while men may opt for variable-hours arrangements to enhance job performance.

While previous research has explored men's and women's motivations for, and consequences of, transitioning to FWAs, much of this work has centered on the division of work-life conflicts. For instance, the "boundary theory" (Nippert-Eng, 1996; Ashforth et al., 2000) has been used to explain gendered motivations for FWAs, while the "flexibility paradox" (Chung, 2022; Lu et al., 2023) and "flexibility stigma" (Williams et al., 2013; Cech and BlairLoy, 2014) have been highlighted to account for the gendered consequences of these arrangements. Despite its importance to wages, FWAs have attracted relatively little attention as an explanation for gender wage gaps.

---

[37] According to the UKHLS, part-time FWAs account for 5.1% of the full sample, while full-time FWAs make up approximately 18.4%.

[38] Additionally, part-time work is associated with significant wage penalties for both men and women (Kalleberg, 2000), potentially overshadowing the effects of other FWA components (e.g., term-time work). Further discussion of these part-time compositions is provided in Appendix F.

Thus, this study extends previous research by examining the role of FWAs in perpetuating gender inequality in the labor market in several ways. First, recent studies on the wage effects of FWAs have largely treated them as a homogeneous category (e.g., Alsulami et al., 2022; Chung, 2020; Lu et al., 2023). In this study, we examine gender disparities across different types of FWAs and their impact on the gender wage gap. As explained below, incorporating gender disparities in FWAs is essential for understanding variations in gendered labor market outcomes, thereby offering deeper insights into the gender structure theory (Risman, 2004) and the gender asymmetry theory (Shauman, 2010). Second, our empirical analysis contributes to the literature by quantifying the working time patterns associated with transitions to various FWAs (e.g., Allen et al., 2013; Chung and van der horst, 2018; Chandola et al., 2019). Our findings underscore the significant differences in working time patterns among the different FWAs, such as regular working, overtime working, weekend working, and paid holiday days. For example, variable-hour arrangements (e.g., flexi-time work) may often involve the rearrangement of working time patterns, while reduced-hours arrangements (e.g., term-time work) may offer differing numbers of paid holiday days, distinguishing them apart from other types of FWAs. Without accurately observing the nuanced differences in working time patterns across FWAs, it is challenging to fully understand their implications for labor market outcomes and gendered inequality. Lastly, while most previous studies have focused on the association between FWAs and labor market outcomes, this paper aims to establish causal effects. By utilizing FWAs adoption and supply data from the UK Household Longitudinal Study (UKHLS) and the larger sample of the Labor Force Survey (LFS), this paper applies the shift-share instrumental variable (SSIV) method to estimate the causal impact of FWAs on labor market outcomes.

Generally, this study shows that gender disparities in transitions to FWAs in the UK have persisted over the past decade. Full-time employed men are more likely to transition to flexi-time work (FTW), a type of variable-hour arrangement, while women primarily shift to term-time work (TTW), a type of reduced-hours arrangement. Transitioning from traditional work to TTW helps women better align their work-life balance by reducing total working hours, allowing for more time spent on household duties. Additionally, compared to women who transition to traditional part-time work (PTW), transitioning to TTW incurs a smaller wage penalty and offers better promotion prospects. Overall, TTW enables women to balance family obligations without significantly sacrificing their wage outcomes and career progression. In contrast, transitions to FTW, primarily by men, involve rearranging their working time patterns without significant reductions in total working hours, allowing them to maintain both their monthly wages and promotions opportunities. Notably, these gender-disparate transition patterns further exacerbate the gender wage gap, resulting in a directly a 5.1% gap between women transitioning to TTW and men transitioning to FTW, which increases to 12.6% when accounting for the mediating effect of total working hours.[39] Additionally, this gap is intensified by occupational segregation, as men are more likely to transition to high-middle skill FTW, while women tend to move into low skill TTW.

2.2. Theoretical background

Current research on FWAs and the gender wage gap has given limited attention to the trends in male and female workers transitioning to different FWAs, particularly FTW and TTW. Amid ongoing debates about whether FWAs mitigate or exacerbate gender wage gaps, it is crucial to gain clearer insights into the mechanisms driving gendered differences in motivations for these transitions and their associated wage outcomes.

[39] The wage penalty for women transitioning to TTW is around six times higher than the penalty for men transitioning to FTW.

A. The gendered restrictions and asymmetry

According to the gender structure theory (Risman, 2004) and the gender asymmetry thesis (Shauman, 2010), the motivations of behind male and female workers' transitions to different FWAs can be highly restricted by gender structures from a macro to micro levels. For instance, social policies that provide significantly longer maternity or parental leave to women than to men can unintentionally reinforce the expectation that caregiving is primarily a woman's responsibility (Tharp and Parks-Stamm, 2021).[40] In terms of restrictions at the meso-level, women are more vulnerable to job discrimination and flexibility stigma in the workplace than men (Rudman and Mescher, 2013; Williams et al., 2013). In this context, Chung (2018) suggested that women's transition to FWAs may amplify the stigma effects. In terms of the restrictions at the micro-level, internalized gender norms further contribute to a gender asymmetry in joint decision-making within the households (Shauman, 2010; Grinza et al., 2022). The concept of 'doing gender' (West and Zimmerman, 1987) provides a framework for understanding individuals' active expression of their gender identity, with men do more paid work to fulfill the role of the 'breadwinner' and women taking on more household responsibilities fulfill the role of the 'homemaker'.

As a result, the gender division of labor persists across societies, with women continuing to shoulder the majority of unpaid work (Henz, 2006), a situation that reinforces the "double jeopardy" (Bratberg et al., 2002). Consequently, even with the availability of FWAs, women are more likely to opt for reduced-hour arrangements to better manage family responsibilities. In contrast, men are more inclined to choose variable-hour arrangements, which allow them to improve job performance and advance their career. In the following section, we further discuss the specific types of FWAs within reduced-hour and variable-hour arrangements, examining their respective

[40] Women are entitled to up to 52 weeks of Statutory Maternity Leave, with up to 39 weeks paid. In contrast, men are eligible for up to 2 weeks of paid Paternity Leave. To address this imbalance, the UK introduced Shared Parental Leave (SPL) in 2015, allowing parents to share up to 50 weeks of leave and 37 weeks of pay. However, uptake of SPL has remained low, with only 1% of eligible mothers and 5% of fathers participating, indicating the persistence of traditional gender norms in caregiving.

advantages and disadvantages.

B. Gender differences in selection of FWAs

Numerous studies over the past decades have shown that women disproportionately engage in reduced-hours arrangements, particularly part-time work (PTW), which is often associated with lower pay (Blau and Kahn, 2017) and limited promotion prospects (Manning and Petrongolo, 2008). However, previous research often overlooks term-time work (TTW) as a distinct category within reduced-hours arrangements.[41,42] Unlike PTW, which typically involves a consistent reduction in hours throughout the year, TTW allows employees to work full-time hours during school terms, potentially reducing the wage penalty caused by reduced working hours and offering better career development opportunities (Gregory and Milner, 2009). Compared to other full-time working arrangements, TTW provides greater flexibility, particularly for parents of school-aged children, by aligning work schedules with the school calendar and helping them manage caregiving duties more effectively. This argument is supported by Gregory and Milner (2009), who discussed that women with dependent children are most likely to take up school term-time working when it is available. Data from the Labor Force Survey (LFS) further corroborates this trend, showing that TTW was significantly more prevalent among full-time women than men between 2000 and 2020, with 9% of women utilizing this arrangement compared to just 1% of men. Conversely, more than 10% of full-time employed men are in flexi-time work (FTW). Similar patterns are observed in the UK Household Longitudinal Study (UKHLS), with further details provided in Appendix A.

Unlike reduced-hours arrangements, which are often adopted for family-related reasons (Connolly and Gregory, 2009), variable-hour arrangements are more likely to

[41] According to UKHLS definition, TTW offers employees the flexibility to work full-time during school terms while reducing their hours outside the academic calendar.

[42] Based on the characteristics of TTW, we define it as full-time employment, as after dropping the part-time TTW respondents, the average weekly working hours are approximately 36.4, which exceeds both the government's threshold of 35 hours and the UKHLS definition of 30 hours.

enhance job performance by enabling workers to tailor their schedules to their personal peak productivity times (Kelly et al., 2014). This structure tends to benefit men more, as gender norms and their motivations for using FWAs often emphasize higher earnings and career advancement (Chung, 2018). Much of the previous research on variable-hour arrangements has focused on flexi-time work (FTW), given their widespread adoption, although access to such flexibility varies by class position.[43] For instance, lower-skilled men often use FTW to extend working hours, driven by financial insecurity and the need to maximize income (Lu et al., 2023; Kossek and Lautsch, 2018). On the other hand, middle- and high-skilled men commonly work longer hours through FTW, reflecting the influence of "ideal worker" norms (Williams et al., 2013; Cech and Blair-Loy, 2014). In general, these workers are more likely to leverage flexibility to work longer, as long hours have traditionally been associated with job dedication and stability (Chatzitheochari and Arber, 2009). In contrast, women frequently face constraints in increasing their working hours due to caregiving responsibilities and societal expectations, which limit their ability to fully utilize FTW for career advancement (Cha, 2010).

Overall, according to the gender structure theory and the gender asymmetry theory, women are more likely to choose work arrangements that ensure enough time and energy for facilitating family demands, while men are expected to leverage flexibility to align with the ideal worker or breadwinner model. Within this framework, we hypothesize that "*women are more likely to transition to TTW over time, as TTW may mitigate some of the disadvantages commonly associated with reduced-hours arrangements*". In contrast, "*men are expected to transition to FTW, as FTW enables them to reorganize their working time patterns to enhance job performance*".

[43] In 2014, the UK introduced a statutory right allowing all employees to request flexible working, including flexi-time, after 26 weeks of continuous service.

C. Gender selection and Gender wage gap

The gender wage gap remains persistent and widespread in labor markets. Explanations put forward include discrimination (Lips, 2013), human capital accumulation (Gayle and Golan, 2012), competitiveness (Dohmen and Falk, 2011), occupational segregation (Busch, 2020), labor force participation rates (Blau and Kahn, 2017), and work-life balance priorities (Goldin, 2014). Although gender inequality/equality in FWAs has frequently been cited as potential factors contributing to labor market outcomes (Cha and Weeden, 2014; Lott and Klenner, 2018; Chung et al., 2021; Leuze and Strauß, 2016), existing evidence remains insufficient to provide a clear picture, largely due to the lack of in-depth investigation into the consequences of FWAs.

FWAs have been proposed as a potential work-family solution to reduce the gender wage gap (Chung, 2018). However, persistent gender structure and gender asymmetry may influence the selection of FWAs differently for men and women (Lott and Chung 2016; Sullivan and Lewis, 2001). As previously discussed, factors from the macro to micro level collectively constrain women, often restricting them to prioritize family demands and work-life balance over income and career prospects (Gregory and Milner, 2009). These restrictions often lead women to choose to reduce working hours and use flexibility to accommodate family demands. Although the TTW within reduced-hour arrangements may mitigate some gender inequalities, this more “family-oriented” approach still creates rigid barriers for women in advancing up the career ladder, resulting in lower long-term payoffs (Blau and Kahn, 2017). Gendered occupational segregation may further widen the gender wage gap. Research consistently shows that women are disproportionately represented in lower-skilled jobs, which typically offer lower pay and fewer opportunities for advancement (Busch, 2020). This pattern is particularly evident in TTW, where women may experience career downgrading or stagnation when transitioning to TTW to accommodate caregiving responsibilities.

In contrast, gender restrictions on men prompt them to use FTW to work longer hours, focusing on building stronger career and income achievements (Williams et al., 2013; Cech and Blair-Loy, 2014). Previous studies have highlighted notable gender disparities in paid job hours, showing that men are more likely to work full-time and engage in overtime more frequently (Presser, 2003; Jacobs and Gerson, 2004). Such 'preference' for long working hours can be attributed to both income effect and the persistent "ideal worker" norms as discussed above (Angrave and Charlwood, 2015; Chatzitheochari and Arber, 2009). The longer working hours and overtime observed among men with time flexibility is summarised as a form of 'flexibility paradox' issue, which emphasized in recent time use studies (Chung and van der Horst, 2018; Lu et al., 2023). Beyond labor outcomes, according to the boundary theory (Ashforth et al., 2000) and flexibility enactment theory (Kossek et al., 2005), FTW can help to integrate work and home roles. Specifically, FTW allows employees to adjust their working hours around personal commitments without reducing overall working hours, thereby reducing conflicts between work and family responsibilities (Chung and van der Lippe, 2020). However, Allen et al. (2013) found different results, showing that while FTW plays a crucial role in enhancing job and promotion satisfaction, but its effect on reducing family-to-work interference is limited. Similarly, Wang and Chen (2024) found that FTW does not necessarily lead to a more equitable division of domestic responsibilities, as women continue to bear the majority of housework and childcare duties. Indeed, if the primary purpose of FTW is to rearrange working hours and even extend them to increase wage, workers may still face challenges in managing family obligations. This raises another important research question: "*Could gender disparities in the use of FWAs further exacerbate gender wage inequality in the labor market?*".

2.3 Data

This paper uses three datasets. The first is the UK Household Longitudinal Study (UKHLS, 2023), a panel survey focused on household dynamics, economic well-being, labor market changes, and family life. Wave 1 collected data from around 80,000 individuals in 40,000 households. We draw on data from waves 1-12, spanning the years 2009/2010 to 2020/2021. Among these waves, only waves 2, 4, 6, 8, 10, and 12 contain information about FWAs. We construct our sample by including both men and women aged 16 to 60, who are neither not students nor self-employed. Next, we restrict the sample to employed women and men, excluding employed individuals earning below the minimum threshold stipulated by the UK's Minimum Labor Wage Act (about 3.7%).[44,45] We also discard individuals with multiple FWAs to avoid composition effects, ensuring a pure interpretation of each FWA influence (about 4.6%). The final sample consists of 27,778 observations, representing 9,622 employed individuals, of whom 44% are men and 56% are women. More details are available in Table 2.1.[46]

We supplement the UK Household Longitudinal Study (UKHLS) with information from the Labor Force Survey (LFS) and the Time Use Survey (TUS). The LFS is a quarterly survey conducted by the Office for National Statistics (ONS), offering extensive data on the labor market status of individuals in the UK, with data spanning from 2000 to 2020. It provides detailed insights into employment types, hours worked, industry. Importantly, it also includes information on the number of paid holiday days, which is missing from UKHLS. However, the LFS has limited longitudinal depth, as each individual is interviewed for only five consecutive quarters before being rotating out of the survey, making it difficult to tracking long-term individuals' transitions. On the other hand, we use data from the 2015–2016 wave of the UK TUS, a cross-sectional dataset that includes detailed information on FWAs. The TUS, conducted periodically, provides details insight into how individuals in the UK allocate their time among various activities, including work, leisure, and domestic responsibilities. This data is

[44] In the UK, the Minimum Labour Wage were set at £5.93, £6.08, £6.19, £6.31, £6.50, and £6.70 for the years 2010, 2012, 2014, 2016, 2018, and 2020 respectively.

[45] Individuals on maternity leave were excluded from the analysis, as they represent only 0.84% of the sample. Moreover, including them could introduce a compound effect. Specifically, women who transition to FWAs may do so in anticipation of having another child and taking maternity leave. Including these cases could therefore conflate the effects of FWAs with fertility-related decisions, making it difficult to isolate the causal impact of FWAs on labour market outcomes.

[46] Since we use individual fixed effects to explore workers' transitions, we exclude individuals who participant in only a single wave of data.

invaluable for analyzing actual working hours outside of standard schedules, such as weekend work, as it offers a detailed perspective on work patterns.

Table 2. 1: Summary Statistics

	All		Men		Women		
2010-2020	Mean	SD	Mean	SD	Mean	SD	Diff.
Flexible							
Traditional	.69	.46	.76	.43	.63	.48	0.1***
Term-time	.11	.32	.041	.20	.17	.38	-0.13***
Flexi-time	.198	.399	.2	.4	.197	.398	0.003
Labor market							
Monthly wage	2,325	1,423	2,742	1,587	1,989	1,172	753***
Working hour	151	32	161	25.3	143.3	33.9	17.***
Overtime hour	18.1	30.6	19.8	31.4	16.7	29.8	3.1***
Firm size	5.6	2.3	5.7	2.3	5.4	2.3	0.4***
Person Chara.							
Age	40.2	11.0	40.0	10.8	40.3	11.1	-0.3***
Urban (Rural)	.79	.41	.80	.40	.78	.42	0.0***
Married Status	.52	.50	.57	.50	.48	.50	0.1***
Num. Children	.90	1.07	.90	1.08	.90	1.05	-0.00
Family income	4,893	2,697	5,003	2,728	4,804	2,660(	198***

NOTE.—Data are sourced from the UKHLS covering from 2010 to 2020. 'Monthly wage' refers to hourly income from the primary job. 'Working hours' and 'Overtime hours' are measured on a monthly basis. 'Firm size' is an ordered variable ranging from 1 to 9, with a mean value of 5 corresponding to firms with 50-99 employees. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

2.4 Empirical analysis

2.4.1 Transitions into FWAs over time

This section explores the evolution of transition patterns into FWAs and the associated gender disparities over time. We use the UKHLS instead of the LFS because the longitudinal nature of the UKHLS panel allows us to explore individual-level transitions, mitigating selection bias and controlling for time-invariant unobserved heterogeneity, such as work-life balance preferences, long-standing caregiving responsibilities, or personality traits. Generally, we employ linear regression analyses to examine these trends. While Logit and Probit models might capture the more precise estimators for dummy-dependent variables, we opt for linear probability regression due to the direct comparability of its results across the different types of FWAs (Mood, 2010).[47,48] Our analysis utilizes six distinct dummy variables, y_{it}, each corresponding to a different type of FWAs, such as part-time, term-time, job sharing, flexi-time, compressed hours, and annualized hours. These dummies take a value of one if the individual i is engaged in the particular FWA at time t, and zero for traditional work. The regression model for each job transition can be formulated as follows:

$$y_{itao} = \alpha_0 + \sum_t \delta_t D_t + \boldsymbol{X}'_{itao}\eta + \rho_i + \gamma_a + \mu_o + \varepsilon_{itao} \tag{2.1}$$

where y_{itao} is a set of six unique dummy variables of FWAs for individual i in the industry a and occupation o at time t; D_t denotes an indicator variable for each year, including years 2012, 2014, 2016, 2018, and 2020, with the base year of 2010 being omitted.[49] $\boldsymbol{X}'_{itao}$ represents a matrix of time varying control variables, such as age, age square, marital status, number of children, family income, and firm size; ρ_i, γ_a and μ_o denote the individual, industry and occupation fixed effect. To address potential correlation of error terms within groups, we cluster the standard errors at the time-industry and time-occupation levels.

Panel (a) in Figure 2.1 displays the estimated coefficients from a series of linear regression models, illustrating the job transition from traditional working to various FWAs among men and women over time. For man, the transition from traditional work

[47] Using Logit or Probit models may exclude observations where some workers do not change jobs over time, which could result in different sample compositions. Comparing across these different samples may bias our results and compromise comparability.

[48] The marginal effects estimated in Logit or Probit models are expected to converge with those obtained from linear estimators as the sample size large substantially.

[49] Due to the inclusion of individual fixed effects, the coefficient δ_t specifically captures the probability of an individual's transition level. Given that time is treated as an exogenous variable and is unlikely to correlate directly with the error term ε_{itao}, we argue that δ_t capture the casual period effect on various FWAs, after controlling for the age and age squared.

to FTW shows a consistent increase by 3, 8, 12, 17, and 21 percentage points in the years 2012, 2014, 2016, 2018, and 2020, respectively, compared to the baseline year of 2010 (see figure a). Similarly, men's transition to TTW also indicates a stable upward trend, growing by approximately 2 percentage points every two years, although the estimates are not statistically significant. Conversely, the transition patterns into FWAs for women significantly differ (see figure b). Among women, the transition to TTW increase consistently by 4 percentage points every two years, reaching 22 percentage points higher in 2020 compared in 2010. Unlike to men, women are less likely to transition to FTW, with a gradual decrease by 13 percentage points between 2010 and 2020. Unlike TTW and FTW, we do not observe any significant changes in PTW over time for either men or women.[50] This finding strongly supports our first hypothesis that, women are more likely to transition to TTW over time, while men are more likely to shift to FTW.

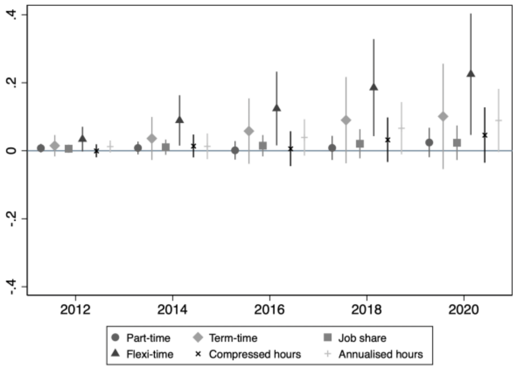
To further explore whether the observed trends in TTW and FTW are driven by part-time or full-time employment, we divide the data accordingly. As shown in figure (c) of Panel B, the trends are primarily driven by full-time male workers. Over time, part-time male workers show no significant change in their employment patterns in either FTW or TTW. Similarly, the upward trend in TTW transition is largely dominated by full-time female workers. While the figure provides some evidence of part-time female workers transitioning to TTW, these changes are not statistically significant. Given that the trends are primarily driven by full-time employees, the subsequent analysis focuses on full-time men and women. Nonetheless, to ensure that the observed trends in FWAs are not a consequence of attrition bias, we employ multiple imputation as a robustness check. Appendix E shows that the application of multiple imputation methods corroborates the increasing trend for full-time female workers in TTW and male workers in FTW.

[50] For part-time workers, according to the UKHLS definition, we classify individuals working fewer than 30 hours per week as part-time; otherwise, they are considered full-time.

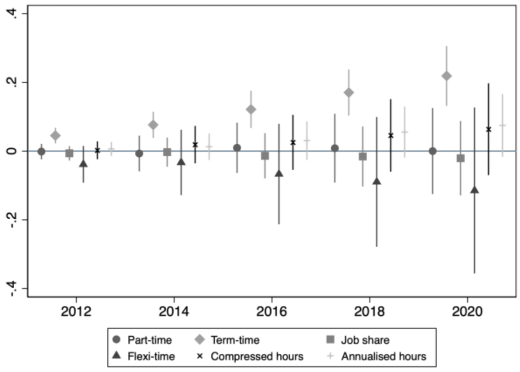
Figure 2. 1: The transition from traditional work to FWAs from 2010 to 2020

Panel A: All types of FWAs

(a) Men

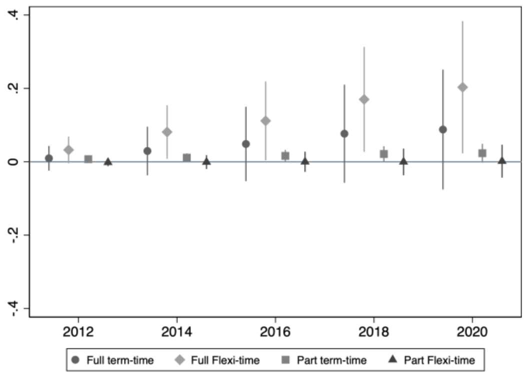


(b) Women

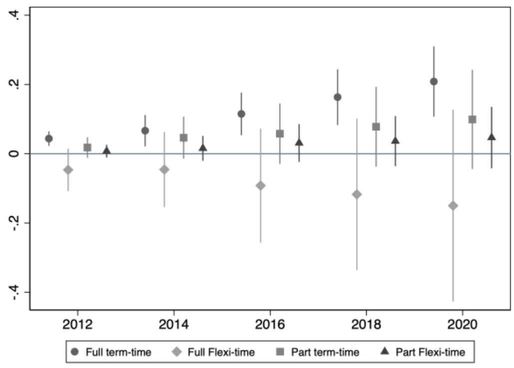


Panel B: Full-time vs. Part-time

(c) Men



(d) Women

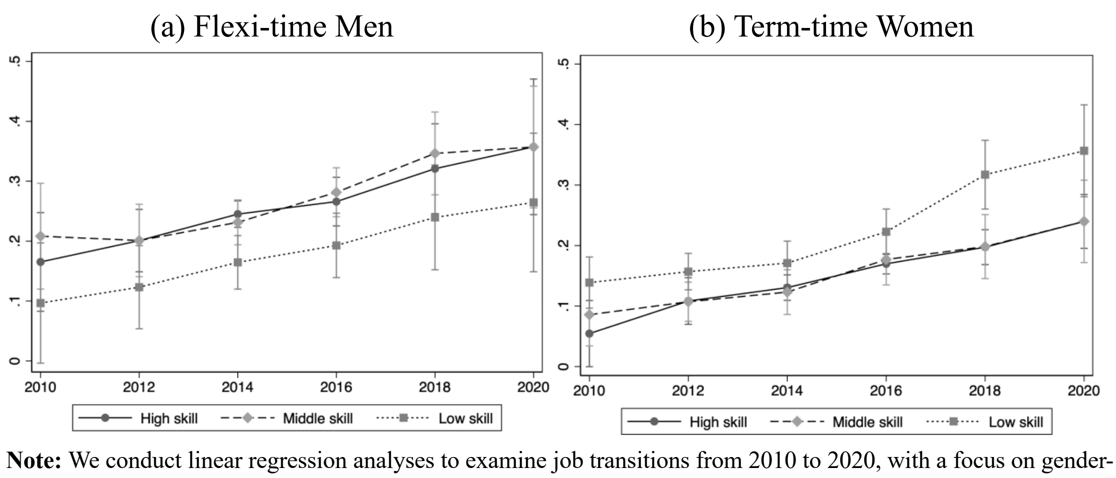


Note: We conduct linear regression analyses to examine job transitions from 2010 to 2020, with a focus on gender-specific patterns. The full set of controls included age, age square, marital status, number of children, family income, and firm size, with confidence intervals set at the 90% level. Panel A shows the transition from traditional work to various FWAs overtime. Panel B shows the transition from traditional work to TTW and FTW overtime, distinguishing between full-time and part-time employment. The x-axis represents the years from 2012 to 2020 with 2010 serving as the reference year. The y-axis indicates the probability of an employee transitioning from traditional working arrangements to various types of FWAs. Standard errors are clustered at the time-industry and time-occupation levels.

The above results highlight a notable increase in the proportion of full-time men and women transitioning from traditional working arrangements to FTW and TTW over the past decade. To better understand these trends, we further explore disparities in the

FWAs transition at the occupational level. For a more structured analysis, we utilize the Three Class National Statistics Socio-economic classification (NS-SEC), which categorizes occupations into three distinct levels: low, middle, and high. Figure 2.2 shows the observed upward trend in transitioning to TTW and FTW across three occupational groups for both men and women. However, the magnitude of these trends varies significantly. As shown in Figure (a), men in high and middle skill occupations are consistently around 10 percentage points more likely to transition to FTW than men in low skill occupations between 2010 and 2020. Conversely, transitions to TTW are more likely among low skilled women throughout the entire study period, with probabilities ranging from 15 to 35 percentage points, which is about 1.6 times higher than those in high and middle skill occupations. Overall, our findings reveal that men are more likely to transition to FTW in high and middle skill (hereafter, referred to as high-mid) occupations over time, whereas women tend to transition to TTW in low skill occupations. As discussed earlier, women may downgrade their occupational position to secure TTW, and such gendered occupational segregation may further exacerbate the gender gap, as we will explore in the following sections.

Figure 2. 2: The transition from traditional work to FWAs from 2010 to 2020 by occupational levels



Note: We conduct linear regression analyses to examine job transitions from 2010 to 2020, with a focus on gender-specific patterns. The coefficients are shown by marginal levels. Figure (a) focuses on the transitions of full-time employed male workers from traditional work to FTW, and Figure (b) focuses on the transitions of full-time employed female workers from traditional work to TTW, differentiated by occupation levels. The x-axis represents the years from 2012 to 2020. The y-axis indicates the probability of an employee transitioning from traditional working arrangements to TTW or FTW. The full set of controls included age, age square, marital status, number of children, family income, and firm size, with confidence intervals set at the 90% level., with confidence intervals set at the 90% level. Standard errors are clustered at the time-industry and time-occupation levels.

2.5 Transition effects on total working hours and monthly wage

This section explores the impact of significant gender disparities in FWAs usage on true working hours and monthly wage, and how these transition patterns further exacerbate the gender wage gap in labor market.

A. Identification strategy

The analysis employs a log-linear model, regressing the natural logarithm of total working time and monthly wages to examine the causal effect of transitioning to TTW and FTW. Unlike previous research that focuses solely on regular working hours (e.g., e.g., Allen et al., 2013; Chung and van der horst, 2018; Chandola et al., 2019), we consider a broader scope of working time, including overtime working hours, weekend working hours and paid holiday days. This is because employees in FTW may often shift their regular working hours to accommodate overtime and weekend work (Kelly et al., 2014). Conversely, TTW is typically associated with considerable paid holiday days, often aligned with the school calendar (Gregory and Milner, 2009). In general, the total working time TWH_{itao} is expressed by following formula:

$$TWH_{itao} = (R_{itao} + O_{itao} + W_{itao}) \times \left(1 - \frac{H_{itao}}{365}\right) \qquad (2.2)$$

where R_{itao}, O_{itao}, W_{itao} and H_{itao} denote regularly, overtime, weekend working hours, and paid holiday days, respectively.[51] The use of TWH_{itao} is motivated by two

[51] Our calculation of total overtime working hours includes both paid and unpaid overtime. This approach is adopted because the monthly wage typically reflects monthly performance. Therefore, even in cases where workers work

reasons. Firstly, transitions to different FWAs involve substantial variation in working time patterns. Secondly, omitting the effect of transition on the total working hours may introduce significant bias into wage estimates. As shown in Table K.1 in Appendix K, controlling only for R_{itao} and O_{itao} underestimate the transition effect on wage outcomes in both TTW and FTW, as they typically involve fewer W_{itao} and more H_{itao}.

To identify TWH_{itao}, we link the Labor Force Survey (LFS) and Time Use Survey (TUS) to UKHLS, because the UKHLS collects information only on R_{itao} and O_{itao} but providing limited information on W_{itao} and H_{itao}. Specifically, while the UKHLS records whether individuals engage in weekend work, it does not capture the duration.[52] To overcome this limitation, we utilize the 2014-2015 TUS, which provides detailed records of participants' activities at 10-minute increments throughout the day. By analyzing this detailed time-use data, we can aggregate the total hours worked on weekends and compute the average weekend working hours by different FWAs, industries, and gender, linking this information to the UKHLS data.[53] Regarding holiday days, we use the LFS (i.e., which captures the number of paid holiday days entitled per year) to fill this gap.[54] Similarly, we maintain the assumption that the number of paid holiday days remains stable within each category (e.g., FWAs, industry, occupation, year, and gender), enabling us to link the information in LFS to UKHLS (for more details on the linking strategy, see Appendix B).[55,56] In general, the model

unpaid overtime, this contribution is potentially factored into the monthly wage.

52 In the UKHLS, weekend work information is collected through the question: "*Do you ever work at weekends?*" The possible responses are: 1) "*Yes - most/every weekend*"; 2) "*Yes - some weekends*"; and 3) "*No weekend working*". We create a dummy variable to capture whether an individual engages in weekend working, categorizing those who respond "*most/every weekend*" as engaging in weekend worker, and all others as non-weekend workers.

53 Since the UKHLS collects information on whether individuals work on weekends, we link only those who report working on weekends.

54 The definition of paid holiday excludes the public holidays.

55 As displayed in Figure B.2 in Appendix B, the number of paid holiday days for those engaged in TTW is typically twice that of other FWAs across most industries.

56 A key issue in calculating total working hours is avoiding double-counting weekend hours. If weekend work is reported as part of overtime in the UKHLS, total hours may be overstated. To verify, we refer to the LFS, which ask: "*Thinking now about the seven days ending Sunday, how many hours did you actually work in your (main) job/business?*". In LFS, full-time workers average 175 monthly hours. By contrast, UKHLS reports 169 hours when weekend work is excluded, rising to 174 hours when included, closely matching the LFS figure (Table B.2.3 in Appendix B). This suggests that weekend work is not included in UKHLS overtime reporting, leading to potential underestimation of total working hours.

focuses on transitions at the individual level to FWAs as outlined in the formula below:

$$
\begin{aligned}
y_{itao} = \alpha_0 + \sum_{j\in\{P,T,F\}} \delta_j TRAN_{itao}^{j} \times G_{itao} + \sum_{j\in\{P,T,F\}} \beta_j TRAN_{itao}^{j} \\
+ \tau G_{itao} + \boldsymbol{X}'_{itao}\eta + \rho_i + \gamma_a + \varphi_t + \mu_o + \varepsilon_{itao}
\end{aligned}
\tag{2.3}
$$

where y_{itao} represents the natural logarithm of outcome of interest for employee i in industry a, and occupation o, at time t. In the following Section B, we first examine work-life balance by analyzing the allocation of working time patterns and housework hours to understand the mechanisms behind the gender disparity in FWA transition patterns. Next, we focus on total working hours and monthly wages to discuss the consequences of the gender wage gap in relation to gendered FWA transition patterns. $TRAN_{itao}^{j}$ captures the employee transition from traditional arrangement to part-time (P), term-time (T) and flexi-time (F) working arrangement. G_{itao} denotes the dummy variable for gender. $\boldsymbol{X}'_{itao}$ represents a matrix that includes a set of time-varying control variables, such as age, age square, marital status, number of children, family income, and firm size. In addition, TWH_{itao} is added when estimating the treatment effects on wages. To eliminate time-invariant unobserved factors and capture the individual transition effect, we include individual fixed effects ρ_i. We also include the industry fixed effect γ_a and occupation fixed effect μ_o to control the industry-occupation-specific unobserved factors. The time fixed effects φ_t control for year variation in the outcome. In general, the coefficients δ_j are crucial for comparing the working time and monthly wage outcomes before and after transitions.

There are three additional concerns regarding the identification of the causal effect of the transition effect on the outcome variables. Firstly, bias may arise because TTW follows the academic calendar, and if the UKHLS data only collects survey in during academic months, it could result in an overestimation of TTW working time and monthly wages. To examine potential fluctuations in the TTW monthly wage from different academic terms, we utilize the detailed information from UKHLS, which records the interview month for each respondent, covering the full calendar year January to December. As shown in Appendix C, both male and female TTW participants exhibit a pronounced cyclical pattern in working time and monthly wage. To accurately capture these seasonal fluctuations, we incorporate seasonal fixed effects φ_s based on the information on interview month. This allows us to further control for systematic variations associated the timing of data collection.

Secondly, individual characteristics may affect both the likelihood of transitioning to different FWAs and subsequent labor outcomes. For instance, employees facing time pressure may self-select into FWAs to reduce their working hours, thereby leading to a decrease in labor income (Chuang, 2018). Additionally, employees with greater autonomy in their work may be more inclined to transition to FWAs in order to rearrange their working time patterns according to personal preferences and needs (Chuang, 2022). To address both forms of self-selection bias, we include two variables as controls "*How much of the time has your job made you feel tense?*", and "*Autonomy over work hours*".[57]

Thirdly, bias may still arise from other unobserved variables, such as family and firm characteristics. To address this potential endogeneity issue, we employ the shift-shared instrumental variable (SSIV) method (Autor et al., 2013; Goldsmith-Pinkham et al. 2020). Our SSIV method leverages variation from two components: regional shifts and industry shares. The shift component reflects overall growth in FWAs adoption, as indicated by regional-level increases in TTW and FTW. These shifts are presumed exogenous to individual region characteristics, driven by broad factors like FWAs policies, which have had a national influence on the labor market. The share component captures the distribution of FWA adoption in the previous year within each region's industry composition. This use of historical data ensures that the share

[57] There are five autonomy questions in the UKHLS: "*Autonomy over job tasks, work pace, work manner, task order, and work hours*". As shown in Appendix I.2, whether we control for each of these factors individually or use factor analysis to construct a composite factor representing overall work autonomy, the results remain consistent. To avoid the overcontrol problem, we only include the "*Autonomy over work hours*".

component is not endogenously determined by current regional characteristics in the labor market. To further strengthen the analysis, we further use FWAs supply data from UKHLS and the larger sample of FWAs adoption from LFS to construct the SSIV. For a more detailed explanation of the construction and results of the SSIV, please refer to Appendix K.

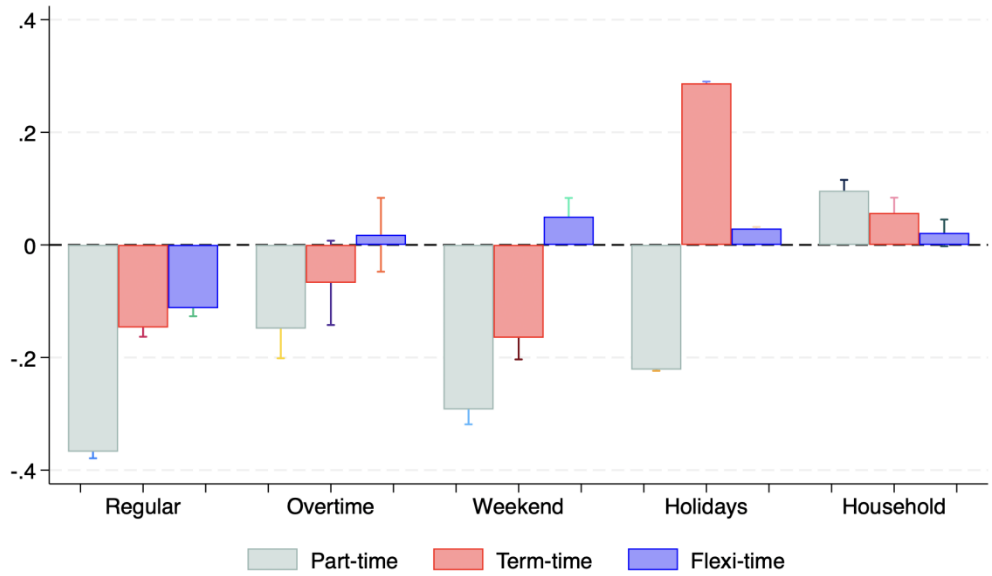
B. Empirical results on work-life time patterns

Figure 2.3 presents the effects of transitioning from traditional work to part-time work (PTW), term-time work (TTW), and flexi-time work (FTW) on various working time patterns and housework hours by gender, controlling for personal and firm characteristics, as well as incorporating time, personal, and occupation fixed effects. Panel (a) reveals that transitioning to PTW leads to significant reductions in working hours for women, including regular working hours (38%; $p<0.01$), overtime (17%; $p<0.01$), and weekend hours (30%; $p<0.01$). Interestingly, transitioning to PTW also results in fewer paid holiday days, likely because entitlement to paid holidays is often proportional to the number of hours worked. However, as discussed earlier, the substantial reduction in working hours under PTW may prompt women to seek alternative reduced-hour arrangements, such as TTW, in order to mitigate the negative impact on their total working time. Our results indicate that the reduction in working time for TTW is only half the size of that PTW. More precisely, women transitioning from traditional work to TTW experience smaller reductions in regular hours (17%; $p<0.01$), overtime (6%; $p<0.01$), and weekend work (18%; $p<0.01$). This is because, although TTW is part of the reduced-hours category, it still requires a full-time commitment during the active work periods. These findings support our hypothesis that women engaged in TTW tend to work more hours than those in PTW, potentially mitigating associated labor market penalties. To further validate this mechanism, we examine the transition wage penalty in the following section. Interestingly, since women transitioning to TTW maintain higher working hours compared to those in PTW, they appear to reduce their time spent on household chore proportionally. Our results support this, showing a 5% ($p<0.01$) reduction in household chores time, suggesting the notion that women with higher working hours under TTW allocate less time to domestic responsibilities. This finding contrasts with previous research suggesting that women engaging in FWAs may experience self-exploitation in both the labor market and household work (Chung, 2018; Henz, 2006), but we find that they often prioritize maximizing work-life balance.

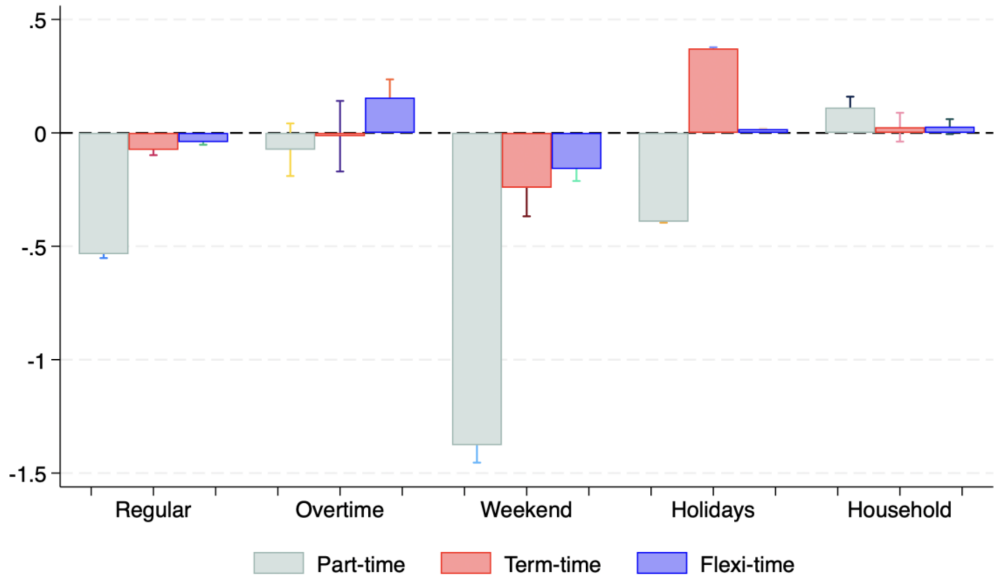
Unlike PTW and TTW, women transitioning to FTW reduce their regular working hours by 10% ($p<0.01$) but compensate by increasing their weekend working hours by 5% ($p<0.01$). In terms of domestic responsibilities, using FTW does not result in a statistically significant increase in time spent on household chores, with only a 2% change ($p>0.1$). A similar pattern is observed for men, where no significant change in household chore hours is found when transitioning to FTW. This is likely because FTW is designed to enhance job performance by rearranging or extending working hours. This type of arrangement is particularly common among men, as societal norms often align with the "ideal worker", which expects men to prioritize work over family responsibilities. As shown in Panel (b), after transitioning to FTW, men experience a slight reduction in regular working hours (4%; $p<0.01$) but a significantly increase their overtime hours (16%; $p<0.01$). Although weekend working hours decrease, the rise in overtime suggests a shift toward more concentrated working hours during the week. Interestingly, unlike women, men transitioning to TTW experience lower regular working hours (8%; $p<0.01$) and weekend working hours (25%; $p<0.01$), while having more holiday days (37%; $p<0.01$), and no change in domestic responsibilities. This finding is consistent with recent research by Wang and Cheng (2024), which shows that men engaged in FWAs do not necessarily increase their share of household work.

Figure 2. 3: Impact of Transitioning on Various Work-Life Time Patterns by Gender

a) Women



b) Men



Notes: Each figure reports the impact of a transition from full-time traditional work to full-time term-time (a) and full-time flexi-time working (b) on various working patterns (e.g., regular working hours, overtime working hours, weekend working hours and paid holiday days) by gender. The control variables included age, age square, marital status, number of children, family income, and firm size. The y-axis represents the percentage change in work-life time when employees transition from traditional working arrangements to PTW, TTW, or FTW.

C. Empirical results on total working hours and labor income

Table 2.2 presents the gender-specific effects of the transition from traditional work to PTW, TTW and FTW on total working hours and monthly wages (see Equation (2.3)). The separate effects by gender are estimated by interacting the main independent variables with gender dummies.[58] As shown in columns 2 and 5, the reduction in total working hours for women transitioning to PTW (36.8%; $p<0.01$) is twice as large as for those transitioning to TTW (18.9%; $p<0.01$), while the transition wage penalty for PTW (14.7%; $p<0.01$) is nearly three times greater than that for TTW (6.3%; $p<0.01$). These findings largely support our argument that more women are transitioning to TTW over time (see Figure 1), possibly because TTW reduces total working hours compared to traditional working arrangements while incurring a smaller transition wage penalty than PTW. Notably, the transition wage penalty of PTW is even more substantial, at 32%, when accounting for lower total working hours.[59] Under these conditions, women may opt for other reduced-hours arrangements, such as TTW, to maximize their work-family balance, as its total transition wage penalty of14.9% is significantly lower than PTW, making it a more attractive option.

Moreover, although TTW is categorized as a reduced-hours arrangement, it is often treated as a full-time role. In this specific context, the stigma associated with flexible work and job discrimination maybe lower for TTW compared to PTW. To provide additional evidence on discrimination, we further show that women transitioning to PTW are associated with lower job satisfaction and higher levels of job-related worry and anxiety, compared to those transitioning to TTW (see Appendix L). The wage penalty for transitioning to FTW is smaller, at only 3.7% ($p<0.05$), and total working

[58] Notably, we cannot estimate the gender wage gap within traditional work arrangements due to the use of individual fixed effects. However, we do capture how transitions from traditional work to different FWAs by men and women contribute to the further widening of the gender wage gap.
[59] The direct effect is 15%, while the indirect effect, mediated through total working hours, is calculated as $37\% \times 45\%$.

hours are reduced by 10.9% ($p<0.01$), However, the reduction in working hours is not efficiently transferred to family responsibilities. In fact, women transitioning to FTW often compensate by overtime hours and weekends hours (see Figure 2.3). Such a schedule does not facilitate a smooth transition to managing domestic duties and may conflict with traditional gender norms. Accordingly, the proportion of women transitioning to FTW has declined over time (see Figure 2.1). In contrast, women transitioning to TTW tend to reduce their regular, overtime, and weekend working hours while increasing their paid holiday days. This enables them to better align their work schedules with children's school schedules and domestic duties, while incurring a smaller wage penalty compared to part-time work.

For men (see columns 1 and 4), the direct wage penalty associated with transitioning to TTW is nearly four times greater than that for FTW, with monthly wage reductions of 4.8% ($p<0.05$) and 1.2% ($p>0.1$), respectively. Similarly, when accounting for indirect transition effects through the total working hours, the total transition wage penalty for TTW is about 9.7% for TTW. Given that societal norms often encourage men to adopt the "ideal worker" role and act as primary breadwinners, they are more likely to opt for FTW. This is because transitioning to FTW typically does not result in a statistically significant wage penalty or a reduction in total working hours, while still allowing men to reschedule their working time more flexibly (see Figure 3). This finding largely supports our hypothesis that men increasingly use FTW to adjust their working time patterns, though we do not find evidence that they use FTW to extend working hours for higher earnings.

Although gender differences in FWAs adoption may be influenced by individual constraints, it is important to recognize that gender inequality in working time and labor income may be further exacerbated by the provision of FWAs. Specifically, the direct effect of gender gap in wage penalty for transitioning to TTW and FTW is about 1.5% and 2.5%, respectively. This disparity becomes even more pronounced when accounting for the indirect transition effect on the total working hours, as men do not experience a significantly reduction when transitioning to FTW. According to the mediation model, which captures both direct and indirect effects of these transitions, the gender gap in wage penalty increases further to 5.2% for transitions to TTW and to 6.3% for transitions to FTW.[60] Nonetheless, given that FWAs reinforce traditional gender roles, women are more likely to transition to TTW to accommodate family responsibilities, while men are more likely to switch to FTW to enhance job performance (see Figure 2.1). This pattern may further widen the gender wage gap. As shown in Table 2.2, women often face substantial reductions in both working time and labor income when transitioning to TTW, while men encounter minimal penalties when switching to FTW. Specifically, the gap in the transitioning wage penalty and total working hours between men transitioning to FTW and women to TTW is 5.1% ($p<0.01$) and 15.8% ($p<0.01$), respectively, as shown in the comparison term "*Man[FTW]−Woman[TTW]*". Furthermore, using the mediation model to account for the indirect effect, the gender wage gap widens further to 12.6% when factoring in the gender disparity in FWAs. These results are robust, as all three of our SSIV estimates in Appendix K demonstrate that, after correcting for endogeneity, the effects of TTW and FTW on total working hours and monthly wages remain largely unchanged. Additionally, Table K.2.2 in Appendix K includes detailed tests that confirm the validity of our SSIV approach.

Table 2. 2: The effect of transition on monthly wage by gender

	Total working hour			Monthly wage		
	Men	Women	Diff	Men	Women	Diff
	(1)	(2)	(3)	(4)	(5)	(6)
PTW	-0.530***	-0.368***	-0.162***	-0.260***	-0.147***	-0.114
	(0.037)	(0.015)	(0.026)	(0.066)	(0.007)	(0.066)
TTW	-0.133***	-0.189***	0.056***	-0.048**	-0.063***	0.014
	(0.026)	(0.014)	(0.018)	(0.026)	(0.017)	(0.035)
FTW	-0.031	-0.109***	0.078***	-0.012	-0.037**	0.025**

60 The total transition effect on the gender wage gap can be decomposed into two components: the difference in the direct effects of the treatment on wages between men and women, and the difference in the indirect effects, which arises from the treatment's differential impact on total working hours and the corresponding returns to working hours on wages for each gender.

	(0.021)	(0.011)	(0.021)	(0.015)	(0.016)	(0.011)
Total working hour				0.369***	0.453***	-0.084**
				(0.054)	(0.020)	(0.039)
Man[FTW] − Woman[TTW]:						
			0.158***			0.051***
			(0.024)			(0.018)
N		40,997			40,997	
R^2		0.7842			0.8953	
Basic controls		Yes			Yes	
Individual & Time FE		Yes			Yes	
Industry & occupation FE		Yes			Yes	
P&A, seasonal		Yes			Yes	

NOTE.—The table shows estimates for the effects of transiting from traditional work to TTW and FTW on total working hours and monthly wage among full-time employees, segmented by gender. We include two interaction terms: TTW × gender and FTW × gender, and recalculate the coefficients for both genders. The "*Man[FTW] − Woman[TTW]*" term refers to the disparity in outcomes when men transition from traditional work to FTW and women to TTW, as analyzed using a t-test. The control variables included age, age square, marital status, number of children, family income, and firm size. Standard errors, clustered at the level of the gender-industry level, are shown in parentheses. *** p < 0.01.** p < 0.05.* p < 0.1.

D. The transitional penalty by occupation levels

So far, our findings indicate that the gender gap in labor outcomes may initially widen in the short term but narrows over the long term as women transition to TTW. However, gendered occupational skill-level segregation may further widen this gender gap. As discussed earlier, women are disproportionately represented in low skill jobs, a pattern particularly evident in TTW, which is heavily concentrated in the education industry (see Appendix D). To secure TTW roles, women may face career downgrading or stagnation, as supported by Appendix H, which shows that when women transition to TTW, there is a 0.715 log-odds ($p<0.01$) increase in the likelihood of moving from high to low skilled roles and a 1.099 log-odds ($p<0.01$) increase in the likelihood of moving from middle to low skill roles. Further evidence is provided in Figure 2.2, which illustrates that women predominantly transition to low skill TTW over time, whereas men are more likely to transition to high-middle skill FTW. This transition pattern may be a key factor further contributing to the widening of the gender wage gap.

As displayed in Table 2.3, we find that within high-mid skill jobs, there is no significant transition wage penalty for men in either TTW or FTW (see column 1 in Panel B). A significant wage penalty is observed only when men in low skill occupations transition to TTW, with a reduction of 17.8% ($p<0.01$), which coincides with a similar reduction in total working hours of 28.8% ($p<0.01$). If men priorities wages and career development over reduced working hours, they are more likely to choose other FWAs, such as FTW, which do not carry a significant wage penalty. This evidence supports the hypothesis that men are more likely to transition to high-mid skill FTW roles, where the transition wage penalty is minimal (see Figure 2.2). For women, however, the transition wage penalty for both TTW and FTW in high-mid skill level jobs is relatively higher, at about 3%-5% ($p<0.05$), compared to the men. Similarly, women who transition to TTW and FTW in this group reduce their total working hours more than men, by around 10%-16% ($p<0.01$). This suggests that the transition to TTW or FTW increases the gender gap within the high-mid skill occupations. Given the observed pattern where women are more likely to transition to low skill TTW, while men are more likely to transition to high-mid skill FTW, a substantial transition wage penalty gap between men and women emerges, amounting to approximately 7.4% ($p<0.01$). Moreover, this gender wage gap grows to 18.8% when accounting for the indirect effects through total working hours and occupation downgrading.

Table 2. 3: The effect of transition on total working hours and monthly wage by gender and occupational levels

	Men		Women		Diff.	
	H-M	L	H-M	L	H-M	L
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. total working hours						
TTW	-0.105***	-0.288***	-0.160***	-0.233***	0.055***	-0.055
	(0.018)	(0.079)	(0.005)	(0.028)	(0.016)	(0.062)
FTW	-0.034*	-0.015	-0.106***	-0.115***	0.072***	0.010**
	(0.017)	(0.039)	(0.013)	(0.016)	(0.020)	(0.039)
Man[FTW_H-M] − Woman[TTW_L]:						

						0.199***
						(0.027)
Panel B. monthly wage						
TTW	-0.023	-0.178***	-0.050***	-0.080***	0.026	-0.098
	(0.015)	(0.060)	(0.015)	(0.018)	(0.025)	(0.065)
FTW	-0.007	-0.038	-0.033**	-0.075***	0.026**	0.037
	(0.016)	(0.027)	(0.015)	(0.009)	(0.012)	(0.028)
Total working hour	0.254***	0.550***	0.391***	0.525***		
	(0.044)	(0.069)	(0.020)	(0.001)		
Man[FTW_H-M] − Woman[TTW_L]:						
						0.074***
						(0.026)
N	40,997		40,997			
R^2	0.7194		0.8704			
Basic controls	Yes		Yes			
I & T & O FE	Yes		Yes			
P&A, seasonal	Yes		Yes			
IV	Yes		Yes			

NOTE.—The table shows IV estimates for effects of transitioning from traditional work to TTW and FTW on total working hours and monthly wages, segmented by gender and occupation skill level. We include two interaction terms: TTW × gender × occupation skill level and FTW × gender × occupation skill level, and recalculate the coefficients for both genders and all occupation skill levels. The "*Man[FTW_H-M] - Woman[TTW_L]*" term refers to the disparity in outcomes when men transition from traditional work to high-mid skill FTW occupations and women to low skill TTW occupations, as analyzed by using a t-test. The control variables included age, age square, marital status, number of children, family income, and firm size. Standard errors clustered at the level of the gender-industry level, are shown in parentheses. *** p < 0.01.** p < 0.05.* p < 0.1.

2.6 Discussion and Conclusion

Despite significant progress over the last half-century, the gender wage gap persists. Although the UK government has promoted FWAs as a social intervention to facilitate women labor participation and reduce the gender wage gap, previous research indicates that their impact may be double-edged. We started by examining gender disparities in transitions to FWAs and found that, over the last decade in the UK, full-time employed men are more likely to transition from traditional work to FTW, while full-time employed women more frequently switch to TTW. To explore the mechanism behind this finding, we explore both working time patterns and transition wage penalties. By integrating LFS and TUS information into the UKHLS, we quantify the total working time patterns, including regular working hours, overtime working hours, weekend working hours and paid holiday days. Our analysis shows that each type of FWAs capitalizes on the distinctive working time patterns it embodies. For instance, women transitioning from traditional work to TTW tend to reduce their regular, overtime, and weekend working hours while increasing paid holiday days. This adjustment allows them to better align their schedules with their children's school calendars and family duties, while incurring a smaller wage penalty compared to PTW. To further support this argument, we track the timeline of these transitions and find that, upon transitioning to TTW, women reduce their total working hours in the short-term, but in the long-term their total working hours almost recover to pre-transition levels. This trend aligns with fertility patterns, as women are more likely to transition to TTW after having a child.

For those transitioning to FTW, neither men nor women experience a significant reduction in total working hours, as they primarily rearrange their working time patterns. Given that transitioning to FTW does not significantly reduce total working hours, its effectiveness in alleviating work-family conflict may be limited. However, within the constraints of gendered structures, men are typically expected to prioritize their roles as ideal workers or breadwinners, making FTW a more viable option for them. This allows men to transition into FTW with no significant wage penalties, while continuing to be perceived as commitment worker.

Overall, although it appears reasonable for men and women to choose different FWAs to optimize their work-life balance, it is important to note that the gender wage gap may further widen with the provision of FWAs. The gap in the transition wage penalty between men transitioning to FTW and women to TTW is 5.1%. This gap grows even larger when accounting for the indirect transition effect through total working hours, resulting in a total transition wage penalty of approximately 12.6%. The gender wage gap is further exacerbated by the observed pattern that men are more likely to transition to high-middle skill FTW roles, while women tend to transition to lower skill TTW roles. That is, we find evidence that women may face career downgrading or stagnation for transitioning to TTW. This gender wage gap is significant, as there is no notable transition wage penalty for men in high-middle skill jobs transitioning to FTW. In contrast, women in low skill jobs face a significant wage penalty of approximately 8.0% when transitioning to TTW. Furthermore, when accounting for the indirect effects through total working hours, this gender wage gap expands to 18.8%. These results are robust, as the application of the shift-share instrumental variable method to estimate the causal impact of FWAs on both working hours and monthly wages yields findings consistent with our baseline models.

This paper has several limitations. First, total working hours are measured using three different datasets, and the use of average imputation may introduce some measurement bias. Second, the analysis focuses only on employed individuals; therefore, it does not capture transitions involving unemployment or movements from unemployment into different jobs. These limitations suggest several directions for future research. First, future studies could use datasets with more consistent and harmonised measures of working hours, or employ alternative methods to handle missing information, in order to reduce potential measurement bias. Second, future research could extend the analysis beyond the employed population to explicitly model labour market entry and exit, including transitions into and out of unemployment.

2.7 References

Allen, Tammy D., Ryan C. Johnson, Kaitlin M. Kiburz, and Kristen M. Shockley. 2013. "Work–Family Conflict and Flexible Work Arrangements: Deconstructing Flexibility." *Personnel Psychology* 66(2):345–76.

Alsulami, Alanoud, Fatma Mabrouk, and Jihen Bousrih. 2022. "Flexible Working Arrangements and Social Sustainability: Study on Women Academics Post-COVID-19." *Sustainability* 15(1):544.

Angrave, David, and Andy Charlwood. 2015. "What Is the Relationship between Long Working Hours, over-Employment, under-Employment and the Subjective Well-Being of Workers? Longitudinal Evidence from the UK." *Human Relations* 68(9):1491–1515.

Ashforth, Blake E., Glen E. Kreiner, and Mel Fugate. 2000. "All in a Day's Work: Boundaries and Micro Role Transitions." *The Academy of Management Review* 25(3):472–91.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6):2121–68.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3):789–865.

Booth, Alison L., and Jan C. Van Ours. 2008. "Job Satisfaction and Family Happiness: The Part-Time Work Puzzle." *The Economic Journal* 118(526):F77–99.

Bratberg, E. 2002. "'The Double Burden': Do Combinations of Career and Family Obligations Increase Sickness Absence among Women?" *European Sociological Review* 18(2):233–49.

Busch, Felix. 2020. "Gender Segregation, Occupational Sorting, and Growth of Wage Disparities Between Women." *Demography* 57(3):1063–88.

Cech, Erin A., and Mary Blair-Loy. 2014. "Consequences of Flexibility Stigma Among Academic Scientists and Engineers." *Work and Occupations* 41(1):86–110.

Cha, Youngjoo. 2010. "Reinforcing Separate Spheres: The Effect of Spousal Overwork on Men's and Women's Employment in Dual-Earner Households." *American Sociological Review* 75(2):303–29.

Cha, Youngjoo, and Kim A. Weeden. 2014. "Overwork and the Slow Convergence in the Gender Gap in Wages." *American Sociological Review* 79(3):457–84.

Chandola, Tarani, Cara L. Booker, Meena Kumari, and Michaela Benzeval. 2019. "Are Flexible Work Arrangements Associated with Lower Levels of Chronic Stress-Related Biomarkers? A Study of 6025 Employees in the UK Household Longitudinal Study." *Sociology* 53(4):779–99.

Chatzitheochari, Stella, and Sara Arber. 2009. "Lack of Sleep, Work and the Long Hours Culture: Evidence from the UK Time Use Survey." *Work, Employment and Society* 23(1):30–48.

Chung, Heejung. 2018. "Dualization and the Access to Occupational Family-Friendly Working-Time Arrangements across Europe." *Social Policy & Administration* 52(2):491–507.

Chung, Heejung. 2020. "Gender, Flexibility Stigma and the Perceived Negative Consequences of Flexible Working in the UK." *Social Indicators Research* 151(2):521–45.

Chung, Heejung. 2020. "Part-time working women's access to other types of flexible working-time arrangements across Europe", in Heidi Nicolaisen, Hanne Kavli, and Ragnhild Steen Jensen (eds), *Dualisation of Part-Time Work: The Development of Labour Market Insiders and Outsiders*

Chung, Heejung, Holly Birkett, Sarah Forbes, and Hyojin Seo. 2021. "Covid-19, Flexible Working, and Implications for Gender Equality in the United Kingdom." *Gender & Society* 35(2):218–32.

Chung, Heejung, and Mariska van der Horst. 2018. "Women's Employment Patterns after Childbirth and the Perceived Access to and Use of Flexitime and Teleworking." *Human Relations* 71(1):47–72.

Chung, Heejung, and Tanja Van Der Lippe. 2020. "Flexible Working, Work–Life Balance, and Gender Equality: Introduction." *Social Indicators Research* 151(2):365–81.

Chung, Heejung. 2022. *The Flexibility Paradox: Why Flexible Working Leads to (Self-)Exploitation*. Bristol, UK: Polity Press.

Connolly, S., and M. Gregory. 2009. "The Part-Time Pay Penalty: Earnings Trajectories of British Women." *Oxford Economic Papers* 61(Supplement 1):i76–97.

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101(2):556–90.

Fuller, Sylvia, and C. Elizabeth Hirsh. 2019. "'Family-Friendly' Jobs and Motherhood Pay Penalties: The Impact of Flexible Work Arrangements Across the Educational Spectrum." *Work and Occupations* 46(1):3–44.

Gayle, George-Levi, and Limor Golan. 2012. "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997." *The Review of Economic Studies* 79(1):227–67.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4):1091–1119.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110(8):2586–2624.

Gregory, Abigail, and Susan Milner. 2009. "Editorial: Work–Life Balance: A Matter of Choice?" *Gender, Work & Organization* 16(1):1–13.

Grinza, Elena, Francesco Devicienti, Mariacristina Rossi, and Davide Vannoni. 2022. "How Entry into Parenthood Shapes Gender Role Attitudes: New Evidence from The UK." *Feminist Economics* 28(4):194–220.

Henz, Ursula. 2006. "Informal Caregiving at Working Age: Effects of Job

Characteristics and Family Configuration.” *Journal of Marriage and Family* 68(2):411–29.

Hilbrecht, Margo, Susan M. Shaw, Laura C. Johnson, and Jean Andrey. 2008. “‘I’m Home for the Kids’: Contradictory Implications for Work–Life Balance of Teleworking Mothers.” *Gender, Work & Organization* 15(5):454–76.

Jacobs, Jerry A., and Kathleen Gerson. 2004. “The Time Divide: Work, Family, and Gender Inequality”. *Harvard University Press*.

Kalleberg, Arne L., Barbara F. Reskin, and Ken Hudson. 2000. “Bad Jobs in America: Standard and Nonstandard Employment Relations and Job Quality in the United States.” *American Sociological Review* 65(2):256–78.

Kelly, Erin L., Phyllis Moen, J. Michael Oakes, Wen Fan, Cassandra Okechukwu, Kelly D. Davis, Leslie B. Hammer, Ellen Ernst Kossek, Rosalind Berkowitz King, Ginger C. Hanson, Frank Mierzwa, and Lynne M. Casper. 2014. “Changing Work and Work-Family Conflict: Evidence from the Work, Family, and Health Network.” *American Sociological Review* 79(3):485–516.

Kossek, E., Brenda A. Lautsch, and S. Eaton. 2005. “Flexibility Enactment Theory: Implications of Flexibility Type, Control, and Boundary Management for Work-Family Effectiveness.”

Kossek, Ellen Ernst, and Brenda A. Lautsch. 2018. “Work–Life Flexibility for Whom? Occupational Status and Work–Life Inequality in Upper, Middle, and Lower Level Jobs.” *Academy of Management Annals* 12(1):5–36.

Leuze, Kathrin, and Susanne Strauß. 2016. “Why Do Occupations Dominated by Women Pay Less? How ‘Female-Typical’ Work Tasks and Working-Time Arrangements Affect the Gender Wage Gap among Higher Education Graduates.” *Work, Employment and Society* 30(5):802–20.

Lier, David. 2010. “Grounding Globalization: Labour in the Age of Insecurity: Webster, E., Lambert, R. and Bezuidenhout, A.” *Journal of Economic Geography* 10(1):163–65.

Lips, Hilary M. 2013. “The Gender Pay Gap: Challenging the Rationalizations. Perceived Equity, Discrimination, and the Limits of Human Capital Models.” *Sex Roles* 68(3–4):169–85.

Lott, Yvonne, and Heejung Chung. 2016. “Gender Discrepancies in the Outcomes of Schedule Control on Overtime Hours and Income in Germany.” *European Sociological Review* 32(6):752–65.

Lott, Yvonne, and Christina Klenner. 2018. “Are the Ideal Worker and Ideal Parent Norms about to Change? The Acceptance of Part-Time and Parental Leave at German Workplaces.” *Community, Work & Family* 21(5):564–80.

Lu, Zhuofei, Senhu Wang, and Wendy Olsen. 2023. “Revisiting the ‘Flexibility Paradox’: Degree of Work Schedule Flexibility and Time Use Patterns across Gender and Occupational Groups.” *Humanities and Social Sciences Communications* 10(1):790.

Lyonette, Clare, and Rosemary Crompton. 2015. “Sharing the Load? Partners’ Relative Earnings and the Division of Domestic Labour.” *Work, Employment and Society* 29(1):23–40.

Manning, Alan, and Barbara Petrongolo. 2008. “The Part-Time Pay Penalty for Women in Britain.” *The Economic Journal* 118(526):F28–51.

McGinnity, Frances, and Patricia McManus. 2007. “Paying the Price for Reconciling Work and Family Life: Comparing the Wage Penalty for Women’s Part-Time Work in Britain, Germany and the United States.” *Journal of Comparative Policy Analysis: Research and Practice* 9(2):115–34.

Nippert-Eng, Christena. 1996. “Calendars and Keys: The Classification of ‘Home’ and ‘Work.’” *Sociological Forum* 11(3):563–82.

Presser, Harriet B. 2003. “Working in a 24/7 Economy: Challenges for American Families”. *Russell Sage Foundation*.

Piasna, Agnieszka, and Anke Plagnol. 2018. “Women’s Job Quality Across Family Life Stages: An Analysis of Female Employees Across 27 European Countries.” *Social*

Indicators Research 139(3):1065–84.

Risman, Barbara J. 2004. “Gender As a Social Structure: Theory Wrestling with Activism.” *Gender & Society* 18(4):429–50.

Rudman, Laurie A., and Kris Mescher. 2013. “Penalizing Men Who Request a Family Leave: Is Flexibility Stigma a Femininity Stigma?” *Journal of Social Issues* 69(2):322–40.

Shauman, Kimberlee A. 2010. “Gender Asymmetry in Family Migration: Occupational Inequality or Interspousal Comparative Advantage?” *Journal of Marriage and Family* 72(2):375–92.

Sullivan, Cath, and Suzan Lewis. 2001. “Home-based Telework, Gender, and the Synchronization of Work and Family: Perspectives of Teleworkers and Their Co-residents.” *Gender, Work & Organization* 8(2):123–45.

Tharp, Derek T., and Elizabeth J. Parks-Stamm. 2021. “Gender Differences in the Intended Use of Parental Leave: Implications for Human Capital Development.” *Journal of Family and Economic Issues* 42(1):47–60.

University of Essex, Institute for Social and Economic Research. (2024). Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 19th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-20.

Van Der Lippe, Tanja, Leonie Van Breeschoten, and Margriet Van Hek. 2019. “Organizational Work–Life Policies and the Gender Wage Gap in European Workplaces.” *Work and Occupations* 46(2):111–48.

Wang, Senhu, and Cheng Cheng. 2024. “Opportunity or Exploitation? A Longitudinal Dyadic Analysis of Flexible Working Arrangements and Gender Household Labor Inequality.” *Social Forces* 102(4):1446–66.

Warren, Tracey. 2004. “Working Part-time: Achieving a Successful ‘Work-life’ Balance? [1].” *The British Journal of Sociology* 55(1):99–122.

West, Candace, and Don H. Zimmerman. 1987. “Doing Gender.” *Gender and Society*

1(2):125–51.

Wheatley, Daniel. 2017. "Employee Satisfaction and Use of Flexible Working Arrangements." *Work, Employment and Society* 31(4):567–85.

Williams, Joan C., Mary Blair-Loy, and Jennifer L. Berdahl. 2013. "Cultural Schemas, Social Class, and the Flexibility Stigma." *Journal of Social Issues* 69(2):209–34.

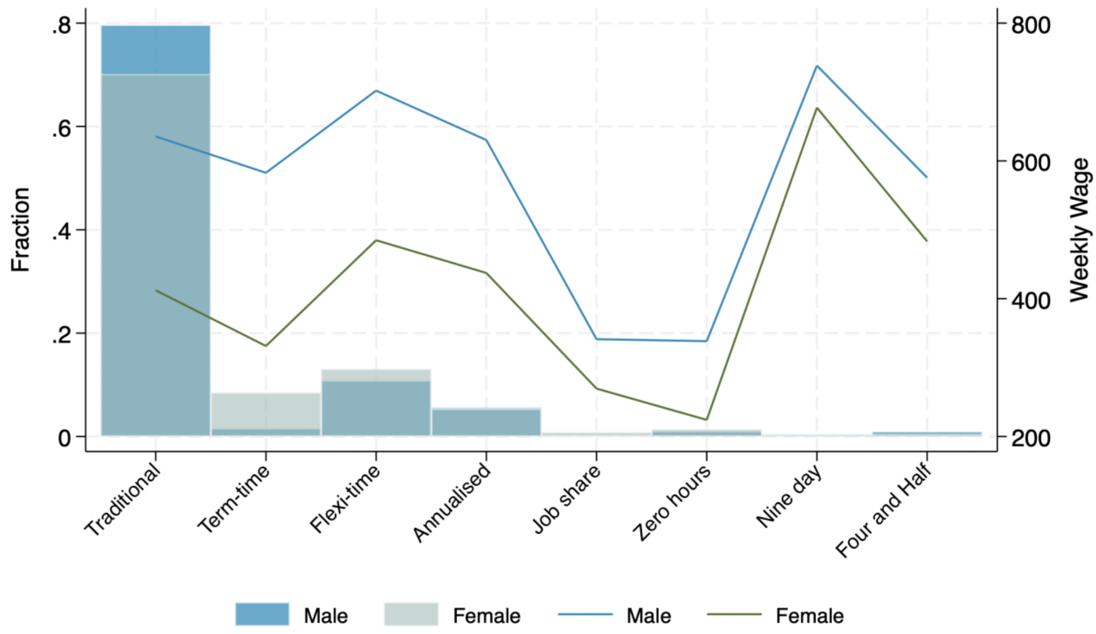
2.8 Appendix

Appendix A: Gender disparities in FWAs within LFS dataset

According to previous literature, there are various types of flexible working arrangements (FWAs), but the proportion of workers using each type varies significantly (Chuang, 2022). To explore these variations, we first analyze data from the UK Labor Force Survey (LFS) from 2000 to 2020. This extensive dataset provides a detailed overview of FWAs in the UK labor market over two decades, allowing us to examine long-term trends and shifts in FWAs adoption. The categorization of FWAs is based on the LFS question, "*In your main job is your agreed working arrangement any of the following*". Respondents could select from eight distinct flexible work categories: term-time, flexi-time, annualized hours, job sharing, zero hours, nine-day fortnight, and a four-and-a-half-day week.

Figure A.2.1 presents the distribution of male and female workers across the eight FWAs and traditional full-time working, alongside average weekly wages for each category. Predominantly, traditional working remains the most common arrangement, constituting 70%-80% of employment for both genders, with an unadjusted weekly wage gap of approximately 33%. In term-time work, a significant gender disparity is observed, with around 9% of women workers engaged in this mode, which is 8 percentage points higher than men. This category also exhibits the largest unadjusted wage gap at 39%, exceeding that of traditional work. Similarly, flexi-time work is common among both women and men, with around 10%. The unadjusted gender wage gap in this category is similar to that in traditional work. Annualized hours are common among both men and women, with 6% in each group. Other FWAs represent a smaller proportion of employment and show smaller unadjusted gender wage gaps, and their limited adoption by both genders restricts further analysis.

Figure A.2. 1: Gender distribution and wage gap in FWAs

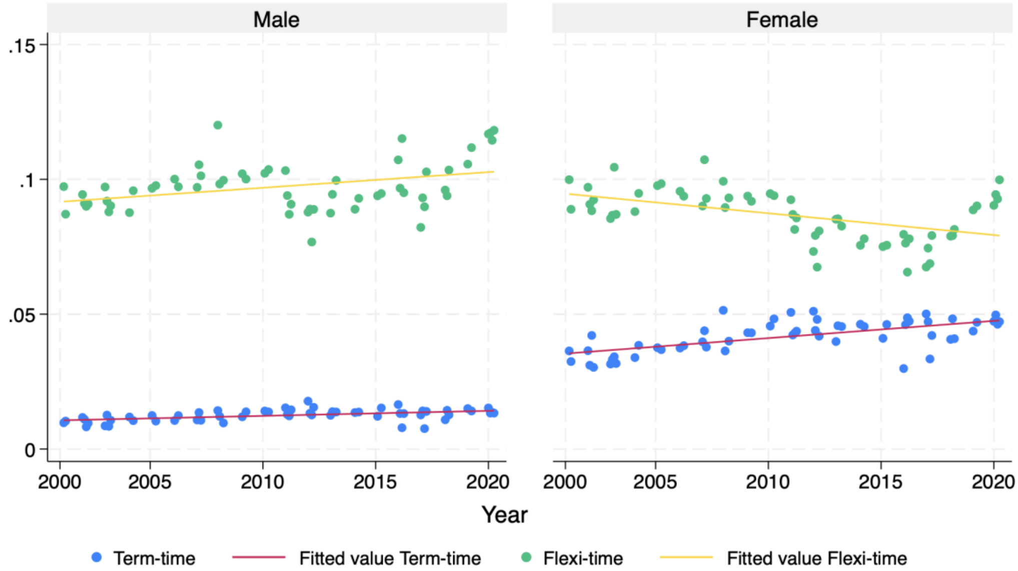


Note: The bars represent the proportion of FWAs in all employment on the left axis, and the lines represent the mean weekly wages on the right axis.

Source: LFS (2000-2020)

Given the significant proportion of both male and female employees utilizing term-time and flexi-time working arrangements, we focus on these two types of FWAs and analyze trends from a longitudinal perspective. Figure A.2.2 reveals a clear gender disparity in the adoption of term-time and flexi-time working arrangements. The left panel shows that over the past two decades, men have increasingly chosen flexi-time work. For term-time work, there is only a gradual increase in adoption among men. Conversely, the right panel illustrates that women have consistently adopted tern-time work. Interestingly, there has been a decline in the number of women opting for flexi-time work. This divergence suggests distinct gender preferences in the types of flexibility sought within the workforce, highlighting the need for a deeper understanding of how these preferences impact gender dynamics in the labor market.

Figure A.2. 2: Gender distribution in Term-time and Flexi-time working (Full-time)



Note: The y-axis represents the proportion of FWAs in all employment.

Source: LFS (2000-2020)

Appendix B: Identification of working time patterns

The UKHLS collects information on both regular and overtime work. However, it provides only limited information on weekend work, recording whether it occurs but not its duration. In the UKHLS, information on weekend work is collected through the question: "*Do you ever work at weekends?*" Respondents can choose from the following options: 1) "Yes - most/every weekend", 2) "*Yes - some weekends*", and 3) "*No weekend working*". We construct a dummy variable to capture whether a worker engages in weekend working, categorizing those who answer "*most/every weekend*" as engaging in weekend work, and all others as not. To address these limitations, we draw to the 2014-2015 Time Use Survey (TUS), which provides detailed information on FWAs and offers a granular record of participants' daily activities in 10-minute intervals. By analyzing this detailed time-use data, we can aggregate the total weekends worked hours for workers who report weekends working and calculate the average weekend working hours by FWAs type, industry, and gender. Additionally, a key consideration is whether workers typically follow a Monday-to-Friday work schedule. This distinction is crucial because employees who work across weekends but maintain a five-day workweek should have their weekend work considered part of their regular working schedule. To address this, we utilize the 2009-2020 Labor Force Survey (LFS), which provides detailed information on the days worked within a week, including whether individuals worked on specific days from Monday to Sunday. We use the LFS to identify occupations that typically follow a Monday-to-Friday schedule. Then, within the UKHLS, if individuals report weekend work and are in occupations classified by the LFS as Monday-to-Friday, we infer that these employees are indeed required to work on weekends.

Table B.2.1, derived from the UKHLS, LFS and TUS, shows the distribution of weekend work and weekend working hours across traditional, term-time (TTW) and flexi-time (FTW) working arrangements, segmented by gender. In traditional working arrangements, the probability of men working on weekend is 29%, compared to 12% for women. Among those who work on weekends, the average work duration is 6.29 hours for men and 6.33 hours for women. Therefore, the expected weekend working hours average about 1.82 hours for men and 0.76 hours for women. Compared to traditional work, TTW involves a less weekend work, with the expected weekend working hours about 0.76 for men and 0.28 for women. Similarly, FTW also involves less weekend work compared to traditional work, with expected weekend working hours of about 0.74 for men and 0.46 for women.

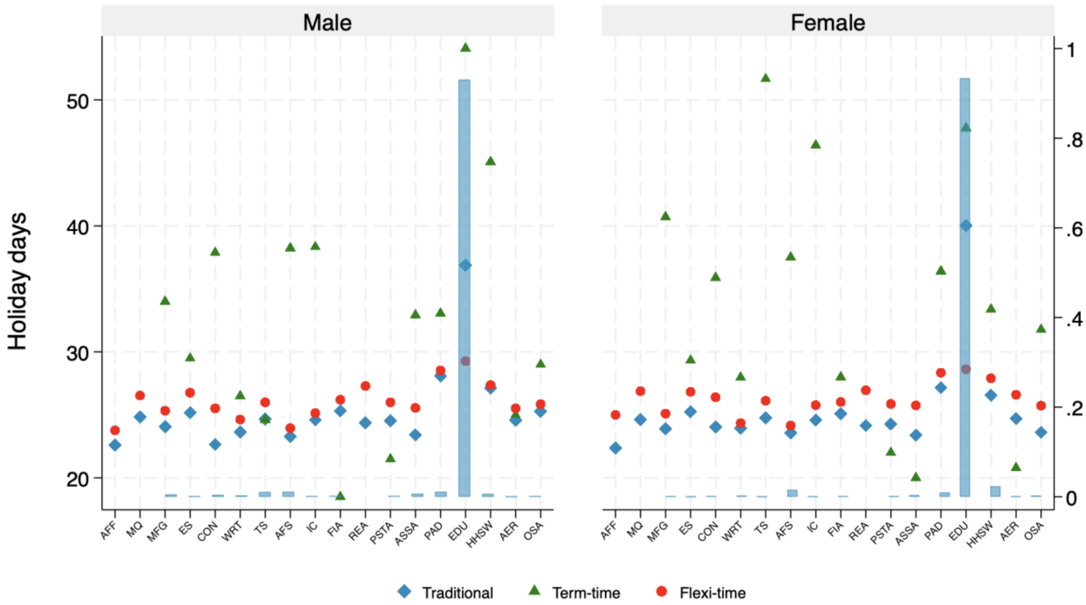
Table B.2. 1: Working hours on weekend (Full-time)

	Full		Male		Female	
	Prob	Hours	Prob	Hours	Prob	Hours
Traditional	0.21	6.30	0.29	6.29	0.12	6.33
Term-time	0.08	4.63	0.19	4.02	0.06	4.70
Flexi-time	0.10	5.76	0.13	5.70	0.08	5.81

Note: "Prob" denotes the probability of engaging in weekend work, and "Hours" represents the duration of weekend working hours.
Source: UKHLS (2010-2020); TUS (2015-2016)

After quantifying the regular, overtime, and weekend working hours, we shift our focus to holiday days, a crucial element in the calculation of working patterns, particularly for TTW. Unfortunately, the UKHLS does not provide data on holiday days, necessitating the use of LFS data (i.e., which capture the number of paid holiday days entitled per year) to fill this gap. Notably, the definition of paid holiday excludes the public holidays. Similarly, we maintain the assumption that the number of paid holiday days remains consistent within each category (e.g., FWAs, industry, occupation and gender), enabling us to link the information of LFS data to the UKHLS. The results are shown in Figure B.2.1. Both men and women, engaged in TTW have a more paid annual holiday days compared to traditional working arrangement and FTW, particularly in the Education (EDU) industry. Since TTW is predominantly clustered within the Education industry, accounting for about 95% of such arrangements, we can further compare the paid holiday days between the non-education and education industries.

Figure B.2. 1: Annual Holiday days by industries, FWAs and gender



Note: The x-axis displays the different industries, the left y-axis represents the paid holidays days, which excludes the public holidays days. The right-axis refers to the proportion of term-time working across the industries. Industry codes are provided in Appendix L.

Source: UKHLS (2010-2020); LFS (2010-2020)

As shown in Appendix Table B.2.2, workers in the non-education industry receive fewer paid holidays, averaging 34 days, compared to 49 days for those in the education industry. Within the education industry, jobs are further categorized into teaching and non-teaching roles. Teaching occupations typically have more holiday days, averaging 56 days, compared to non-teaching roles, which averaging 41 days. Within teaching occupations, primary and secondary teachers have the most holiday days, averaging 56-58 days, while tertiary teaching roles have fewer holiday days, averaging 40 days. Notably, no difference is observed between male and female teachers in terms of holiday days.

Table B.2. 2: Holiday days per year in the education industry by gender

	No-Edu	Edu					
	Tot	Tot	No-Tea	Tea	Pri	Sec	Ter
Total	34	49	41	56	56	58	40
Male	33	54	45	58	58	59	40
Female	34	48	40	55	56	58	41

Note: "Tot" represents the total sample; "Tea" denotes the teaching occupation, and "No-Tea" refers to the non-teaching occupation; "Pri", "Sec", and "Ter" are the primary, secondary, and tertiary education level in teaching occupation.

Source: LFS (2010-2020)

Finally, to comprehensively understand the impact of FWAs on work patterns and accurately assess the potential wage penalties associated with these transitions, we calculate the total working hours (TWH). The TWH is computed by summing up the regular (R), overtime (O), and weekend working hours (W). This sum is then adjusted according to the proportion of the year the individual is actually working:

$$TWH = (R + O + W) \times \left(1 - \frac{H}{365}\right) \tag{B.1}$$

As displayed in Table B.2.3, men working under FTW have the lowest total working hours, without adjusting for holiday days, at 171 hours, while women working under TTW record the lowest total working hours at 134 hours. After adjusting for holiday days, TTW continues to be the arrangement with the lowest working hours for women. In contrast, for men, the total working hours in TTW and FTW become comparable. This suggests that men may choose FTW for reasons other than reducing working hours.

Nonetheless, a key consideration in calculating total working hours is avoiding double-counting weekend hours. For instance, if participants in the UKHLS report weekend work hours as part of either regular or overtime hours, the total working hours may be overestimated. To verify this, we refer to the LFS to conduct a comparative analysis, where the explicitly asks for the "*Thinking now about the seven days ending Sunday the …, how many hours did you actually work in your (main) job/business –*

please exclude meal breaks?". In the LFS, the mean value of full-time workers' total monthly working hours is 175. In contrast, the UKHLS reports an average of only 169 hours when weekend hours are excluded, However, when we manually increase the weekend working hours in UKHLS, the average rises to approximately 174 hours (see Table B.2.3), closely aligning with the result in LFS. This comparison strongly suggest that, in UKHLS, the workers are do not account for weekend work in their overtime work. As a result, omitting weekend working hours leads to an underestimation of total working hours, which could bias our estimation of the transition effect on working time patterns.

Table B.2. 3: Total monthly working hours (Full-time)

	Full		Men		Women	
	TWH	THW-Ad	TWH	THW-Ad	TWH	THW-Ad
Traditional	183	170	191	179	174	162
Term-time	140	123	183	157	134	118
Flexi-time	159	148	171	159	149	138
Mean	174	160	187	174	162	150

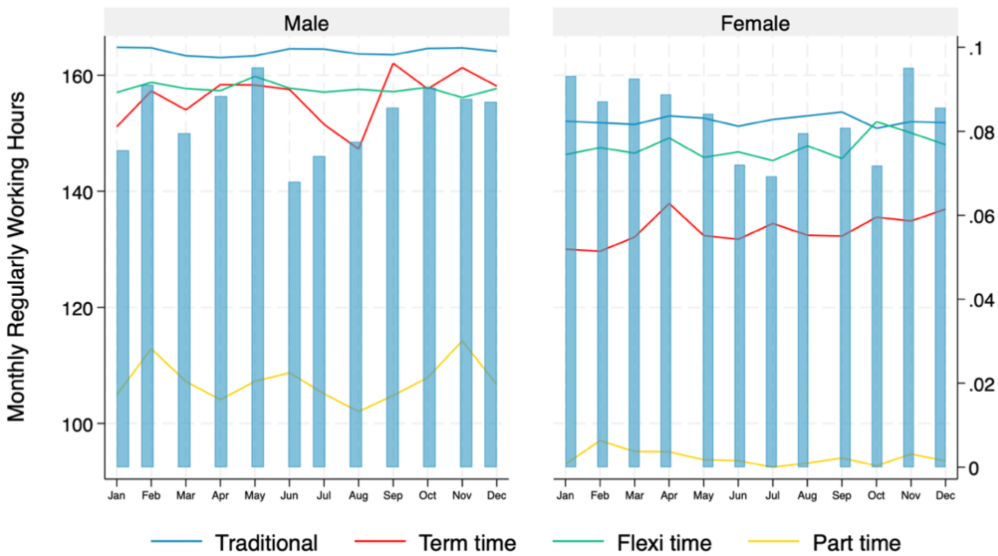
Note: "TWH" denotes the total monthly working hours without adjustment for holiday days, and "THW-Ad" represents the total monthly working hours with adjusted for holiday days.

Source: UKHLS (2010-2020); LFS (2010-2020); TUS (2015-2016)

Appendix C: Further details on seasonal fixed effects

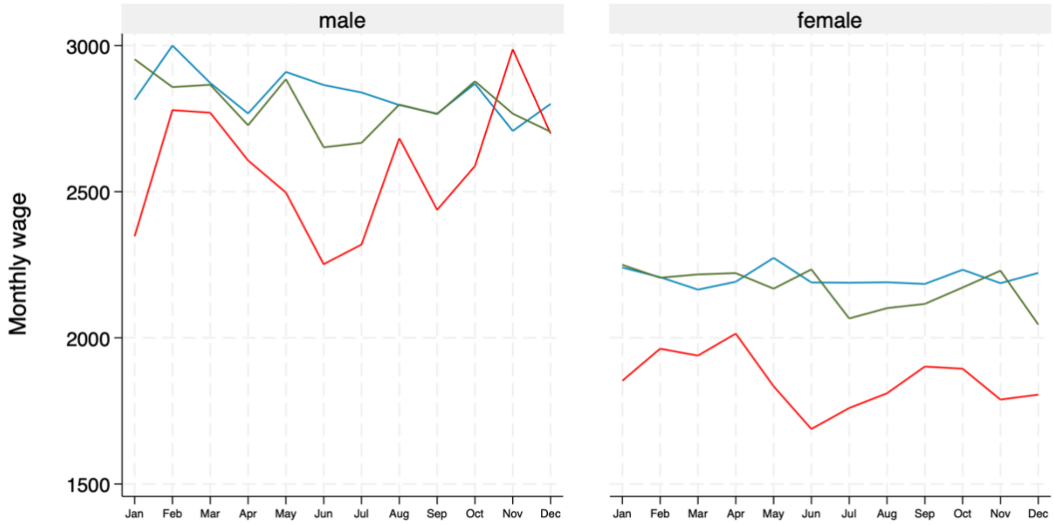
This section provides a detailed analysis of fluctuations in monthly working hours and wages across various working arrangements (i.e., traditional work, term-time work, flexi-time work, and part-time work) over the course of different months. The analysis is based on data from the UKHLS, which includes participant information collected on a monthly basis.. As shown in Figure C.2.1 and Figure C.2.2, both male and female participants exhibit a distinct cyclical pattern in working hours and monthly wages, but only among those engaged in term-time work (TTW).

Figure C.2. 1: Working hours in a year



Note: On the x-axis, we display the interview month for each participant. The left y-axis represents the monthly regular working hours, which excludes overtime and weekend work. Meanwhile, the right y-axis indicates the proportion of participants interviewed in each respective month.

Figure C.2. 2: Monthly wage in a year



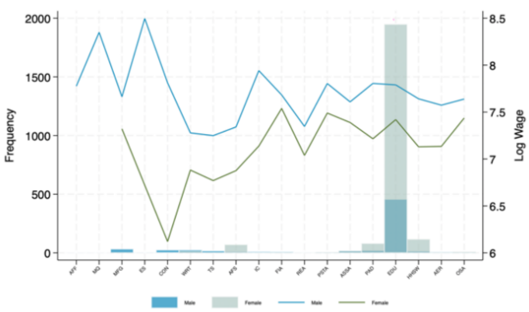
Note: On the x-axis, we display the interview month for each participant, and the y-axis represents the monthly wage.

Appendix D: Distribution of TTW and FTW across Industries

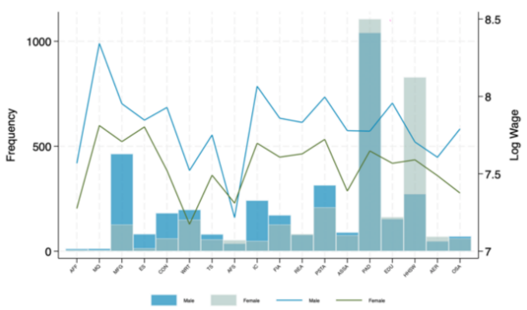
As Figure D.2.1 displayed, TTW are predominantly observed within specific industry clusters, notably in education. Moreover, the line graph, which plots the natural logarithm of wages, shows that the wage disparity in the education industry is relatively narrower compared to other industries. In terms of FTW, both men and women cluster in the Public Administration and Defense (PAD), and the frequency of women is higher than that of men. This indicates that public sector is more likely to offer more "family-friendly" policies than other industries.

Figure D.2. 1: Gender distribution and wage gap in Term-time and Flexi-time work by industry

a) TTW



b) FTW



Note: The figures provide a dual representation of gender distribution and wage disparities across various industries in TTW and FTW. The bar chart illustrates the frequency of men (in blue) and women (in green) within each industry, while the line graph depicts the natural logarithm of wages for each gender various industries.

Appendix E: Attrition problem of Gender Disparities in FWA Patterns

In this section, we first provide a description of the percentage of participants who dropped out (attrition) in each year. Next, we perform multiple imputation as a robustness check to assess whether attrition introduces bias into our baseline model presented in Figure 2.1.

E.1. Statistical Overview of Attrition rates

The Table E.2.1 shows attrition rates across waves in a longitudinal study, starting at 0% in 2012 and rising to 15.1% in 2014, 25.7% in 2016, 23.6% in 2018, and 25.9% in 2020. The overall attrition rate is 17.7%, reflecting cumulative dropout over time. While attrition remains low in early years, its steady increase raises concerns about potential attrition bias in later waves if the attrition is systematic.

Table E.2. 1: Statistical description of attrition probability

	Full	2012	2014	2016	2018	2020
Attrition	17.7%	0%	15.1%	25.7%	23.6%	25.9%

Note: Attrition rate refers to the probability of participants dropping out of the UKHLS in the next wave.

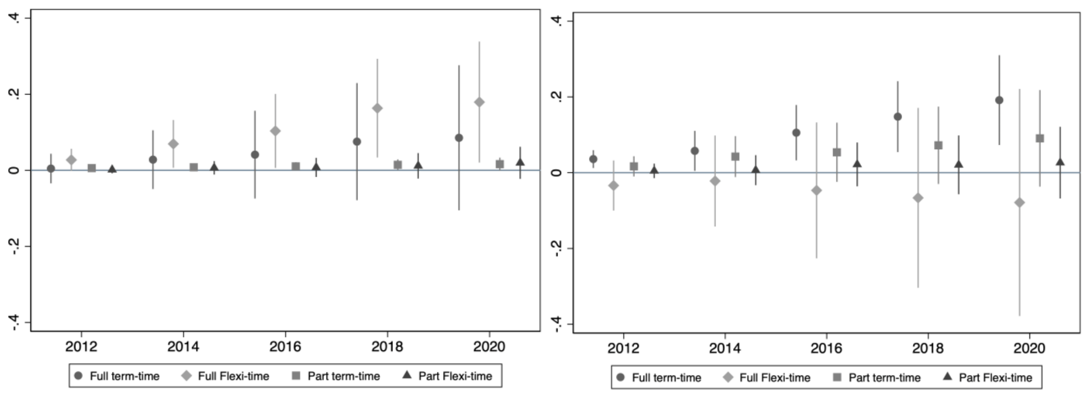
Source: UKHLS (2010-2020)

E.2. Robustness check for the gender disparity patterns in FWAs

Figure E.2.1 replicates our baseline regression from Figure 2.1 using multiple imputation (MI) to address the attrition issue. Specifically, we first generate the missing observations for participants who dropped out of the dataset and then apply MI to re-estimate the baseline model. In general, the results in Figure E.2.1 are consistent with those in Figure 2.1, suggesting that attrition does not drive our findings.

Figure E.2. 1: Robustness check for the transition of Women's to TTW and Men's to FTW from 2010 to 2020

(a) Men (b) Women



Note: We conduct linear regression analyses to examine job transitions from 2010 to 2020 by using Multiple Imputation to fill the missing waves for each individual, with a focus on gender-specific patterns. The figure focuses solely on transitions to term-time and flexi-time working arrangements, categorized into part-time and full-time categories. The control variables included age, age square, marital status, number of children, family income, and firm size, with confidence intervals set at the 90% level. Standard errors are clustered at the time-industry and time-occupation levels. The x-axis represents the years from 2012 to 2020 with 2010 serving as the reference year. The y-axis indicates the probability of an employee transitioning from traditional working arrangements to various types of FWAs compared to the year of 2010.

Appendix F: Impact of Transitioning to Part-time Working Arrangements on Total working hours and Monthly wage

In this section, we first present the statistical description of total working hours and monthly wages across different FWAs. Next, we examine the composition effect of transitioning from full-time traditional work to part-time TTW and FTW on total working hours and wage outcomes for both men and women. We find that the effects of part-time work generally overshadow those of part-time TTW, while the impact of FTW often outweighs that of part-time FTW.

F.1. Statistical Overview of Total Working Hours and Monthly Wages across Various FWAs

Table F.2.1 highlights the differences in total monthly working hours and monthly wages across various FWAs. Both men and women in part-time traditional work have similar total working hours and monthly wages similar to those in part-time TTW. In contrast, both men and women in part-time FTW exhibit total working hours and monthly wages comparable to those in full-time FTW.

Table F.2. 1: Total monthly working hours (Full-time)

	Part-time Trad		Part-time TTW		Part-time FTW		Full-time FTW	
	Men	Women	Men	Women	Men	Women	Men	Women
TWH	106	97	121	108	172	148	174	161
MW	1,253	1,121	1,503	1,306	2,633	1,878	2,676	2,049

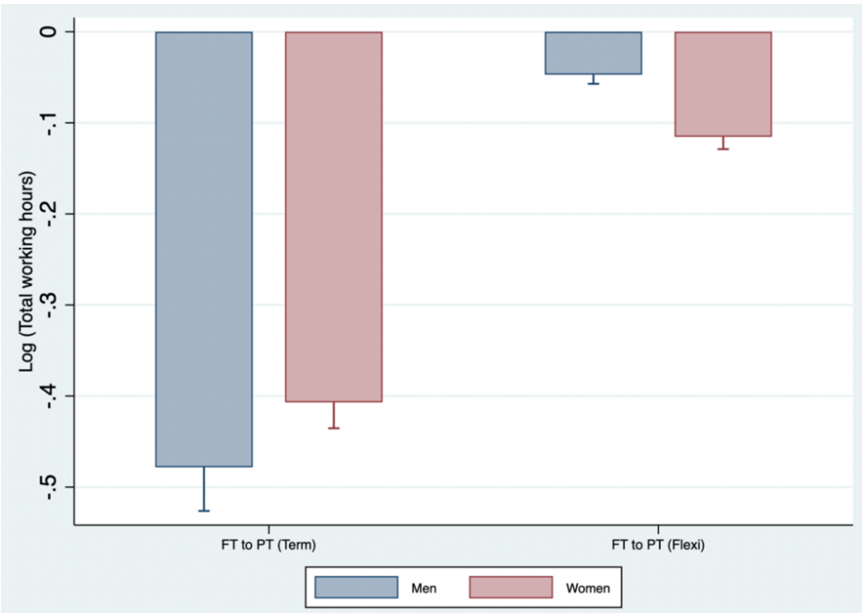
Note: "TWH" denotes the total monthly working hours with adjusted holiday days, as defined in Formula (2.2), and "MW" refers to the monthly wage.

Source: UKHLS (2010-2020)

F.2. The effect of transitions to part-time FWAs on total working hours

As illustrated in Figure F.2.1, when full-time employees transition from traditional work to part-time TTW, there is a decrease of 45% ($p<0.01$) for men and 40% ($p<0.01$) for women in total working hours. These transitions lead to a substantial reduction in working hours compared to other types of transitions. Combining this with the evidence that transitioning to full-time TTW only reduces working hours by 13.3% for men and 18.9% for women (see Table 2.2), it can be inferred that the effects of part-time work generally overshadow those of part-time TTW. For transitions from full-time traditional work to part-time FTW, the reduction in total working hours is relatively modest, amounting to only about 5% ($p<0.01$) for men and 10% ($p<0.01$) for women. This level of reduction is similar to the transition from full-time traditional work to full-time FTW, indicating that transitioning to part-time FTW does not significantly reduce working hours as much as transitioning to other part-time arrangements does.

Figure F.2. 1: Impact of Transitioning to Part-time Working Arrangements on Total Working Hours

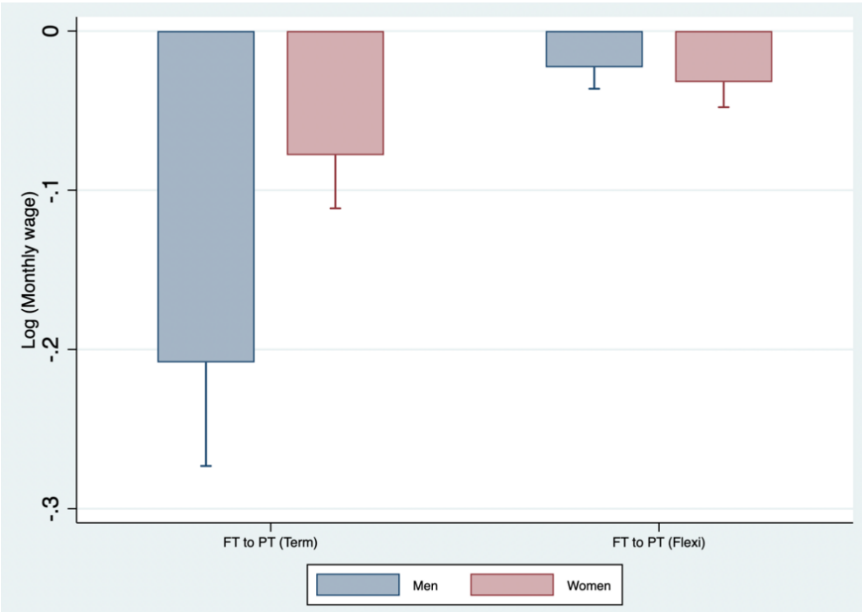


Notes: This figure presents estimates illustrating the impact of employee transitions from traditional work to part-time TTW and FTW on total working hours by gender. Total working hours are calculated according to Equation 2.3. The control variables included age, age square, marital status, number of children, family income, and firm size. The y-axis represents the percentage change in total working hours when employees transition from traditional working arrangements to PTW, TTW, or FTW.

F.3. The effect of transitions to part-time FWAs on monthly wage

As illustrated in Figure F.2.3, and consistent with the effect on the total working hours, when full-time employees transition from traditional work to part-time TTW, there is a substantial decrease in their monthly wage. Specifically, the reductions is approximately 20% ($p<0.01$) for men and 8% ($p<0.01$) and women. For transitions from full-time traditional work to part-time FTW, the reduction in monthly wage is relatively modest, amounting to only about 2% ($p<0.01$) for men and 3% ($p<0.01$) for women.

Figure F.2. 2: Impact of Transitioning to Part-time Working Arrangements on Monthly wage



Notes: This figure presents estimates of the impact of employee transitions from traditional work to part-time TTW and FTW on monthly wages by gender. The control variables included age, age square, marital status, number of children, family income, and firm size. The y-axis represents the percentage change in total working hours when employees transition from traditional working arrangements to PTW, TTW, or FTW.

Appendix G: Short-run and Long-run Impacts of Transition on Outcomes

Beyond estimating the overall average treatment effect (ATE), we are also interested in the immediate and long-term impacts on the pre- and post-transition dynamics. To clearly capture these dynamics effects, we employ an event-study Difference-in-Differences (DiD) model, which allows us to estimate the treatment effects at various time points relative to the transition. This approach enables the delineation of a timeline around the transition event, where periods prior to the transition are coded as negative values (e.g., -5+ to -1), the transition point is coded as 0, and post-transition periods are coded as positive values (e.g., 1 and 4+).[61,62] In general, the event-study DiD regressions for $k \in \{TTW, FTW\}$ can be formulated as follows:

$$y_{itao} = \sum_{j=-5, j\neq -1}^{4} \delta_j \, D_{itao,j}^{k} + \boldsymbol{X}'_{itao}\eta + \rho_i + \varphi_t + \gamma_a + \mu_o + \varepsilon_{itao} \qquad \text{(G.1)}$$

where y_{itao} represents the outcome of interest for employee i in industry a and occupation o at time t, namely total working hour and monthly wage. We estimate the regression separately for the gender-specific sub-samples, where $g \in \{w, m\}$. $D_{itao,j}^{k}$ is a year-specific treatment indicator, capturing the transition effect from traditional work to TTW or FTW for each year j. δ_j are the coefficients of main interest, as they identify the effect of transitioning from traditional to TTW or FTW on labor outcomes, relative to the matched counterfactual group. We omit the year prior to the event ($j = -1$), which means that all estimates of δ_j are relative to the year before the transition.

[61] For any periods prior to the transition that are earlier than at -5, we code as -5, and for any periods beyond 4 after the transition, we code them as 4+.

[62] However, using this approach presents a significant challenge. The UKHLS questionnaire on FWAs is administered biennially, which may lead to massing data on job transitions in the years when the questions were not asked. To address this issue, we extend our analysis to include all available waves of the UKHLS, not just those containing FWAs waves. We bridge the biennial gaps by using on responses to question about whether an individual's employer, occupation and industry remained the same as in the previous wave. To ensure the robustness of this method, we report the overall transition effects based on Formula 2.3.

t indicates the year in which working transition occurred.

After examining overall transition effects, we also explore the short-term and long-term impacts of transitions on total working hours and monthly wages. The advantage of TTW is that it allows employees to align their work schedules with school calendars, and this arrangement is often driven by the needs of employees with children (Chung and van der Horst, 2017). This arrangement may have a significant impact on the labor outcomes in the short term, but as children grow older and caregiving demands decrease, long-term labor outcomes may improve. Figure G.2.1 displays the event-study coefficients derived from Formula G.1, which are obtained by interacting all independent variables with gender dummies. Moving forward, we will focus exclusively on the transition effects of TTW and FTW to better capture gender-divergent patterns.

As displayed in Panel (a), in the short-term (the first to third years) following the post-transition from traditional work to TTW, both men and women experience a reduction in their total working hours, but their working hours patterns differ in the long run (beyond the third year). Specifically, at $t = 0$, both men and women reduce their total working hours compared to pre-transition levels ($t = -1$). However, beyond the third year, only women's total working hours generally recover to pre-transition levels, while men's hours remain reduced. One potential reason for this pattern is that women often transition to TTW due to childbirth. As shown in Figure G.2.2 in Appendix G, there is evidence indicating that women are 15 percentage points ($p<0.01$) more likely to have an additional child in the first year following their transition. Additionally, as shown in Figure G.2.3 in Appendix G, the recovery pattern post-transition is primarily driven by mothers, whereas non-mothers who transition to TTW see their total working hours remain constant over time. These patterns are aligned with our hypothesis that TTW is primarily driven by women seeking to align their work schedules with their children's calendars. For men, the transition to TTW may not be related to fertility behavior (also see Figure G.2.1. in Appendix G), and the reduction in total working hours remains constant over time, at about 12% (p<0.01) from the second to fourth years post-transition ($t = 1$ to $t = 4 +$). For those transitioning to FTW (see Figure B in Panel A), both men (3%; p<0.01) and women (6%; p<0.01) experience a reduction in total working hours, both in the short term and long term. However, given the limited effect of the transition on reducing working time, FTW does not appear to significantly increase the housework hours (see Figure 3).

In terms of wage penalties, figure c in Panel B shows that in the first year after the transition ($t = 0$), women (10%; p<0.01) experience a reduction in monthly wage twice as large as that of men (5%; p<0.01) when transitioning to TTW, compared to pre-transition levels ($t = -1$). However, after one-year post-transition, the reduction in monthly wages for women decreases to 5% (p<0.01), which is similar to the reduction experienced by men. These results align with those from the working time analysis, showing that women tend to recover their total working hours beyond one-year post-transition (see Panel A). Interestingly, in the long term (after four years post-transition), there is no significantly observable wage penalty for women transitioning to TTW, and their monthly wages nearly recover to pre-transition levels, with only an average reduction of 2.5% (p<0.05). This wage pattern is similar for transitions to FTW. In the first year of transition, the wage penalty for women is twice as high as that for men, with a reduction of 4% (p<0.01). However, after three years post-transition, the wage penalty for women falls to 2.5%, while for men, the wage penalty remains constant and even increases to 5% after four years.

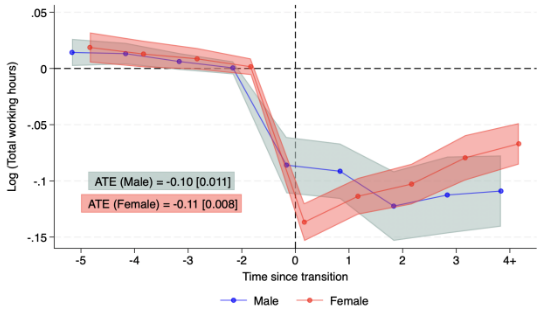
Overall, we observe a significant gender wage gap in the first-year post-transition, with women facing larger wage penalties when transitioning to TTW compared to men. However, this gender wage gap decreases over time, with women experiencing a smaller wage penalty in the long-term, particularly after three-years post-transition ($t = 2 +$). This may be because, upon transitioning to TTW, women initially reduce their total working hours in the short-term, often to accommodate having an additional child.

However, in the long-term, their total working hours nearly recover to pre-transition levels. In contrast, men transitioning to TTW experience a substantial reduction in total working hours that remains constant in the long-term. Moreover, for those transitioning to FTW, neither men nor women experience a significant reduction in total working hours and monthly wages.

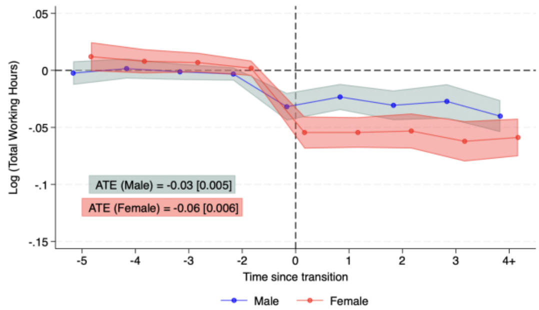
Figure G.2.1: Short-run and Long-run Impacts of Transition on Outcomes by Gender

A. Total working hours

a) TTW

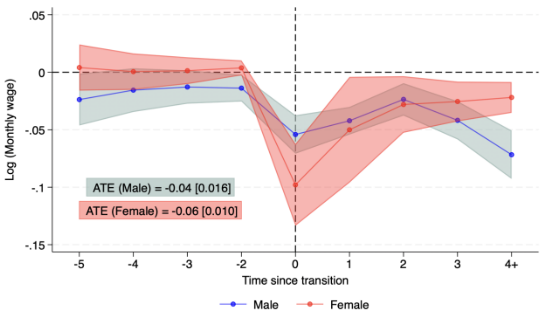


b) FTW

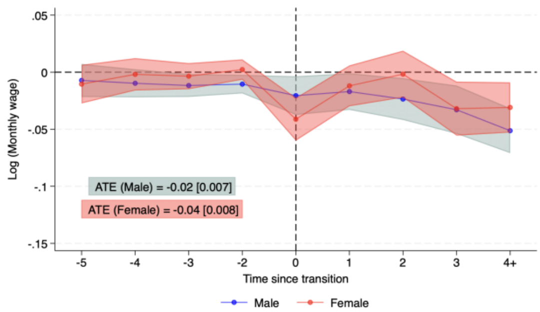


B. Monthly wage

c) TTW



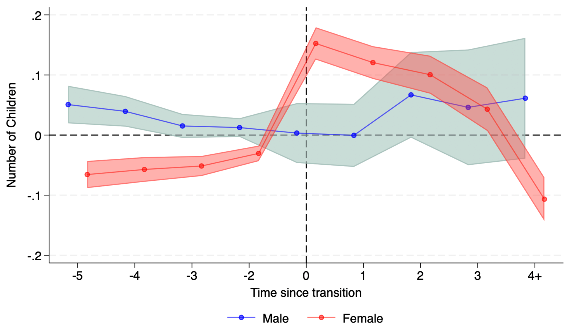
d) FTW



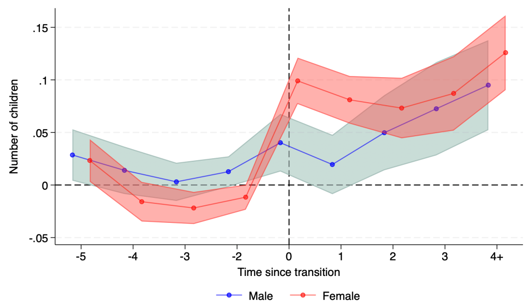
Notes: Each figure reports the estimates of the transition from traditional work to TTW and FTW on total working hours (Panel A) and monthly wages (Panel B). The estimates use the matched control to identify effects 5 years before and 5 years after a transition (see Equation 2.4). The control variables included age, age square, marital status, number of children, family income, and firm size. The y-axis represents the percentage change in monthly wage or total working hours when employees transition from traditional working arrangements to PTW, TTW, or FTW. The total number of observations becomes 23,327 for males and 28,957 for females, as we bridge the biennial gaps in FWAs information by relying on responses to whether occupation and industry codes accurately describe the participants' main jobs. The overall ATE reported in each figure aligns with Table 3, showing that the robustness of imputation method.

Figure G.2.2: Impact of Transitioning from Traditional to Term-time and Flexi-Time Working Arrangements on Number of children

a) Term-time working



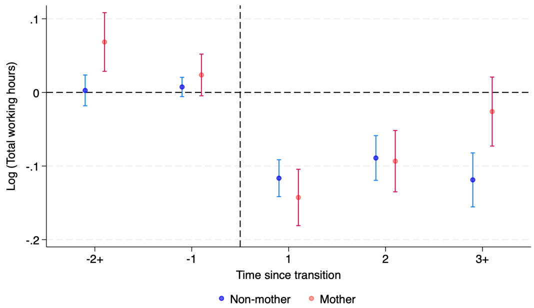
b) Flexi-time working



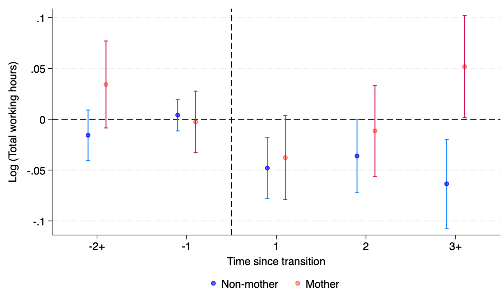
Notes: Each figure reports the effect of transition from traditional working to TTW and FTW on number of children. The estimates use the matched control to identify effects 5 years before and 5 years after a transition (see Equation 5). The complete set of controls described in equation (5) is included but not reported. The total number of observations is 23,327 for males and 28,957 for females, as we bridge the biennial gaps in FWAs information by relying on responses to whether occupation and industry codes accurately describe the participants' main jobs.

Figure G.2.3: Short-run and Long-run Impact of Transition on Outcomes by Mother

a) Total working hours



b) Monthly wage



Notes: Each figure reports the IV estimates of a transition from traditional working to TTW and FTW on total working hours (Panel A) and monthly wage (Panel B). The estimates use the matched control to identify effects 5 years before and 5 years after a transition (see Equation 5). The complete set of controls described in equation (5) is included but not reported. The total number of observations is 23,327 for males and 28,957 for females, as we bridge the biennial gaps in FWAs information by relying on responses to whether occupation and industry codes accurately describe the participants' main jobs. The overall ATE reports in each figure align to Table 3, showing that the imputation method is robustness.

Appendix H: Transition Effects on Occupational Downgrading

This section shows that when women transition to TTW, they may face career downgrading or stagnation. As Table I.2.1 displayed, when women transition to TTW, there is a 0.715 log-odds ($p<0.01$) increase in the likelihood of moving from high to low skilled roles and a 1.099 log-odds ($p<0.01$) increase in the likelihood of moving from middle to low skill roles.

Table H.2. 1: Transition Effects on Occupational Downgrading by Gender

	High → Low		Mid → Low	
	Men	Women	Men	Women
	(1)	(2)	(3)	(4)
TTW	1.330**	0.715***	2.121**	1.099***
	(0.592)	(0.264)	(0.911)	(0.302)
FTW	0.105	-0.188	-0.009	-0.179
	(0.248)	(0.231)	(0.290)	(0.240)
N	2,208	3,330	2,208	3,330
I & T & O FE	Yes	Yes	Yes	Yes
P&A, seasonal	Yes	Yes	Yes	Yes
Personal controls	Yes	Yes	Yes	Yes

NOTE.—Estimates are derived from regressions on the effect of transitioning to term-time or flexi-time working on the probability of moving from high and middle to low occupation levels by multinomial logit models. The coefficients report on Logit levels. The control variables included age, age square, marital status, number of children, family income, and firm size. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

Appendix I: Transition effects on total working hour and monthly wage

I.1. The Effect of Transition on Monthly Wages, Controlling for Total Working Hours

As shown in Appendix Table I.2.1, controlling only for regular and overtime working hours underestimate the transition effect on wage outcomes in both TTW and FTW, as they typically involve fewer weekend working hours and more paid holiday days.

Table I.2. 1: The effect of transition on monthly wage

	Baseline model	Plus regular, overtime	Plus total working hours
	(1)	(2)	(3)
TTW	-0.148***	-0.052***	-0.066***
	(0.010)	(0.010)	(0.010)
FTW	-0.061***	-0.007	-0.027***
	(0.007)	(0.007)	(0.007)
N		41,012	
Basic controls		Yes	
I & T & O FE		Yes	
P&A, seasonal		Yes	

NOTE.—The table shows effects of transiting from traditional to TTW and FTW on monthly wage among full-time employees (without IV). The control variables included age, age square, marital status, number of children, family income, and firm size. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

I.2. The Effect of Transition on Working Hours and Monthly Wages, Controlling for Working Autonomy Factor

There are five autonomy questions in the UKHLS: "*Autonomy over job tasks, work pace, work manner, task order, and work hours*". As shown in Table I.2.2, whether we control for each of these factors individually or use factor analysis to construct a composite factor representing overall work autonomy, the results remain consistent.

Table I.2. 2: The effect of transition on total working hours and monthly wage

	Total working hour			Monthly wage		
	Baseline model	Plus Industry FE	Plus work autonomy factor	Baseline model	Plus Industry FE	Plus work autonomy factor
	(1)	(2)	(3)	(4)	(5)	(6)
TTW	-0.123^{***}	-0.121^{***}	-0.123^{***}	-0.056^{***}	-0.055^{***}	-0.056^{***}
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
FTW	-0.042^{***}	-0.041^{***}	-0.041^{***}	-0.022^{***}	-0.023^{***}	-0.022^{***}
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
N		28,301			28,301	
Basic controls		Yes			Yes	
I & T & O FE		Yes			Yes	
P&A, seasonal		Yes			Yes	

NOTE.—The table shows effects of transiting from traditional to TTW and FTW on total working hours and monthly wage among full-time employees (without IV). The control variables included age, age square, marital status, number of children, family income, and firm size. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Table I.2. 3: The effect of transition on hourly wage

	Men		Women	
	FE	SSIV	FE	SSIV
	(1)	(2)	(4)	(5)
TTW	0.032	-0.049^{**}	0.036^{**}	-0.035^{***}
	(0.026)	(0.023)	(0.014)	(0.013)
FTW	0.003	-0.015	0.013	-0.026^{**}
	(0.011)	(0.010)	(0.011)	(0.011)
N	12,646		16,059	
Basic controls	Yes		Yes	
I & T & O FE	Yes		Yes	
P&A, seasonal	Yes		Yes	

NOTE.—The table shows effects of transiting from traditional to TTW and FTW on hourly wage among full-time employees (without IV). The hourly wage is calculated by dividing the monthly wage by the total working hours. The control variables included age, age square, marital status, number of children, family income, and firm size. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Appendix J: Industry codes

AFF: Agriculture, Forestry and Fishing

MQ: Mining and Quarrying

MFG: Manufacturing

ES: Energy Supply

CON: Construction

WRT: Wholesale and retail trade

TS: Transportation and storage

AFS: Accommodation and food service activities

IC: Information and communication

FIA: Financial and insurance activities

REA: Real estate activities

PSTA: Professional, scientific and technical activities

ASSA: Administrative and support service activities

PAD: Public administration and defense; compulsory social security

EDU: Education

HHSW: Human health and social work activities

AER: Arts, entertainment and recreation

OSA: Other service activities

Appendix K: SSIV identification and results

The SSIV formula is defined as:

$$SSIV_{tra}^{k} = s_{(t-1)ra}^{k} \times \left(1 + \Delta FWA_{tr}^{k}\right) \quad \text{(K.1)}$$

where $s_{(t-1)ra}^{k} = \sum_i s_{ira}^{k}$ (i.e., $k \in \{TTW, FTW\}$) is the share component, capturing the FWAs adoption in industry a within region r in the previous year. The shift component, $\Delta FWA_{tr}^{k} = \frac{s_{tr}^{k} - s_{(t-1)r}^{k}}{s_{(t-1)r}^{k}}$ (i.e., $s_{tr}^{k} = \sum_a \sum_i s_{itra}^{k}$), reflects the overall growth rate of FWA adoption in each region.

The effectiveness of the shift-share IV relies on two key conditions: relevance and the exclusion restriction. To support the exclusion restriction of our SSIV in estimating the causal effect of FWAs on individual labor outcomes, it is essential to consider both the shift and share components separately. On the one hand, the shift component, which represents the region-level trend in FWAs adoption, captures a broad, exogenous increase in FWAs that is expected to influence regional FWAs trends uniformly rather than respond to specific individual characteristics or localized wage determinants. This shift is typically driven by factors such as national policy changes that promote FWAs across regions and occupations, reducing the likelihood that these trends are influenced by individual-level factors. On the other hand, the share component reflects the level of FWAs adoption across industries within each region in previous year. Given this, it is unlikely that the historical distribution of FWAs adoption within each region by industry composition would directly affect individual-level current labor outcomes.[63]

Relevance can be assessed through both theoretical and empirical approaches. On the theoretical perspective, the combination of industry structure within each region

[63] Given that there are over 200 industry-region combinations, and the variation across these groups is substantial.

and broad FWAs trends serves as a strong predictor of individual FWAs adoption, as these factors capture how each region's workforce is positioned for FWAs adoption under an exogenous trend. On the empirical side, we perform a first-stage test to formally evaluate the instrument's relevance. Nonetheless, to further strengthen the robustness of our SSIV, we use the supply of FWAs rather than actual adoption (i.e., equilibrium condition).[64] This approach reinforces the exclusion restriction, as the regional growth rate in FWA supply and the historical distribution of FWA supply are even less likely to directly impact individual working hours and current wages. Additionally, since the shift-share variables are derived from the UKHLS, the sample size is relatively smaller, although it remains representative. For an additional robustness check, we use the LFS, which covers the entire population in UK. In this approach, we construct the shift-share instrument variable using the LFS and then match it back to our primary dataset, the UKHLS. By obtaining the instrument variable from a separate dataset, we employ a two-sample instrumental variable (TSIV) method to obtain more reliable standard errors (Angrist and Krueger, 1992; Inoue and Solon, 2010; Thomas et al., 2016).

Moreover, given that both of our potential endogenous variables, namely TTW and FTW are dummies, applying a linear IV model may biased. Hence, it would be better to consider the control function (CF) method to obtain a more efficient estimator (see Wooldridge, 2015). Overall, in the first-stage regression, two Logit models predict R_{its}^{j} for job-finding analysis:[65]

$$TRAN_{itao}^{j} = \alpha_0 + \sum_{k} \zeta_k SSIV_{tra}^{k} + \boldsymbol{X}_{itso}'\eta + \rho_i + \varphi_t + \sigma_a + \mu_o + \varepsilon_{itao} \qquad \text{(K.2)}$$

[64] In the UKHLS, in addition to gathering information on FWA adoption through the question "*Do you currently work in any of the following FWAs?*" (multiple choices), the survey also collects data on FWA supply (availability) at the workplace. Respondents are asked, "*I would like to ask about working arrangements at your place of work. If you personally needed any, which of the following arrangements are available at your workplace?*" (multiple choices).

[65] Logit is preferred over Probit in fixed effects models because the conditional likelihood approach in Logit avoids the incidental parameters problem, enabling consistent estimation even with short panel data.

Following this, we calculate the inverse Mills ratios imr_{itao}^{k} and incorporate them into the primary regression model (3) to address endogeneity problem.[66]

Empirical results

This section conducts a robustness check to assess whether our baseline model suffers from endogeneity issues and to confirm that the baseline model in the main text demonstrates the causal effect of transitioning from traditional work arrangements to TTW and FTW on total working hours and monthly wages. As shown in Panel A of Table K.2.1, transitioning from traditional work to TTW leads to a significant reduction of 18.7% (p<0.01) in total working hours, while transitioning to FTW results in a smaller decrease of 7.8% (p<0.01), after controlling for basic variables along with year, individual, and occupation fixed effects (see column 1). This pattern in working hours aligns with the observed transition wage penalty. In Panel B, our FE model shows that transitioning to TTW incurs a substantial wage penalty of 6.5% (p<0.01), whereas transitioning to FTW results in a much smaller penalty of 2.7% (p<0.01).[67] Given potential bias may arise from unobserved variables related to firm characteristics, we employ the SSIV method to address this endogeneity issue. The SSIV estimates presented in column 2 reveal that, after correcting for endogeneity, the transition effect of both TTW and FTW on total working hours and monthly wages remains largely unchanged. Importantly, the significance of the inverse Mills ratios (imr_TTW and imr_FTW) in both Panels A and B suggests potential endogeneity in the working time and wage analyses. However, despite the presence of endogeneity, the coefficients of interest remain relatively stable, suggesting that any bias introduced is minimal. These results are robust, as using the supply of FWAs to construct the SSIV and applying the

[66] In logistic regression contexts, the IMR is computed using the formula $imr = \frac{\phi(\Phi^{-1}(\mathrm{pcdf}))}{\mathrm{pcdf}}$, where pcdf is the logistic cumulative distribution function of the estimated probabilities.

[67] In the wage regression, we include the same basic controls along with year, individual, and occupation FE as in the working hour regression, but also include total working hours as an additional control.

SSIV from the LFS with TSIV yield similar outcomes (see columns 3 and 4).

Table K.2. 1: The effect of transition on total working hours and monthly wage

	FE	IV		
	Basic controls and FE	SSIV (adoption)	SSIV (supply)	TSIV (LFS)
	(1)	(2)	(3)	(4)
Panel A. total working hours				
TTW	-0.187^{***}	-0.187^{***}	-0.187^{***}	-0.187^{***}
	(0.009)	(0.009)	(0.009)	(0.009)
FTW	-0.078^{***}	-0.078^{***}	-0.078^{***}	-0.078^{***}
	(0.006)	(0.006)	(0.006)	(0.006)
imr_TTW		-0.111	-0.099	-0.002
		(0.104)	(0.088)	(0.053)
imr_FTW		-0.006	-0.027	-0.003
		(0.071)	(0.078)	(0.014)
Panel B. monthly wage				
TTW	-0.065^{***}	-0.063^{***}	-0.062^{***}	-0.065^{***}
	(0.010)	(0.010)	(0.010)	(0.010)
FTW	-0.027^{***}	-0.028^{***}	-0.027^{***}	-0.027^{***}
	(0.007)	(0.007)	(0.007)	(0.007)
imr_TTW		-0.120^{***}	-0.118^{***}	0.011^{*}
		(0.024)	(0.024)	(0.007)
imr_FTW		-0.400^{***}	-0.785^{***}	-0.025^{**}
		(0.078)	(0.120)	(0.011)
N	40,997	40,997	40,997	40,997
Basic controls	Yes	Yes	Yes	Yes
I & T & O FE	Yes	Yes	Yes	Yes
P&A, seasonal	Yes	Yes	Yes	Yes

NOTE.—Panels A and B shows estimated effects of transiting from traditional work to TTW and FTW on total working hours and monthly wage for full-time employees by using linear regression, respectively. Column 1 show a simple two-way FE regression on the endogenous variables. Columns 2 and 3 present IV estimates using shift-share instrumental variables identified by FWA adoption and FWA supply, respectively. Column 4 displays TSIV estimates using IV based on FWA adoption from the LFS. The control variables included age, age square, marital status, number of children, family income, and firm size. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Moreover, to validate the relevance of the instruments used in these regressions, we examine the correlations in the first-stage regressions. As the Table K.2.2 displayed, the positive correlations between $SSIV^{TTW}$ and $TRAN^{TTW}$, as well as between $SSIV^{FTW}$ and $TRAN^{FTW}$, confirm the validity and relevance of these instruments across most columns. Additionally, F-test results in each first-stage regression exceed the critical threshold of 10, indicating no issues with weak instruments. Furthermore, the p-value of 0.000 in the over-identification test confirms that the instruments are sufficiently correlated with the endogenous explanatory variables. Overall, we find that while the FE model exhibits endogeneity issues, the coefficients remain stable after using SSIV, suggesting that any endogeneity-related bias is minimal. This consistency across models reinforces the robustness of our findings. Thus, our results provide reliable evidence on the effect of transitioning to TTW and FTW on working hours and wages in our main text.

Table K.2. 2: The IV estimates of the first-stage

	SSIV		SSIV (supply)		SSIV (LFS)	
	$TRAN^{TTW}$	$TRAN^{FTW}$	$TRAN^{TTW}$	$TRAN^{FTW}$	$TRAN^{TTW}$	$TRAN^{FTW}$
	(1)	(2)	(3)	(4)	(5)	(6)
$SSIV^{TTW}$	0.098***	0.013	0.072***	-0.014	0.154***	-0.006
	(0.019)	(0.027)	(0.018)	(0.029)	(0.058)	(0.008)
$SSIV^{FTW}$	0.013	0.077***	0.012	0.042*	0.030	-0.000
	(0.015)	(0.021)	(0.015)	(0.025)	(0.067)	(0.014)
N	40,997	40,997	40,997	40,997	40,997	40,997
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes
I & T & O FE	Yes	Yes	Yes	Yes	Yes	Yes
P&A, seasonal	Yes	Yes	Yes	Yes	Yes	Yes
F-test	19.92	19.89	19.57	19.49	16.38	7.65
Overidentification	0.000	0.000	0.000	0.000	0.000	0.000

NOTE.—The table reports the first-stage results for the IV specifications. The control variables included age, age square, marital status, number of children, family income, and firm size. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

Appendix L: The effects of transition on job discrimination

The table presents the relationships between different FWAs and job discrimination, including job satisfaction, job-related worry, and job-related anxiety for both men and women. For women, part-time work is associated with a statistically significant decrease in both job satisfaction and job-related worry. In contrast, term-time work shows no significant impact on job satisfaction or job-related anxiety for women.

Table N.2. 1: The effect of transition on job discrimination

	Men			Women		
	Job satisfaction	Job-related worried	Job-related anxiety	Job satisfaction	Job-related worried	Job-related anxiety
	(1)	(2)	(3)	(4)	(5)	(6)
Part-time work	-0.074	-0.068	0.098	-0.059**	-0.052**	0.046
	(0.065)	(0.045)	(0.116)	(0.030)	(0.021)	(0.056)
Term-time work	-0.028	0.087	-0.183	-0.031	-0.017	-0.011
	(0.096)	(0.066)	(0.170)	(0.040)	(0.029)	(0.075)
Flexi-time work	-0.019	0.001	0.018	0.029	-0.013	0.072
	(0.040)	(0.028)	(0.072)	(0.035)	(0.025)	(0.065)
N		12,646			16,059	
Basic controls		Yes			Yes	
I & T & O FE		Yes			Yes	
P&A, seasonal		Yes			Yes	

NOTE.—The table shows effects of transiting from traditional to TTW and FTW on job discrimination among full-time employees (without IV). The control variables included age, age square, marital status, number of children, family income, and firm size.*** p < 0.01.** p < 0.05.* p < 0.1.

Chapter 3

Understanding the Fertility Puzzle: How Bargaining Power Deviates from Traditional Models

Abstract

This study examines why traditional intra-household bargaining models often fail to accurately predict fertility behavior. To better understand this puzzle, we identify specific proxies of bargaining power that influence fertility decisions and highlight which groups deviate from the model's predictions. Utilizing a discrete-time survival model and data from HILDA, we find that both tangible bargaining power and intangible bargaining power statistically significantly affect birth intervals, with their combined presence strengthening this effect. Greater female bargaining power tends to align fertility timing with the woman's preferences when there is a disagreement between partners, which is consistent with bargaining theory. However, deviations from the bargaining model's predictions occur: career-oriented women delay childbearing despite having higher fertility preferences, whereas norm-conforming women have children earlier even when they have lower fertility preferences.

3.1 Introduction

Fertility behavior, defined as the timing, number, and spacing of childbearing, is widely recognized as one of the most complex household decisions to predict, particularly in developed countries. While early economic models conceptualized households as unitary decision-making entities with shared preferences, more recent theoretical developments have shifted toward *collective bargaining power models*. In these models, household decisions arise from negotiation processes shaped by each partner's bargaining power, which refers to their relative ability to influence outcomes based on access to economic resources, labor market opportunities, and perceived outside options (Chiappori, 1992). Within this framework, bargaining power plays a central role in shaping household outcomes, including fertility-related choices. For instance, a woman's bargaining power should theoretically enable her to align fertility timing with her preferences. However, despite the robust theoretical frameworks, empirical studies consistently show mixed and sometimes contradictory findings (Schultz, 1990; Lundberg and Pollak, 1996; Brodmann et al., 2007; Doss, 2013). Three key challenges contribute to this: first, the measurement of bargaining power, often approximated through various proxies; second, the failure to account for heterogeneity in couples' fertility preferences; and third, the existence of subgroups (e.g., career-oriented or norm-conforming women) whose fertility behavior diverge from the model.

Prior research has identified numerous proxies for intra-household bargaining power, which can be broadly categorized into two types: *tangible* and *intangible*. Tangible bargaining power, derived from observable economic resources and measured through objective indicators, reflects how such resources strengthen a woman's influence over household decisions (Chiappori, 1992; Thomas, 1994). However, much of the existing literature has focused primarily on tangible bargaining power, often overlooking the intangible dimensions of bargaining, leading to a limited understanding of how non-objective factors influence fertility outcomes. Intangible bargaining power, on the other hand, captures less easily quantified forms of influence and is typically measured using subjective indicators (Mason, 1986; Kabeer, 1999; Hogan et al., 1999; Ambrosetti et al., 2021). This form of power reflects a woman's ability to control economic resources within the household, regardless of her economic contribution. Building on this distinction, our analysis emphasizes that both tangible and intangible aspects of economic factors, whether objectively observed or subjectively perceived, are considered the primary determinants of bargaining power (e.g. Hoddinott and Haddad, 1995; Thomas, 1994; Allendorf, 2007; Klawon and Tiefenthaler, 2001). To distinguish between these two dimensions of bargaining power related to economic factors, we use *shared wage* as a proxy for tangible bargaining power and *financial control* as a proxy for intangible bargaining power.[68] This dual-measure approach allows us to capture the different dimensions of economic influence and examine their separate effects on fertility outcomes.

Despite advances in measuring bargaining power, much of the previous empirical research on bargaining theory has only focused on developing countries and often directly assumes women have lower fertility preferences than their male partners, without accounting for preference heterogeneity within couples (Bankole and Singh, 1998; Manser and Brown, 1980; Doepke and Kindermann, 2019). This assumption does not always hold, especially in developed countries. In our Household, Income and Labor Dynamics in Australia Survey (HILDA) sample, we found that nearly 75% of women have fertility preferences that are either higher than or equal to those of men.[69] This highlights the importance of explicitly incorporating preference heterogeneity in empirical analyses, particularly in developed country contexts where traditional assumptions may not apply. Another often-overlooked challenge in the fertility

[68] Tangible bargaining power is measured by women's share of the couple's total monthly wage income (hereafter referred to as *shared wage*), while intangible bargaining power is measured by the extent of women's control over household finances (hereafter referred to as *financial control*).

[69] In our sample, about 50% of women share the same fertility preference as men, and approximately 25% of women have a higher fertility preference than men.

preference literature is the issue of reverse causality (Rosenzweig and Wolpin, 1980; Angrist and Evans, 1998; Clarke, 2018). For instance, individuals with more children might experience a decreasing preference for additional offspring, which could create an artificial negative correlation.[70] This reverse causality concern extends to the study of bargaining power, where fertility outcomes might affect the distribution of bargaining power within the household. To mitigate this issue, our study employs survival analysis, focusing on birth interval as the dependent variable. This approach focuses on the transition in status, namely the duration until a subsequent birth, rather than the states itself.[71]

Nonetheless, while bargaining theory provides insights into intra-household fertility outcomes, it often neglects external factors such as labor market conditions and social norms. Career considerations, particularly the motherhood wage penalty, can diminish the effect of bargaining power, leading career-oriented women to delay childbearing. Hener's (2015) theoretical model suggests that women's increased bargaining power can have dual effects on fertility: it facilitates childbearing when preferences align, but it also encourages delay when career-related costs are high. Beyond the influence of career orientation, research on 'gender expectation' (West and Zimmerman, 1987) highlights that traditional role expectations can compel even women with significant bargaining power to conform to societal or spousal norms regarding motherhood, especially among those adhering to traditional gender roles. In this context, cultural pressures may dilute the effect of women's bargaining advantages, causing them to act contrary to their own fertility preferences. Overall, by examining both career-driven postponements and norm-driven accelerations of childbearing, this study offers new insights into the fertility puzzle, explaining why greater female bargaining power does not always lead to the fertility outcomes predicted by classical bargaining models.

---

[70] A clear example is observed in the changing fertility preferences of individuals. Initially, a woman without children might express a high preference for having children (e.g., a score of 10 on a preference scale). However, once she has a child, her preference for additional children might significantly decrease (e.g., dropping to 0). This shift can create an artificial negative correlation in regression analyses, suggesting that lower fertility preferences lead to more children, which is a misinterpretation due to the reverse causality effect.

[71] If an individual has a high fertility preference, it can directly shorter the birth interval.

Building on this framework, this paper sets out to examine how women's bargaining power within the household influences birth interval. Using longitudinal data from the Household, Income and Labor Dynamics (HILDA) survey in Australia and a discrete-time survival model, we found that both tangible bargaining power (measured by the shared wage) and intangible bargaining power (measured by the financial control) significantly affect birth interval, with their combined presence further reinforcing this influence. Greater female bargaining power tends to influence birth interval in line with woman's preferences when partners differ in fertility preference. Despite the overall support for the bargaining model logic, our findings reveal some deviations from its predictions. Some career-oriented women delayed childbearing even when they expressed a desire for children and had stronger bargaining power, implying that career considerations can override the predicted influence of bargaining power on fertility decisions. Conversely, some women with low fertility preference had children earlier than their own preferences would predict, likely due to traditional social and gender norms. These cases highlight that fertility outcomes are influenced not only by individual fertility preferences but also by personal career goals and social pressures, suggesting that traditional bargaining models should incorporate both career aspirations and social norms to better capture the complexities of fertility decisions. Nonetheless, couples with aligned fertility preferences exhibit no significant bargaining-based effects on fertility outcomes, as there is no need for negotiation in these cases. When both partners have high fertility preferences, the fertility behavior tends to reflect their mutual desire. However, when both partners prefer fewer children, their shared family planning results in longer postponement of birth intervals.

This study notably advances the field of fertility behavior and intra-household bargaining power research in several key aspects. First, we identify the specific facets of bargaining power that meaningfully influence fertility decisions, highlighting that both tangible and intangible female bargaining power significantly affect the birth interval. To the best of our knowledge, there are only two studies that explored the relationship between bargaining power and birth interval by using discrete-time survival model, namely Upadhyay and Hindin (2005) and Feldman et al. (2009). However, both studies, along with the majority of previous fertility research, primarily focusing on tangible measures of bargaining power, overlook the role of intangible bargaining power. Second, we account for preference heterogeneity by explicitly examining the fertility preference gap between partners when analyzing the effects of bargaining power on fertility behavior, rather than assuming that women universally prefer fewer children, as often assumed in studies of developing countries (Gipson and Hindin, 2009; Sarnak and Gemmill, 2022; Tong et al., 2024). Notably, nearly 75% of women in our HILDA sample report fertility preferences that are equal to or higher than those of their partner, emphasizing the importance of considering heterogeneity in fertility preferences within developed countries.

Third, we move beyond documenting an association to explain why some women deviate from the fertility behavior predicted by bargaining models. Our introduction of 'career considerations' and the 'gender expectation' as mediating factors is novel in the context of fertility decision-making. We show that career-oriented women postpone births when faced with high career costs and that traditional gender role pressures can counteract women's bargaining advantages, leading them to have children against their initial inclination. These insights help resolve the fertility puzzle by identifying the conditions under which the standard bargaining model fails. To our knowledge, this is the first study to systematically document how such factors can override the impact of bargaining power in fertility decisions, shedding light on mechanisms that previous bargaining models did not fully capture. Fourth, we contribute to the literature by focusing on a developed country context, specifically Australia, where fertility decisions are influenced under different economic and social dynamics than those typically studied in developing countries (e.g., Chiappori 1992; Agarwal 1997; Upadhyay and Hindin 2005; Doss 2013; Ashraf et al. 2014; Hener 2015; Feldman et al. 2009). In developed settings, children are more often viewed as a lifestyle choice rather than as a form of economic security or old-age support, as is common in developing contexts. Additionally, lower child mortality rates, widespread access to contraception, and greater career opportunities for women fundamentally reshape the nature of fertility preferences and investments in children. These distinctions highlight the importance of studying bargaining power and fertility behavior in developed countries.

3.2 Literature review and Hypotheses

Research on household decision-making and fertility has moved beyond the unitary model of the family, which assumed a single set of preferences and joint utility maximization (Becker, 1960). Instead, collective bargaining models propose that spouses often have different fertility preferences and resources, and that family outcomes reflect a bargaining process between partners (Chiappori, 1992). This shift has significantly influenced how researchers approach bargaining power and its impact on fertility behavior. This following section reviews the literature across four key themes: (1) the concept of bargaining power; (2) the role of intra-household bargaining in fertility decisions; (3) deviations from the predictions of bargaining power models; and (4) similarity in couples' fertility preferences.

A. The concept of Bargaining power

Intra-household bargaining power refers to an individual's ability to influence decision-making and outcomes in the family setting. Notably, bargaining power is inherently unobservable, making it necessary for researchers to rely on proxies to capture its effects. In the literature, these proxies are typically categorized into two types: tangible (objective, resource-based) and intangible (subjective, perception-based). Tangible bargaining power refers to influence rooted in quantifiable economic resources and assets, whereas intangible bargaining power derives from less observable factors like perceptions of authority or decision-making. A classic view is that greater access to or control over resources strengthens an individual's fallback position, thereby enhancing her bargaining power within the family (Manser and Brown, 1980; Lundberg and Pollak, 1993; Chiappori, 1992). Empirically, a wide range of objective proxies has been used to capture tangible bargaining power. Examples include a wife's share of household income or unearned income (e.g. Hoddinott and Haddad, 1995; Thomas, 1994; Allendorf, 2007; Klawon and Tiefenthaler, 2001), relative labor supply or working hours (Blundell et al., 2005; Antman, 2014), comparative education or the employment status of spouses (Musick et al., 2009). Use of these indicators is predicated on the assumption that economic resources confer greater influence in household decisions.

In contrast, intangible bargaining power refers to decision-making authority within the household, such as control over money, expenditures, and family planning decisions, which are not directly linked to tangible resources. Intangible bargaining power is typically measured through self-reported assessments of decision-making authority and other subjective indicators (Mason, 1986; Kabeer, 1999; Hogan et al., 1999; Ambrosetti et al., 2021). Common survey-based measures ask who has the final say on various household decisions, for example, who controls daily expenditures, large purchases, children's education, or family planning decisions (Allendorf, 2007). Such self-rated measures try to capture a person's perceived influence within the household (Reggio, 2011; Mason, 1986; Kabeer, 1999; Ambrosetti et al., 2021). Moreover, tangible and intangible forms of bargaining power are often interrelated rather than independent. Agarwal (1997) and Kabeer (1999) argue that increased material resources can also enhance influence over household decision-making, thereby shaping bargaining outcomes. This insight aligns with evidence suggesting that women's empowerment is multidimensional (Agarwala and Lynch, 2006).

B. Bargaining power in Fertility Choices

The traditional unitary model (Becker, 1960) assumed that the household acts as a single decision-maker with a unified fertility preference. However, accumulating evidence contradicts this assumption. Studies from diverse settings show that spouses often have disparate reproductive goals. In a study of 18 developing countries, Bankole and Singh (1998) found that husbands generally want more children and at earlier time than their wives, with 10-26% of couples having differing fertility preference. Given that spouses may differ in their desired family size or birth interval, the dynamics of bargaining power between couples are crucial in determining fertility outcomes (Manser and Brown, 1980; Lundberg and Pollak, 1993). According to the Chiappori's (1992) collective model formalizes, when partners have different utility trade-offs, the final decision on fertility depends on their relative bargaining power and the weight each partner's preference carries in household choices.

Several empirical studies find results that are consistent with the above bargaining model in developing countries. For instance, Thomas (1990) reported that in Brazil, women with higher non-labor income (a proxy for greater tangible bargaining power) had significantly fewer children on average. This suggests that empowered women were able to limit fertility, presumably reflecting their lower optimal family size. Subsequent studies in other settings found similar negative relationships between women's bargaining power and fertility outcomes. For example, using data from India, Varanasi (2009) showed that increases in women's bargaining power (measured through a composite index) were associated with declines in the number of children born. However, other literature documents cases that challenge the assumption of universally lower female fertility preferences. In some contexts, women's desired number of children equals or even exceeds that of men. When this is true, empowering women need not reduce fertility, it may have neutral or positive effects on childbearing. Schultz (1990) provides an illustrative example: studying Thailand, he found that higher unearned income for women (proxy for tangible bargaining power) was associated with higher fertility rates. He argued that Thai women with greater economic resource opted to have more children, treating children as a 'normal good' or a form of leisure consumption. Similarly, Mason and Smith (2000) observed instances in which women's decision-making power correlated with slightly increased likelihood of additional births.

On the other hand, research on intangible bargaining power remains more limited and primarily contains experimental papers. One notable example comes from Ashraf et al. (2014), who conducted a randomized experiment in Zambia. They found that when women were granted greater intangible bargaining power through increased privacy in contraceptive use, effectively reducing their husbands' involvement, fertility rates declined. This outcome suggests that when women have greater decision-making freedom, they may choose to avoid unwanted births, reinforcing the importance of intangible bargaining power in shaping fertility outcomes. Conversely, when husbands dominate decisions, outcomes skew toward husbands' preferences (Bankole and Singh, 1998). More broadly, who has the final say on reproductive matters is often a reflection of underlying intangible bargaining power. Mason and Smith (2000), in their analysis of decision-making in Asian countries, found that in certain communities, greater female involvement in household decisions, particularly when women had lower fertility preferences, was associated with lower fertility. Overall, building on the findings from previous research in developing countries, we hypothesize that this relationship also holds in developed countries such as Australia: *women with greater tangible or intangible bargaining power are associated with longer birth intervals.*

C. Deviations from Bargaining Power Models in Fertility Choices

As discussed above, the bargaining model does not always predict fertility choices well. Agarwal (1997) argues that women's fertility decisions are shaped by broader economic and social contexts. Even when women have higher bargaining power and stronger fertility preference, they may choose to delay childbearing if early motherhood conflicts with their career aspirations. Classic economic models highlight that women with greater earning power face higher opportunity costs when interrupting their careers, which in turns incentivizes longer birth intervals and later ages at first birth (Becker. 1960; Mincer, 1963). The motherhood wage penalty and associated career setbacks often encourage career-oriented women to postpone childbearing until they are more professionally established. Importantly, even stronger bargaining power within the household cannot offset the high career costs of childbearing when women prioritize their careers. This theoretical framework aligns with Hener's (2015) model, which posits that women's higher wages, as a form of bargaining power, can simultaneously encourage both increased fertility and delayed childbearing, depending on the interaction between fertility preferences and career development.

The process of intra-household bargaining is also shaped by social and gender norms. Even a woman with substantial economic resources (tangible bargaining power) and formal decision authority (intangible bargaining power) may face social pressures regarding her role as a mother or wife. Sociological research on 'gender expectation' suggests that individuals often behave in ways that reinforce traditional gender roles, sometimes even against their immediate self-interest. In the context of fertility, this implies that a woman may acquiesce to having a child or refrain from using contraception not because she lacks bargaining power per se, but because such behavior aligns with expected gender norms. For instance, qualitative studies have observed that high-achieving women in the United States often left the workforce to become stay-at-home mothers, influenced by cultural or gender expectations of good motherhood, despite their strong economic position (Hochschild, 1989). This pattern supports the idea that even as women's economic independence increases, it still "does not solely dictate their roles and behaviors". In some cases, women with greater resources may still defer to traditional norms in family decision-making, thereby limiting their extent of bargaining advantage.

Building on these insights, we propose two hypotheses. First, *even when career-*

oriented women have higher fertility preferences and greater bargaining power, they may delay childbearing due to the high opportunity costs of early motherhood. Second, the effect of women's bargaining power on fertility outcomes is moderated by traditional gender attitudes; specifically, *women with lower fertility preference but strong gender roles may exhibit fertility behaviors consistent with those attitudes, resulting in shorter birth intervals despite their greater bargaining*.

D. Aligned Fertility Preferences Between Partners

In cases where spouses have identical fertility preferences, the typical predictions of bargaining theory may not hold, as there is no conflict to resolve (Chiappori, 1992; Hener, 2015). As a result, changes in tangible or intangible bargaining power (e.g., through changes in shared wage or financial control) may not significantly influence fertility behavior. Previous work has supported this view, with studies by Rasul (2008), Riederer et al. (2019), and Hener (2015) showing that in couples with identical preferences, fertility outcomes remained unaffected by shifts in bargaining power. This scenario aligns with the Signal Utility Model, which posits that households will achieve their fertility goals independently of bargaining dynamics when preferences are perfectly aligned (Becker, 1981; Hener, 2015). According to this model, when couples hold aligned fertility preferences, they are more likely to make cooperative decisions without requiring negotiation. As a result, such couples may even have higher fertility rates than those with discordant preferences, particularly when both partners desire larger families. Moreover, the Signal Utility Model suggests that the signaling effect of household preferences plays a central role when both partners have the same goals (Mason and Smith, 2000). For instance, spouses are more likely to support each other's preferences, leading to joint action on fertility decisions. However, when both partners share a low fertility preference, the outcome may be the opposite. In such cases, mutual agreement to delay or limit childbearing is reinforced, leading to longer birth intervals than those observed in couples with divergent preferences (Hener, 2015; Doepke and

Kindermann, 2019).

3.3 Data and Stylized facts

A. HILDA Survey

The data for this study comes from the Household, Income and Labor Dynamics in Australia (HILDA) Survey. HILDA is a household-based panel study that collects information about economic and personal well-being, labor market dynamics and family life. It started in 2001 and, follows the lives of more than 17,000 Australians each year. Each wave/questionnaire comprises a set of core and rotating content, where core content is carried at each wave, but rotating content is carried at intervals. Information on intangible bargaining power measured by financial control is collected in waves 5-14, 16, 18 and 20. Data on tangible bargaining power such as shared wage, has been collected in every wave throughout the study period. Based on the research objectives, the sample was modified in the following ways. First, women younger than 16 and older than 49 are dropped (38.7%).[72] The minimum legal working age in Australia is 16, and the fertile period of women is considered to be between 14 and 49.[73] Additionally, men older than 60 are excluded, as they may be retired and earning zero wage, which would artificially inflate the shared wage for women. Second, this paper only focuses on heterosexual couples, which includes couples who were married or cohabiting (dropping about 28.3% of the sample).[74] Then, this paper further excludes all couples having children before living together (4.0%), because it is possible that these children are inherited from a previous marriage or relationship. Finally, couples

[72] Since our data is longitudinal, we use the minimum age of women as a reference for data exclusion. This approach ensures that we can consistently match and track women at the household level over time, avoiding situations where some women are excluded because of aged from the sample. Additionally, this method helps mitigate the issue of right-censoring.
[73] We consider the minimum legal working age because one of the important independent variables is tangible bargaining power, constructing by the monthly wage shared.
[74] Same-sex marriage was legalized in Australia in 2017. While we do not have sufficient data to analyse same-sex couples, we expect that gender dynamics would not apply in the same way in a non-heteronormative relationship.

in which both partners report zero earnings are excluded, because bargaining power in these households may be determined by factors beyond earnings. After matching the male and female at the household level, the sample comprises 3,804 couples and 37,951 observations. Nonetheless, there are 2,406 individuals with 5,916 observations that, although unmatched, can still represent the household level, as our main dependent and independent variables can be identified based on information from one partner alone. Overall, the sample comprises 5,677 couples and 42,867 observations, where the women are aged 16 to 49. Among them, 70.5% have at least one child.

B. Dependent and independent variables

Birth interval. Birth interval is typically measured by the time between consecutive births. In our sample, pregnancy histories were not collected; however, the age of each child was provided, enabling the calculation of birth spacing based on the duration of marriage or cohabitation and the ages of the children. Specifically, the first birth interval is calculated by subtracting the age of the first child from the total duration of marriage or cohabitation. Subsequent birth intervals are derived from the age gaps between successive children. Table C.3.1 in Appendix C shows that the average number of children is approximately 1.64, and the mean birth interval is 15.9 years (see Table A.3.1 in Appendix A). Among these observations, around 15% represent first births with an average birth interval of 5.7 years, and approximately 30% represent second births with a shorter average interval of 2.5 years. Additionally, Figure A.3.1 in Appendix A illustrates the Kernel Density Estimation (KDE) of the birth intervals, showing that the second birth interval is more concentrated around shorter durations, while the first birth interval displays a longer duration.

Shared wage. To measure tangible bargaining power, we focus on shared wages. This measure is frequently cited in intra-household bargaining literature and is considered the key indicator of tangible bargaining power (Hoddinott and Haddad, 1995; Thomas, 1990; Schultz, 1990). Specifically, we calculate female relative monthly wages as her earnings divided by the total earnings of both her and her partner, adopting a female-centric perspective in our analysis.[75] This ratio is also known as "shared wage":

$$shared\ wage = \frac{female\ wage}{male\ wage + female\ wage} \tag{1}$$

With this approach, couples in which both partners have zero wage were excluded.[76] Additionally, this approach to measuring female economic independence is favored for its simplicity and ease of interpretation, as affirmed by foundational studies such as Hoddinott and Haddad (1995) and Thomas (1990). A similar method is used to construct measures of shared working hours and shared household chore hours as a robustness check (Doepke and Kindermann, 2019). For shared household chore hours, we use one minus the ratio to reflect the idea that as women take on fewer household chores, their bargaining power increases.[77] Table B.3.1 in Appendix B shows a strong correlation between shared wage and shared working hours, with a coefficient of 0.71, suggesting that these two measures may reflect the same dimension of tangible bargaining power.

Financial control. Following the methodology of Rammohan and Johar (2009), we use self-rated measures of financial control as a proxy for intangible bargaining power, focusing on decision-making processes related to control over resource distribution

[75] In the HILDA dataset, about 25.0% of the sample cannot be matched at the household (couple) level within the same wave. To retain these observations and calculate couples' monthly wages, we use two variables: "*total family monthly wage*" and "*family type*". Based on the family type, we determine the household composition. If a household consists solely of a couple, we estimate the partner's monthly wage by subtracting an individual's monthly wage from the total family monthly wage. However, this method cannot be applied to families that include other relatives, since our primary focus is on couples, and the presence of additional family members may influence shared wage.
[76] However, we include them in the following analysis to compare with couples who have an equal shared wage, in order to distinguish between these two types of equal shared couples and to examine the effects of bargaining on fertility.
[77] While several studies have identified alternative proxies for bargaining power, such as the share of unearned income (Schultz, 1990), dowry or bride price (Anderson and Bidner, 2015), individual land or property ownership (Allendorf, 2007), access to credit or formal banking, and ownership of productive assets like livestock or business capital (Duflo and Udry, 2004), we are unable to replicate these measures due to limitations in the HILDA dataset. Specifically, the data do not consistently provide disaggregated information on unearned income sources or detailed asset ownership for both partners.

within household. Specifically, we utilize the decision module from the HILDA survey, which inquiries about who is responsible for making decisions in *making large household purchases*. Participants could select from: (1) *always me*, (2) *usually me*, (3) *decisions shared equally with my partner*, (4) *usually my partner*, or (5) *always my partner*.[78,79] These responses were then recoded into a scale from 0 to 1, where 0 represents decisions are made entirely by the male partners and 1 represents decisions are made entirely by the women. There are two additional decision-making variables related to finance, such as paying bills, as well as investment and saving, but we consider that they may not fully reflect the concept of financial control, as paying bills may simply reflect routine financial management rather than a true exercise of control, while saving decisions might be influenced by broader financial goals or external factors unrelated to intra-household power dynamics.[80] Table B.3.2 in Appendix B shows that the correlations among the three financial decision-making variables, large purchases, bill payments, as well as investment and saving, are relatively low, ranging from 0.37 to 0.49. Moreover, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is 0.66, which falls below the commonly accepted threshold of 0.70 for reliable factor analysis.

Figure 3.1 presents the Kernel Density Estimation (KDE) of shared wage and financial control. The solid line represents the distribution of shared wage, which peaks around 0.5, indicating that a large proportion of couples have relatively balanced monthly wage contributions. However, the KDE also show a high density at lower

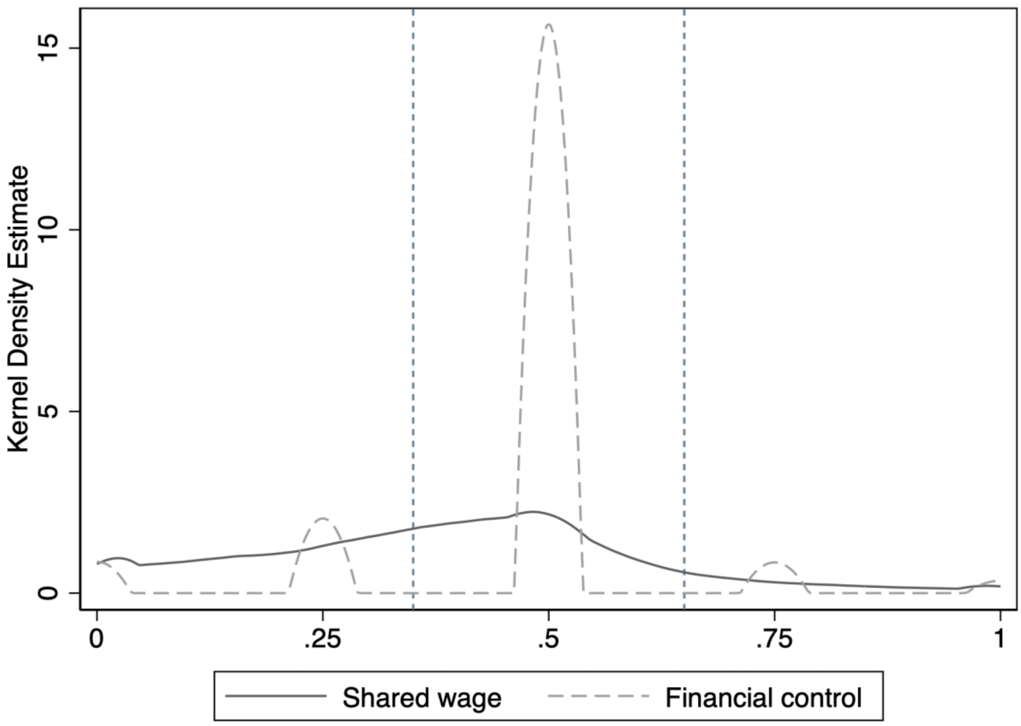
[78] We excluded those couple who reported that the decision was made by their relative (about 2.7%), as such responses do not reflect the idea of bargaining power among the couples. Additionally, cases in which couples gave conflicting reports about who made decision were also excluded, representing only 0.3% of total responses.

[79] Similar to the shared wage variable, some couples cannot be matched within the same wave, which may result in significant omissions of the financial control variable. However, since only 0.3% of couples report differing perceptions regarding economic decision-making, we rely on responses from the available partner when couple-level information is missing in a given wave. Additionally, since economic decision-making data is only available in waves 5–14, 16, 18, and 20, and these roles within households are generally stable over time, we assume that the decision-making variables remain consistent in the nearest previous and subsequent years. This assumption allows us to impute missing values for the waves where data was not collected.

[80] There are two more financial decision-making questions, such as (1) *managing day-to-day spending and paying bills*, (2) *saving, investment and borrowing*.

shared wage values, suggesting that in some households, women still contribute less to household income. On the other hand, the dashed line represents the distribution of financial control, which follows a similar pattern. It shows higher density around the mid-range 0.5-0.7. This suggests that financial control is also more likely to be equally shared within the household. To simplify the continuous variables and enhance interpretability, we categorize shared wage into three groups: male breadwinner, equal breadwinner, and female breadwinner, and financial control into three categories: male controller, equal controller, and female controller. This classification defines male breadwinner or controller as the index of shared wage or financial control below 35% (see Figure 3.1), equal breadwinner or controller as between 35% and 65%, and female breadwinner or controller as above 65% (similar to Browning et al., 2013). Also, these categorizations are visually represented in Figure 3.1, where the thresholds for each category are marked along the distribution curves for shared wage and financial control.

Figure 3. 1: The Kernel Density Estimation of the Shared wage and Financial control



Note: The Figure presents the Kernel Density Estimation (KDE) of shared wage and financial control. The x-axis represents the value of shared wage and financial control within the household, ranging from 0 to 1. The y-axis represents the kernel density estimate, which illustrates the distribution of these variables across the sample. The solid line corresponds to the distribution of shared wage, while the dashed line represents financial control. The vertical dashed lines at 0.35 and 0.65 indicate thresholds used to classify different levels of shared wage and financial control. The KDE provides a smoothed estimate of the distribution, representing the probability density at each value on the x-axis, and thereby indicating the concentration of data points around that value.

Following the examination of the KDE, Table 3.1 presents descriptive statistics for shared wage and financial control across all households, as well as for households with and without children. Male breadwinner households make up 51.4% of the subsection of household with children, compared to 28.5% of subsection of households without children. This finding aligns with previous research suggesting that, in male breadwinner or controller households, men with higher tangible (proxied by shared wage) or intangible (proxied by financial control) bargaining power are more likely to have children.[81,82] However, the bargaining power theory does not fully explain why the proportion of female breadwinner household or female controller household is similar in household with children and those without, suggesting that higher female tangible or intangible bargaining power does not necessarily reduce the probability of childbearing. Additionally, the theory struggles to account for equal couples, in which either tangible or intangible bargaining power is relatively balanced. This is important, as the proportion of equal breadwinner households accounts for 33.1% and equal controller households accounts for 73.6% of our full household sample.

Table 3. 1: Descriptive Statistics

	All households		With child		Without child		
Variables	Mean	SD	Mean	SD	Mean	SD	Diff
Shared wage							
Male	.449	.497	.514	.5	.285	.451	0.230***
Equal	.331	.471	.266	.442	.495	.5	-0.23***
Female	.22	.414	.22	.414	.22	.415	-0.00100
Financial control							
Male	.193	.395	.2	.4	.177	.382	0.022***

[81] However, it is also possible that having children itself reduces women's bargaining power in these households. The following empirical results present the findings after addressing this potential two-way causality problem.
[82] In the previous paper, the assumption was that women always have a lower fertility preference than men in developing countries (e.g., Thomas, 1990; Schultz, 1990; Mason and Smith, 1999). However, this dynamic may differ in developed countries. Therefore, in the following empirical section, we further divide the sample based on the couples' fertility preferences and provide a more comprehensive explanation of the bargaining power theory.

Equal	.736	.441	.731	.444	.75	.433	-0.02***
Female	.07	.256	.069	.254	.073	.26	-0.00300

NOTE.—Data are sourced from the HILDA Survey, covering the period from 2005 to 2020. The table reports descriptive statistics on the probability of having at least one child across different household types, categorized by shared wage (male breadwinner, equal breadwinner, and female breadwinner) and financial control (male controller, equal controller, and female controller), for three groups: all households, households with children, and households without children. Shared wage is based on annual income from the primary job, calculated as the female's share of the couple's total monthly wage. Financial control is determined by who makes decisions regarding large household purchases.

3.4 Emprical analysis

Our analysis is organized into three distinct phases in this section. First, we discuss the identification strategy. Second, we examine how shared wage and financial control influence fertility behavior. Third, we explore the underlying mechanisms associated with bargaining power theory.

A. Identification strategy

The analysis is conducted using a discrete-time survival model, where the birth interval is regressed to examine the effects of shared wage and financial control. There are three key reasons for using the birth interval rather than the number of children as the dependent variable. First, using the number of children as the dependent variable sacrifices important temporal detail about fertility behavior. In contrast, modeling birth intervals allows us to capture not just whether, but when, couples have additional children, providing a more detailed understanding of the fertility behavior in relation to bargaining power. Second, fertility research often encounters right censoring issues, particularly when individuals have not yet completed childbearing by the end of the observation period. To address this, we construct retrospective birth intervals for each parity by utilizing information on the duration of marriage or cohabitation and the age of each child, allowing us to reconstruct the sequence and timing of births during the observed period and to more accurately capture variation in birth spacing. Third, the current shared wage and financial control may be influenced by fertility outcomes when using number of child as dependent variable, which is also known as reverse causality problems. For instance, it is possible that women anticipating future additional childbearing intentionally select into lower-earning or more flexible jobs that facilitate work–life balance. This preemptive adjustment in labor market behavior could reduce their measured bargaining power, thereby confounding the direction of causality. By focusing on birth interval as the dependent variable in a survival analysis, we emphasize the timing of transition in status, specifically the duration until a subsequent birth, rather than the outcome of the status itself. It is unlikely to consider that couples adjust their current wage-sharing arrangements or financial control in response to the exact timing of a future birth, which helps mitigate concerns about reverse causality.

The remaining concern is interval censoring. To address this, we use the Complementary Log-log (Cloglog) model instead of the standard Logit model, as it is better suited for discrete-time survival analysis and more effectively handles interval censoring.[83] However, a potential concern is that, apart from the variability in hazard rate across couples that can be attributed to couples' bargaining power, other factors influencing the variability in hazard rate could introduce bias into our findings. To address this, we include a comprehensive set of both time-constant and time-varying covariates. Despite this, our dataset may not capture all relevant variables, which raises the issue of unobserved heterogeneity. To mitigate this, we incorporate frailty γ_i through random effects in our model. In general, the random-effect discrete-time Cloglog survival model is specified as follows:

$$\log(-\log(1 - h_{ijtw})) = \rho S_{ijtw} + \tau F_{ijtw} + \boldsymbol{X}'_{ijtw}\eta + \gamma_i + \zeta_j + \eta_t + \phi_w + \varepsilon_{ijtw} \quad (2)$$

[83] This is because the underlying hazard process occurs continuously over time, but we only observe events at discrete intervals, such as annually or at each survey wave. The Cloglog model accommodates this by approximating the continuous process more accurately than the Logit model.

where h_{ijtw} denotes the discrete-time hazard function, given the covariate values, for each i^{th} household of birth parity j, at survival time t, in wave w; S_{ijtw} and F_{ijtw} represent the category variables of shared wage (i.e., male breadwinner, equal breadwinner, female breadwinner) and financial control (i.e., male controller, equal controller, female controller); $\boldsymbol{X}'_{ijtw}$ is the matrix of control variables, including the average of male and female values for age, education, self-rated health, and fertility preference, as well as marital status, and household income;[84] γ_i is the household random effect. We further include the wave fixed effect ϕ_w to capture common shocks across households in a given survey wave and survival time fixed effect η_t to account for the temporal variation in hazard rate, and ζ_j denotes birth parity fixed effects, which capture the unobserved heterogeneity across different birth parities; and ε_{ijtw} denotes the error term.

B. The effects of shared wage and financial control

To assess how different forms of bargaining power influence fertility behavior, we examine the effects of shared wage and financial control on birth intervals. Results are presented in Table 3.2. The dataset includes 5,677 couples, with a total of 42,867 observations across waves. When the data is expanded through the application of discrete-time survival analysis, considering the duration of couples' cohabitation, the full sample consists of 78,252 observations. A more detailed explanation of this methodological approach is provided in Appendix A. On average, the observed duration of marriage or cohabitation in our sample is approximately 15.9 years (see the Table A.3.1 in Appendix A). Table 3.2 presents the results of the impact of shared wage and financial control on the hazard rate of having a child (i.e., birth interval). The results are presented in three specifications: column (1) includes only the main variables of

[84] Appendix C provides detailed descriptive statistics for each control variable in the full sample, separately presented for households with and without children.

interest and household fertility preference; column (2) adds financial management variables (e.g., bill payments, investments and savings) to isolate the effect of financial control; and column (3) further incorporates control variables (e.g., average values of male and female for age, education, and self-rated health, as well as marital status, and household income) along with fixed effects as a robustness check.

Specifically, column (1) shows that male breadwinner households exhibit the highest hazard rate of having a birth, with coefficients 0.032 ($p < 0.01$) and 0.009 ($p < 0.01$) higher than those of equal and female breadwinner households after controlling for fertility preference.[85] This suggests that men with greater tangible bargaining power are more likely to have children sooner. Similarly, when measuring intangible bargaining power through financial control, a similar pattern emerges, with equal-shared controller households exhibiting a lower hazard rate (coefficient = -0.005; $p < 0.1$) compared to male controller households. Interestingly, in female controller households, the hazard rate is comparable to that of male controller households (coefficient = -0.000; $p > 0.1$). This suggests that in female controller households (where women hold greatest intangible bargaining power), the birth interval tends to be short. This contrasts with much of the existing literature that finds female breadwinner households (where women hold greatest tangible bargaining power) are typically associated with lower fertility probability (Thomas 1990; Varanasi, 2009; Schultz, 1990; Mason and Smith, 2000).

Moreover, to further isolate the effects of financial control, we include two types of financial management variables in the model: bill payments, investment and savings (see column 2). Consistent with our hypothesis, paying bills reflects routine financial management rather than a true exercise of control. As displayed in Table F.3.1 in Appendix F, similar to the effect of involvement in household chores in previous research (Kan et al., 2019; Erfani and Pilon, 2025), when women engage more in paying

[85] Notably, controlling for fertility preference is important as unlike many of these earlier studies that focus on developing countries, our analysis examines Australia, a developed country, where the distribution of fertility preferences is similar between men and women as shown in Figure D.3.1 in Appendix D.

bills activity, their bargaining power is diminished, resulting in shorter birth intervals. On the other hand, saving decisions might be influenced by broader financial goals or external factors unrelated to intra-household power dynamics, and consequently, they have an insignificant effect on the hazard rate. Overall, the results of the effect of shared wage and financial control on birth interval remain robust even when controlling the impact of money management. These findings also hold after accounting for personal and family characteristics and incorporating household and duration fixed effects (see column 3). Additionally, the results are consistent when using alternative proxies for tangible bargaining power, such as shared working hours and shared housework hours (see columns 4 and 5), both of which are negatively associated with the hazard rate of births when women hold more bargaining power.

Overall, these findings suggest that women with higher shared wage or greater financial control generally experience longer birth intervals. Figure H.3.1 in Appendix H further illustrates that the combined presence of higher shared wage and financial control amplifies this effect. In particular, the highest hazard rate is observed among men with the highest shared wage and financial control. In contrast, the lower hazard rates are found in female breadwinner and female controller households. Notably, we found that the lowest hazard rates are found in equal breadwinner or equal controller households. This result is particularly significant, given that equal controller households make up 73.6% of the sample (Table 3.1). While equal breadwinner households represent a smaller proportion (33.1%), this proportion has been rising over time (see Table E.3.1 in Appendix E). Taken together, these findings suggest that the high prevalence and growing trend of equal breadwinner and equal controller households may lead to longer birth intervals, potentially playing a role in the decline of Australia's total fertility rate.

Table 3. 2: The average marginal effect of shared wage and financial control on the birth interval

	Baseline	Money management	Controls and RE	Shared working hours	Shared housework hours
	(1)	(2)	(3)	(4)	(5)
Shared wage					
Equal vs. Male	-0.034***	-0.033***	-0.032***		
	(0.002)	(0.002)	(0.002)		
Female vs. Male	-0.019***	-0.018***	-0.009***		
	(0.003)	(0.003)	(0.003)		
Financial control					
Equal vs. Male	-0.005*	-0.008***	-0.007**	-0.009***	-0.008***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Female vs. Male	-0.000	-0.008*	-0.003	-0.006	-0.008
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Shared working hours					
Equal vs. Male				-0.040***	
				(0.002)	
Female vs. Male				-0.028***	
				(0.004)	
Shared housework hours					
Equal vs. Male					-0.008***
					(0.002)
Female vs. Male					-0.022***
					(0.004)
N	78,252	78,252	78,252	78,252	78,252
Controls	No	No	Yes	Yes	Yes
Household RE	No	No	Yes	Yes	Yes
Parity & duration & wave FE	No	No	Yes	Yes	Yes
Money management	No	Yes	Yes	Yes	Yes

NOTE.—Estimates from discrete-time survival regressions on the birth interval. In survival analysis, the dependent variable is a binary variable indicating whether an event has occurred or not within a time interval. As coefficients in non-linear models are not directly comparable, we present average marginal effects instead. The sample includes 5,677 households with women aged 16-49 and 78,252 observations. The complete set of controls includes average of male and female values for age, education, self-rated health, and fertility preference, as well as marital status, and household income. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

C. Mechanisms

The previous sections show that women with higher shared wage or greater financial control generally experience longer birth intervals. To better understand the underlying mechanisms driving this pattern, we explore its implications through the lenses of bargaining theory. The core principle of bargaining theory rests on the heterogeneity of partners' preferences. For example, when the woman holds greater bargaining power and has a lower fertility preference, she may be more likely to influence fertility decisions, resulting in a longer birth interval (Chiappori, 1992; 1997). However, as discussed in the literature review, gender attitudes can reinforce traditional roles, prompting women to acquiesce to childbearing, not due to a lack of bargaining power, but because such behavior aligns with socially expected norms of motherhood. Therefore, we expect that when women have lower fertility preferences, higher bargaining power will be associated with a lower hazard rate of births interval, except among those with strong adherence to traditional gender attitudes. In contrast, when women have higher fertility preferences, bargaining theory predicts a shorter birth interval with greater bargaining power; however, this pattern is primarily observed among part-time employed women, as full-time employed women may delay childbearing due to career commitments. Last but not least, for women who share equal fertility preferences with their partners, we expect that the bargaining power model will have limited predictive power, as there is no preference conflict to negotiate and fertility decisions.

To explore the potential mechanisms, it is crucial to first categorize couples based on whether the female partner's fertility preference is higher, lower, or equal to that of her partner. This step is often overlooked in previous research, which frequently assumes that women have lower fertility preferences than men, an assumption that does not consistently hold. In developed countries such as Australia, this is particularly evident.

As shown in Figure D.3.1 in Appendix D, fertility preferences are generally similar between men and women in the Australian context. To construct the fertility preference gap, we rely on information in the HILDA questionnaire: (1) "About any future children. *Picking a number between 0 and 10 to show how likely are you to have a child or more children in the future*".[86,87] We calculate the fertility preference gap by subtracting the women's fertility preference from the men's. Based on this measure, we categorize the sample into three groups: women with lower, equal, and higher fertility preferences than their male partners.

As shown in Table 3.3, columns (1) and (2) present results for the subsample in which the male partner's fertility preference is higher than the female's. Specifically, we find that in these households, when men hold greater shared wage and financial control, the birth interval tends to be shorter. This is evidenced by the higher hazard rates observed in male breadwinner (coefficient = -0.019; $p < 0.01$) or male controller households (coefficient = -0.016; $p < 0.01$) compared to their equal households. However, the bargaining theory does not fully explain those female breadwinner or female controller households since they do not exhibit significantly lower hazard rates. In fact, when measured by financial control, female controller households display slightly higher hazard rates, though this effect is not statistically significant (coefficient = 0.007; $p > 0.1$). These results challenge the expectation that women with stronger bargaining power and lower fertility preferences would necessarily experience longer birth intervals. To further explore this pattern, we divide the sample according to female

[86] There are additional questions regarding fertility preferences, such as: "*Picking a number between 0 and 10 to show how likely is it that you will have a child in the future*". However, as shown in Panels (a) and (b) of Figure D.1 in Appendix D, the results remain consistent, with similar distributions across responses.

[87] While some participants consistently reported their fertility preferences across all survey waves, others did not provide this information in certain waves, resulting in missing values. To address this, we assume that for respondents with the missing value, their reported fertility preference remains constant until the birth of another child. We imputed the missing data with the average fertility preference for the same birth parity. Despite this assumption, approximately 69% of observations are lost, primarily because 54% of couples cannot be matched, making it difficult to identify the gap in fertility preferences.

adherence to traditional gender attitudes.[88,89] Prior research suggests that this may be linked to the concept of 'gender expectation', in which women, even when holding greater bargaining power, conform to conventional gender roles by aligning their fertility behavior with societal expectations. Column (2) supports this explanation: female controller households exhibit higher hazard rates (coefficient = 0.049; $p < 0.05$) when women adhere strongly to traditional gender attitudes, despite having lower fertility preferences than their male partners.[90] However, we do not observe a significant effect when bargaining power is measured by shared wage in column (2), implying that intangible bargaining power may be more sensitive to the influence of gender attitudes in shaping fertility behavior.

Table 3. 3: The average marginal effect of shared wage and financial control on birth interval through the bargaining power mechanism

	male's preference > female's		male's preference = female's		male's preference < female's	
	Full	High attitudes	Full	High Pref.	Full	Part-time job
	(1)	(2)	(3)	(4)	(5)	(6)
Shared wage						
Equal vs. Male	-0.019***	-0.009	-0.006	-0.008	-0.010	0.013
	(0.006)	(0.013)	(0.006)	(0.025)	(0.009)	(0.018)
Female vs. Male	-0.008	-0.027	-0.007**	-0.004	0.000	0.019**
	(0.008)	(0.016)	(0.003)	(0.031)	(0.007)	(0.009)
Financial control						
Equal vs. Male	-0.016***	0.011	-0.001	-0.004	-0.004	0.016
	(0.005)	(0.012)	(0.005)	(0.015)	(0.010)	(0.013)
Female vs. Male	0.007	0.049**	-0.003	-0.038	-0.009	0.018
	(0.009)	(0.024)	(0.009)	(0.036)	(0.015)	(0.020)
N	6,444	642	23,170	4,310	6,585	3,615
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Household RE	Yes	Yes	Yes	Yes	Yes	Yes

88 Gender norms are measured using by two the questions "*Mothers who don't really need the money shouldn't work*"; and "*It is better for everyone involved if the man earns the money and the woman takes1 care of the home and children*". Item 1 concern women's employment and earning roles, while Item 2 relate to family versus market roles. Given that, among women with high gender attitudes, the two items exhibit similar bargaining effects on the birth interval, we combine them into a single index using factor analysis to facilitate interpretation.

89 Based on the distribution (see Figure G.3.2 in Appendix G), we divide the sample into low and high gender attitude groups using a cut-off value of approximately 0.7.

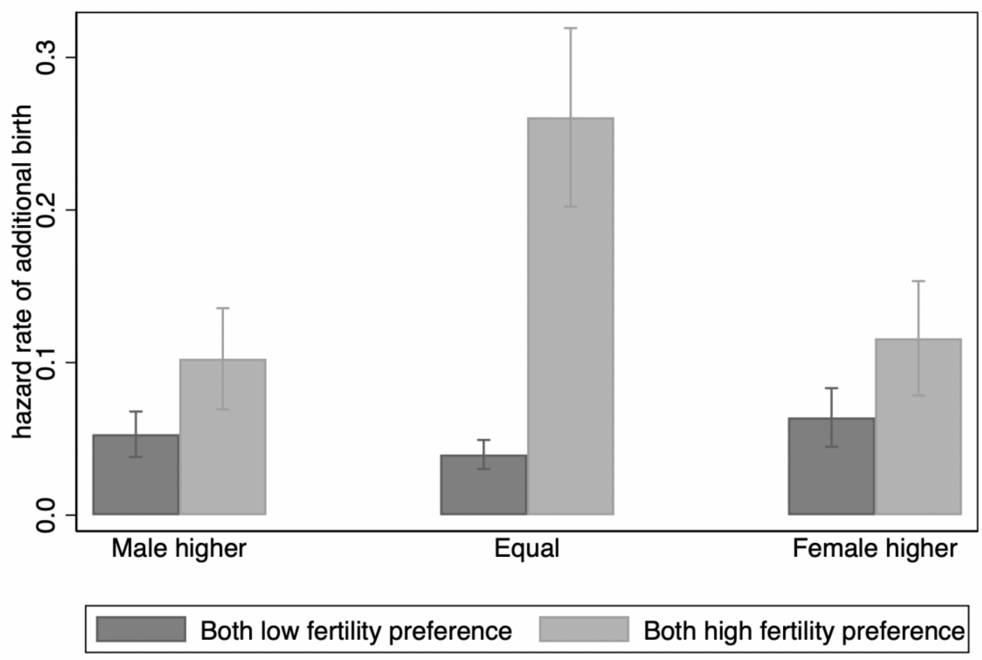
90 The smaller sample size in column (2) is due to the relatively low proportion of women with high gender norms in Australia. For more detail, see the distribution of female gender norms in Appendix G.

P & W & D FE	Yes	Yes	Yes	Yes	Yes	Yes

NOTE.—Estimates are from discrete-time survival regressions on the birth interval. In survival analysis, the dependent variable is a binary indicator reflecting whether a birth occurred within a given time interval. As coefficients in non-linear models are not directly comparable, we present average marginal effects instead. Column (1) reports the results for the subsample in which the male partner has a higher fertility preference than the female. Building on this, column (2) presents results for the further subset of these households in which the female exhibits high gender attitudes. Column (3) reports the results for couples with equal fertility preferences, and column (4) further narrows this to couples in which both partners have high fertility preferences. Column (5) focuses on the subsample where the female partner has a higher fertility preference than the male, while column (6) further refines this group to those in which the female partner is employed part-time. The full sample includes only 36,199 observations, as 54% of couples cannot be matched, making it difficult to identify the gap in fertility preferences. The complete set of controls includes average of male and female values for age, education, and self-rated health, as well as marital status, and household income. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

For couples with equal fertility preferences, bargaining theory provides limited explanatory power, as there is no preference heterogeneity to be negotiated. In such cases, fertility decisions are more likely to reflect joint decision-making, mutual goals, relationship quality, or shared expectations about family life. Column (3) supports this perspective, showing that the majority of cases exhibit no significant differences in hazard rates across the three household types, whether using tangible bargaining power (male breadwinner, equal breadwinner, and female breadwinner) or intangible bargaining power (male controller, equal controller, and female controller) as the measures. These results remain robust even when we narrow the sample to couples with high fertility preferences, where again no significant differences in hazard rate are observed across household types (see column 4). Nonetheless, the hazard rate for couples with identical high fertility preferences is 0.25 ($p < 0.01$), which is approximately 1.5 times higher than that of couples with heterogeneous fertility preferences (see Figure 3.2). This aligns with previous research suggesting that fertility intentions are more likely to be realized when partners share similar goals and values, which reinforces the importance of preference alignment over bargaining dynamics in shaping fertility behavior (Thomson, 1997; Miller et al., 2004).

Figure 3. 2: Predicated hazard rate by Couples' Preference Gap and Fertility Preference Intensity



Note: Estimates are derived from discrete-time survival regressions on the birth interval, categorized by the fertility preference gap between male and female partners. The results report the predicted hazard rate. "Male higher" refers to the subsample where the male partner has a higher fertility preference, "Equal" refers to couples with matching fertility preferences, and "Female higher" refers to cases where the female partner's fertility preference is higher. Within each subsample, we further divide couples into "low" (both partners have low fertility preferences) and "high" (both have high fertility preferences). The full sample only includes 36,199 observations due to 54% of couples cannot be matched, making it hard to identify the gap in fertility preferences. The analysis includes the full set of controls as outlined in equation (2), with confidence intervals set at the 90% level. The y-axis represents the hazard rate. The complete set of controls includes average of male and female values for age, education, and self-rated health, as well as marital status, and household income.

Lastly, for women with higher fertility preferences than their male partners, column (5) shows no significant differences across the three types of households, which contradicts the predictions of bargaining power theory. According to this theory, women with greater bargaining power and higher fertility preferences should have shorter birth intervals. However, beyond bargaining power, career pursuits and time allocation play a crucial role in shaping the fertility outcomes. In female breadwinner households, women may face the challenge of balancing career demands with the time required for childcare, leading to high motherhood penalty. As a result, for those women who pursue careers, even when they have higher fertility preferences and greater bargaining power, the demands of career could result in longer birth intervals. Despite this, we still expect the bargaining theory to hold for the part-time women, as they experience lower family-career conflict. The results in column (6) support for this argument, focusing on the subsample of households where women work part-time. In this group, the hazard rates increase when women have more tangible or intangible bargaining power. This suggests that when women have higher fertility preferences and greater bargaining power, they experience a shorter birth interval. Nonetheless, we do not observe a significant effect when bargaining power is measured by financial control in column (6), implying that tangible bargaining power may be more sensitive to the influence of career development in shaping fertility behavior.

3.5 Conclusion

Our findings highlight the complex ways in which intra-household bargaining power influences birth interval in a developed-country. Drawing on longitudinal Australian data and a discrete-time survival model, we show that both tangible and intangible bargaining powers play pivotal roles in determining birth intervals. These results validate fundamental insights from collective bargaining theory (Chiappori, 1992), which holds that the partner with greater bargaining power is more likely to achieve their fertility preferences. However, the impact of bargaining power on fertility is not linear: couples with roughly equal bargaining power experience the longest birth intervals. This suggests that in egalitarian households, decision-making may involve more protracted negotiation or mutual career investment, delaying childbearing. Given the growing prevalence of dual-earner and equal-sharing partnerships in developed contexts like Australia, these dynamics may help explain broader trends toward later and fewer births.

Moreover, we identify two groups of women whose behavior deviates from traditional bargaining model predictions. First, career-oriented women who have high tangible bargaining power and stronger fertility desires than their partners do not accelerate childbearing as classical bargaining models would predict. Instead, career considerations, especially the high costs of motherhood, take precedence and lead these women to postpone births. This finding provides empirical support for Hener's (2015) theoretical model, which argues that women with greater bargaining power may delay childbearing when career costs are high. Second, we uncover a paradox among norm-conforming women who hold greater bargaining power but lower fertility preferences than their partners. Under classical bargaining logic, these women should be able to prevent or delay additional births. However, our empirical results show that they have shorter birth intervals. Qualitative research on 'gender expectation' suggests that traditional role expectations can compel even empowered women to conform to societal or gender norms around motherhood. In these cases, cultural pressures blunt the impact of women's bargaining advantages, driving them to act against their own fertility intentions. This finding reinforces the importance of situating intra-household power within its normative context, an insight stressed by feminist economists (Agarwal, 1997) but often absent in formal bargaining frameworks. To our knowledge, this is the first study to systematically show how the motherhood penalty and gender norms can influence a woman's autonomy in fertility decisions, even in developed countries, egalitarian context. By pinpointing both career-driven postponements and norm-driven accelerations of childbearing, we shed new light on the fertility puzzle, explaining why greater female bargaining power does not always produce the fertility outcomes classical bargaining models predict.

We also find that nearly half of the couples in our Australian sample report aligned fertility preferences. In these cases, bargaining power has no significant effect on birth interval. Indeed, when preferences coincide, couples make fertility decisions as a unit, consistent with the classic unitary model (Rasul, 2008; Riederer et al., 2019; Hener, 2015). Importantly, the content of the aligned preference matters: couples who both strongly desire children tend to have children more rapidly than those with heterogeneous preferences, whereas couples with low fertility preferences exhibit the longest birth intervals. By empirically documenting the prevalence and impact of preference alignment in a developed country, we demonstrate that traditional bargaining models alone are insufficient to explain fertility behavior and must be complemented by unitary decision-making frameworks.

In conclusion, our study explains why traditional bargaining models only partially account for fertility behavior and offers a more nuanced framework. By identifying when women's bargaining power does or does not translate into actual childbearing, we move the literature on household decision making and fertility beyond a dimensional model. Although we validate core predictions of collective bargaining theory, we also expose its limits: in contemporary developed societies the broader context of bargaining, characterized by egalitarian arrangements, societal norms and career related penalties, can override individual fertility preferences.

However, one important limitation of this study concerns the potential for reverse causality in the observed relationship between women's relative wage and fertility behavior. While we interpret relative wage as a proxy for bargaining power within the household, it is possible that women anticipating future childbearing intentionally select into lower-earning or more flexible jobs that facilitate work–life balance. This preemptive adjustment in labor market behavior could reduce their measured bargaining power in the following cohabiting/married duration, thereby confounding the direction of causality even we control for the average fertility preference. In addition, fertility decisions may be influenced not only by relative wage shares but also by individuals' absolute wage levels. If absolute wage directly affects the affordability of an additional child, omitting it from the main specification may lead to an overstatement of bargaining-related effects. However, directly controlling for individuals' absolute wage is not straightforward, as absolute wage is embedded in the construction of wage shares and highly collinear with relative measures. Including absolute income may therefore absorb part of the variation that reflects bargaining power itself, obscuring the interpretation of relative wage effects and potentially leading to over-control bias. Moreover, the analysis implicitly assumes that non-response is conditionally random given observed covariates. If survey non-response is systematically related to unobserved characteristics that are also correlated with fertility behavior or intra-household bargaining dynamics, the estimates may be subject to selection bias. To address this limitation, future research will consider employing instrument variable (IV) method or experimental designs to better capture the causal effect of intra-household bargaining power on fertility decisions. Additionally, exploring other factors such as partners' bargaining power, policy interventions, and workplace flexibility could offer a more comprehensive understanding of the interplay between fertility behavior and women's labor market outcomes.

References

Agarwal, Bina. 1997. "Bargaining" and Gender Relations: Within and Beyond the Household." *Feminist Economics* 3(1):1–51.

Agarwala, Rina, and Scott M. Lynch. 2006. "Refining the Measurement of Women's Autonomy: An International Application of a Multi- Dimensional Construct." *Social Forces* 84(4):2077–98.

Allendorf, Keera. 2007. "Do Women's Land Rights Promote Empowerment and Child Health in Nepal?" *World Development* 35(11):1975–88.

Ambrosetti, Elena, Aurora Angeli, and Marco Novelli. 2021. "Childbearing Intentions among Egyptian Men and Women: The Role of Gender-Equitable Attitudes and Women's Empowerment." *Demographic Research* 44:1229–70.

Anderson, Siwan, and Chris Bidner. 2015. "Property Rights over Marital Transfers *." *The Quarterly Journal of Economics* 130(3):1421–84.

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review* 88(3):450–77.

Antman, Francisca M. 2014. "Spousal Employment and Intra-Household Bargaining Power." *Applied Economics Letters* 21(8):560–63.

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.

Bankole, Akinrinola, and Susheela Singh. 1998. "Couples' Fertility and Contraceptive Decision-Making in Developing Countries: Hearing the Man's Voice." *International Family Planning Perspectives* 24(1):15.

Becker, Gary S. 1960. *An Economic Analysis of Fertility*. Cambridge, Mass: National Bureau of Economic Research.

Becker, Gary S. 1981. *A Treatise on the Family*. beck81-1. National Bureau of Economic Research.

Blundell, Richard, Pierre-André Chiappori, and Costas Meghir. 2005. “Collective Labor Supply with Children.” *Journal of Political Economy* 113(6):1277–1306.

Brodmann, S., G. Esping-Andersen, and M. Guell. 2007. “When Fertility Is Bargained: Second Births in Denmark and Spain.” *European Sociological Review* 23(5):599–613.

Chiappori, Pierre-André. 1992. “Collective Labor Supply and Welfare.” *Journal of Political Economy* 100(3):437–67.

Clarke, Damian. 2018. “CHILDREN AND THEIR PARENTS: A REVIEW OF FERTILITY AND CAUSALITY.” *Journal of Economic Surveys* 32(2):518–40.

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

Doss, Cheryl. 2013. “Intrahousehold Bargaining and Resource Allocation in Developing Countries.” *The World Bank Research Observer* 28(1):52–78.

Duflo, Esther, and Christopher Udry. 2004. “Intrahousehold Resource Allocation in Cote d’Ivoire: Social Norms, Separate Accounts and Consumption Choices.”

Erfani, Amir, and Leandra Pilon. 2025. “Gender Equality in the Division of Housework and Fertility Intentions in Canada: The Moderating Effect of Employment and Education.” *Journal of Family Issues* 46(5):784–804.

Feldman, Becca S., Alan M. Zaslavsky, Majid Ezzati, Karen E. Peterson, and Marc Mitchell. 2009. “Contraceptive Use, Birth Spacing, and Autonomy: An Analysis of the *Oportunidades* Program in Rural Mexico.” *Studies in Family Planning* 40(1):51–62.

Gipson, Jessica D., and Michelle J. Hindin. 2009. “The Effect of Husbands’ and Wives’ Fertility Preferences on the Likelihood of a Subsequent Pregnancy, Bangladesh 1998–2003.” *Population Studies* 63(2):135–46.

Hener. 2015. “Do Couples Bargain over Fertility?” *Annals of Economics and Statistics*

(117/118):211.

Hochschild, Arlie Russell. 1989. *The Second Shift : Working Parents and the Revolution at Home /*. New York, N.Y. : Viking,.

Hoddinott, John, and Lawrence Haddad. 1995. “DOES FEMALE INCOME SHARE INFLUENCE HOUSEHOLD EXPENDITURES? EVIDENCE FROM CÔTE D’IVOIRE.” *Oxford Bulletin of Economics and Statistics* 57(1):77–96.

Hogan, D. P., B. Berhanu, and A. Hailemariam. 1999. “Household Organization, Women’s Autonomy, and Contraceptive Behavior in Southern Ethiopia.” *Studies in Family Planning* 30(4):302–14.

Kabeer, Naila. 1999. “Resources, Agency, Achievements: Reflections on the Measurement of Women’s Empowerment.” *Development and Change* 30(3):435–64.

Kan, Man-Yee, Ekaterina Hertog, and Kamila Kolpashnikova. 2019. “Housework Share and Fertility Preference in Four East Asian Countries in 2006 and 2012.” *Demographic Research* 41:1021–46.

Klawon, Emily, and Jill Tiefenthaler. 2001. “Bargaining over Family Size: The Determinants of Fertility in Brazil.” *Population Research and Policy Review* 20(5):423–40.

Lundberg, Shelly, and Robert A. Pollak. 1993. “Separate Spheres Bargaining and the Marriage Market.” *Journal of Political Economy* 101(6):988–1010.

Lundberg, Shelly, and Robert A. Pollak. 1996. “Bargaining and Distribution in Marriage.” *Journal of Economic Perspectives* 10(4):139–58.

Luz, Luciana, and Victor Agadjanian. 2015. “Women’s Decision-Making Autonomy and Children’s Schooling in Rural Mozambique.” *Demographic Research* 32:775–96.

Manser, Marilyn, and Murray Brown. 1980. “Marriage and Household Decision-Making: A Bargaining Analysis.” *International Economic Review* 21(1):31.

Mason, Karen Oppenheim. 1986. “The Status of Women: Conceptual and Methodological Issues in Demographic Studies.” *Sociological Forum* 1(2):284–300.

Mason, Karen Oppenheim, and Herbert L. Smith. 2000. “Husbands’ versus Wives’ Fertility Goals and Use of Contraception: The Influence of Gender Context in Five Asian Countries.” *Demography* 37(3):299–311.

Mincer, Jacob. 1963. “Market Prices, Opportunity Costs, and Income Effects.” *Measurement in Economics : Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*.

Musick, K., P. England, S. Edgington, and N. Kangas. 2009. “Education Differences in Intended and Unintended Fertility.” *Social Forces* 88(2):543–72.

Rammohan, Anu, and Meliyanni Johar. 2009. “The Determinants of Married Women’s Autonomy in Indonesia.” *Feminist Economics* 15(4):31–55.

Rasul, Imran. 2008. “Household Bargaining over Fertility: Theory and Evidence from Malaysia.” *Journal of Development Economics* 86(2):215–41.

Reggio, Iliana. 2011. “The Influence of the Mother’s Power on Her Child’s Labor in Mexico.” *Journal of Development Economics* 96(1):95–105.

Riederer, Bernhard, Isabella Buber-Ennser, and Zuzanna Brzozowska. 2019. “Fertility Intentions and Their Realization in Couples: How the Division of Household Chores Matters.” *Journal of Family Issues* 40(13):1860–82.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 1980. “Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models.” *Journal of Political Economy* 88(2):328–48.

Sarnak, Dana O., and Alison Gemmill. 2022. “Perceptions of Partners’ Fertility Preferences and Women’s Covert Contraceptive Use in Eight Sub-Saharan African Countries.” *Studies in Family Planning* 53(3):527–48.

Schultz, T. Paul. 1990. “Testing the Neoclassical Model of Family Labor Supply and Fertility.” *The Journal of Human Resources* 25(4):599–634.

Smith. 1999. “Female Autonomy and Fertility in Five Asian Countries.” *PSC Research Report* 04-556.

Steele, Fiona, Harvey Goldstein, and William Browne. 2004. “A General Multilevel

Multistate Competing Risks Model for Event History Data, with an Application to a Study of Contraceptive Use Dynamics." *Statistical Modelling* 4(2):145–59.

Steele, Fiona, Constantinos Kallis, Harvey Goldstein, and Heather Joshi. 2005. "The Relationship between Childbearing and Transitions from Marriage and Cohabitation in Britain." *Demography* 42(4):647–73.

Thomas, Duncan. 1990. "Intra-Household Resource Allocation: An Inferential Approach." *The Journal of Human Resources* 25(4):635.

Thomas, Duncan. 1994. "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height." *The Journal of Human Resources* 29(4):950.

Tong, Yuying, Yiqing Gan, and Chunxue Zhang. 2024. "Whose Preference Matters More? Couple's Fertility Preferences and Realization in the Context of China's Two-Child Policy." *Journal of Family Issues* 45(2):471–501.

Upadhyay, Ushma D., and Michelle J. Hindin. 2005. "Do Higher Status and More Autonomous Women Have Longer Birth Intervals?" *Social Science & Medicine* 60(11):2641–55.

Varanasi, 2009., "Intrahousehold Bargaining and Fertility", unpublished manuscript, University of. Washington. 16.

Appendix

Appendix A: Constructing the discrete-time sample

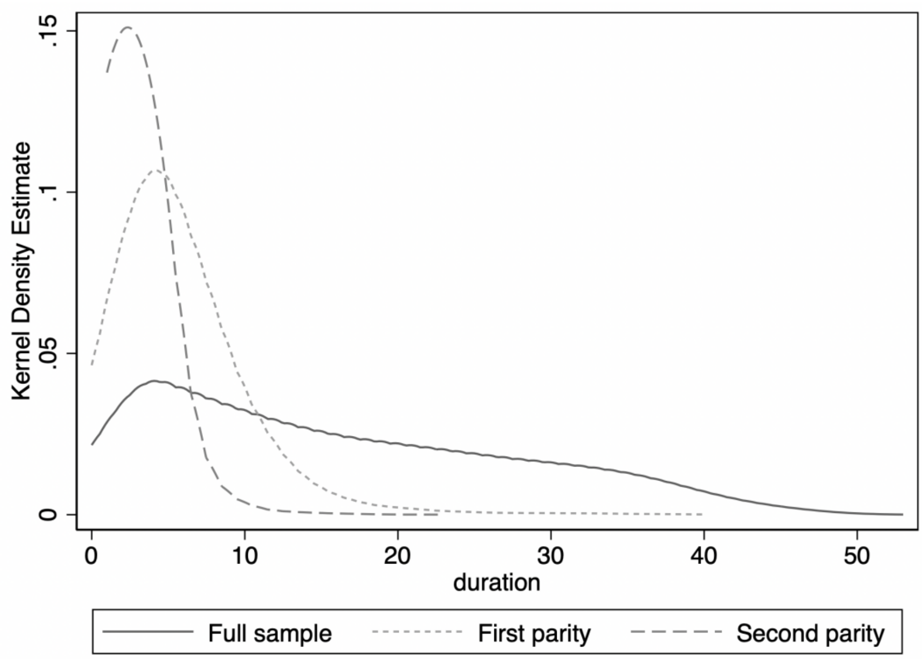
Birth interval. Birth interval is typically measured by the time between consecutive births. In our sample, pregnancy histories were not collected; however, the age of each child was provided, enabling the calculation of birth spacing based on the duration of marriage or cohabitation and the ages of the children. Specifically, the first birth interval is calculated by subtracting the age of the first child from the total duration of marriage or cohabitation. Subsequent birth intervals are derived from the age gaps between successive children. Table A.3.1 presents the mean values of the birth intervals, with the full sample showing an average duration of 15.9 years. Among this, the first birth interval averages 5.7 years, while the second birth interval averages 2.5 years. Figure A.3.1 illustrates the Kernel Density Estimation (KDE) of the birth intervals, showing that the second birth interval is more concentrated around shorter durations, while the first birth interval displays a longer duration.

Table A.3. 1: Mean value of the birth interval

	Full sample	First parity	Second parity
Birth interval	15.9	5.7	2.5

Note: The table presents the mean value of married or cohabiting duration in the full sample, subsample of first birth interval, and subsample of second birth interval.

Figure A.3. 1: The Kernel Density Estimation of the birth interval



Note: The figure presents the Kernel Density Estimation (KDE) of married or cohabiting duration, first birth interval, and second birth interval. The x-axis represents the proportion of share of earning and financial control within the household, ranging from 1 to 55. The y-axis represents the kernel density estimate, which illustrates the distribution of these variables across the sample.

Once the birth intervals are constructed, we expand our sample based on these intervals. Following the approach outlined in previous research (Steele et al., 2004; 2005), we begin by replicating the initial wave information for each participant according to the duration of their cohabitation. We then generate a duration variable, $t = 1,2,3 \ldots$, where each value represents a distinct time period within the observation window. Afterward, we create an event variable to capture the occurrence of a birth, with the variable taking a value of 1 when the household has a newborn and 0 otherwise, representing the event of interest in the discrete-time survival model. Next, we generate the birth parity variable to indicate the number of births that have occurred within each household. This variable is critical for distinguishing between first, second, or subsequent births, allowing us to analyze birth intervals for different parity groups separately.

However, unlike previous research that primarily relies on cross-sectional or retrospective fertility histories, our dataset is longitudinal. In addition to the initial wave information, we incorporate data from subsequent waves, which enables us to more effectively address the issue of right censoring. Specifically, we combine information from all available follow-up waves with the discrete-time survival dataset constructed from the initial wave. To do this, we update the duration variable to continue across waves, ensuring that duration time t is incremented sequentially rather than restarting with each wave. The event variable is updated. If a birth occurs during a later wave, the event variable is set to 1 in the corresponding time period; otherwise, it remains 0. Moreover, the birth parity variable is also updated dynamically. This method follows similar practices used in longitudinal fertility studies (e.g., Kravdal, 2001), but with the advantage of capturing time-varying covariates and events across multiple observation points. By integrating follow-up waves into the survival dataset, we can reduce right censoring and obtain more accurate estimates of the timing and likelihood of subsequent births.

Appendix B: Correlation analysis

Table B.3.1 presents the correlation coefficients among four proxies of tangible bargaining power: share of earning (SMI), shared working hours (SWH), shared household working hours (SHWH), and education gap. The strongest correlation is observed between SW and SWH (0.71), indicating that couples with a more equal division of monthly income also tend to share working hours more evenly. In contrast, correlations between SHWH and the other variables are relatively weak, suggesting that shared household working hours may capture a different dimension of bargaining power.

Table B.3. 1: Correlation analysis of tangible bargaining power variables

	SoE	SWH	SHWH	Edu gap
SoE	1			
SWH	0.71	1		
SHWH	0.13	0.14	1	
Edu gap	0.23	0.17	0.004	1

Note: This table presents the correlation coefficients among five proxies for bargaining power: share of earning (SoE), shared working hours (SWH), shared household working hours (SHWH), and education gap.

There are three financial decision-making questions, such as (1) *making large household purchases*. (2) *managing day-to-day spending and paying bills*, (3) *saving, investment and borrowing*. Participants could select from: (1) *always me*, (2) *usually me*, (3) *decisions shared equally with my partner*, (4) *usually my partner*, or (5) *always my partner*. Table B.3.2 shows that the correlations among the three financial decision-making variables, large purchases, bill payments, as well as investment and saving, are relatively low, ranging from 0.37 to 0.49. Moreover, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is 0.66, which falls below the commonly accepted threshold of 0.70 for reliable factor analysis.

Table B.3. 2: Correlation analysis of three financial decision-making variables

	Large buy	Bills	Saving
Large buy	1		
Bills	0.37	1	
Saving	0.48	0.49	1
KMO	**0.66**		

Note: Table presents the correlation analysis of three financial decision-making variables: large purchases, bill payments, and savings. Kaiser-Meyer-Olkin (KMO) is a measure of sampling adequacy that indicates the proportion of variance among variables that might be common variance.

Appendix C: Control variables

We control for several socio-demographic characteristics at the individual and household levels, such as age, education, household income, household size, and self-rated health. All these characteristics are expected to influence fertility intentions according to previous literature and empirical evidence (Upadhyay and Hindin, 2005; Seebens, 2006; Feldman et al., 2009). Given the potential for a nonlinear relationship between age and birth interval, we incorporate age squared. To control for family characteristics, we first add household income, expressed on a logarithmic scale. Next, we include the number of household members. Self-rated health status is measured using an ordered variable ranging from 1 to 5. Nonetheless, financial management may overlap with financial control in influencing birth intervals. To isolate this effect, we include two additional controls: paying bills and saving and investment activities. Specifically, Table C.3.1 shows that the mean proportion of women responsible for paying bills is 0.589, with a significant difference of -0.08 between households with children and without children. Conversely, saving and investment responsibilities show a smaller yet significant positive difference (0.018) between households with and without children, indicating that financial control in long-term savings may operate differently from routine financial tasks.

Table C.3. 1: Descriptive Statistics of control variables

	All households		Without child		With child		
Variables	Mean	SD	Mean	SD	Mean	SD	Diff
Financial management							
Bills	.589	.395	.528	.365	.614	.404	-0.08***
Saving	.449	.306	.462	.299	.444	.308	0.018***
Personal Chara.							
Age (year)	39.5	10.4	32.535	9.751	42.36	9.398	-9.82***
Education (1-9)	2.738	1.621	3.115	1.372	2.584	1.689	0.531***
Health (unhealth)	2.492	.761	2.394	.702	2.533	.778	-0.13***
Num. Child	1.642	1.359	-	-	-	-	-

One Child					0.15	0.002	
Two child					0.31	0.002	
Three Child					0.24	0.002	
Family Chara.							
Household Income	11.4	.58	11.439	.59	11.51	.574	-0.07***
Siblings	4.61	2.86	4.894	2.746	4.502	2.899	0.391***

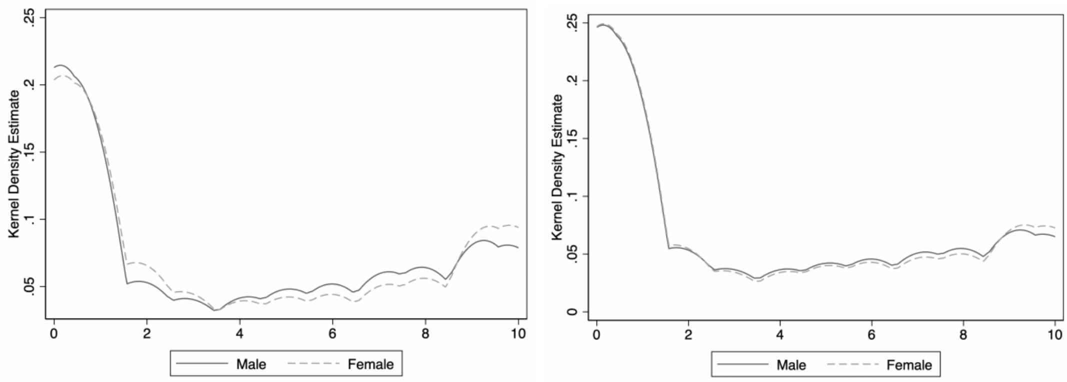
NOTE.—Data is sourced from the HILDA Survey, covering the period from 2002 to 2020. The table presents descriptive statistics for shared wage and financial control across different samples: All households, With child, and Without child. Shared wage refers to the monthly wage from the primary job, calculated as the female monthly wage divided by the total couple monthly wage. Financial control is measured based on three family financial decision-making questions.

Appendix D: Fertility preferences

There are two questions used to measure fertility preferences: a) "*How likely are you to have a child/more children in the future?*" and b) "*Is it likely that you will have a child in the future?*". Both questions are answered on a scale from 0 to 10, where higher values indicate stronger childbearing intentions. As shown in Figure D.3.1, the distributions of male and female fertility preferences are broadly similar for both questions. In particular, male and female responses show a high density around zero, with slight differences in the upper tail, suggesting that extreme preferences are not strongly gendered in this context.

Figure D.3. 1: Distribution of male and female child preferences

a) Q1 b) Q2

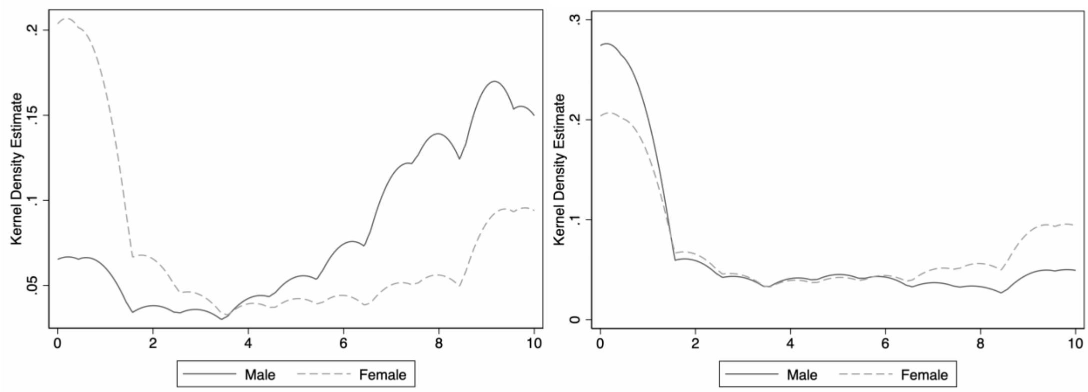


Note: The Figure presents the Kernel Density Estimation (KDE) of male and female child preferences. The x-axis represents the self-ranking child preference, ranging from 0 to 10. The y-axis represents the kernel density estimate, which illustrates the distribution of these variables across the sample. The solid line corresponds to the distribution of male child preference, while the dashed line represents female child preference.

Figure D.3.2 shows the distribution of male and female child preferences by birth parity. In panel (a), women display a higher density at zero and lower overall fertility preferences for the first parity compared to men, whereas in panel (b), men show a weaker preference for subsequent children. Overall, men tend to express higher fertility preferences than women for the first birth, while women exhibit stronger fertility preferences than men for subsequent births.

Figure D.3. 2: Distribution of male and female fertility preferences by birth parity

a) First b) Subsequence

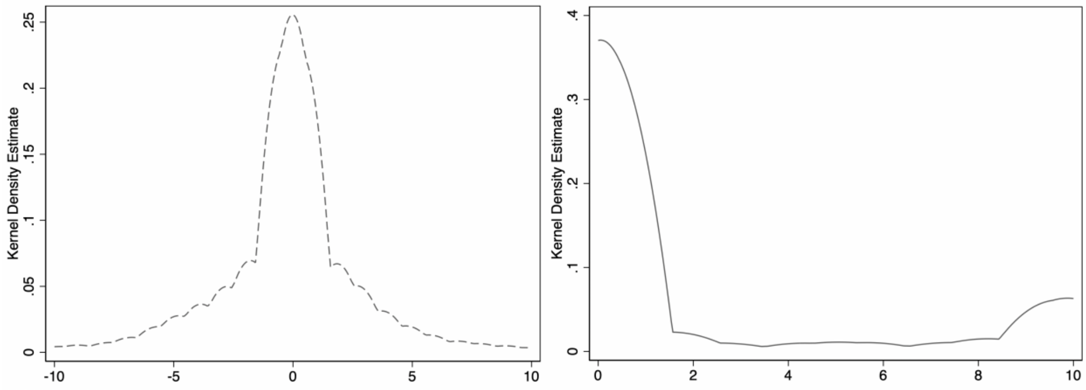


Note: The Figure presents the Kernel Density Estimation (KDE) of male and female child preferences by birth parity. The x-axis represents the self-ranking child preference, ranging from 0 to 10. The y-axis represents the kernel density estimate, which illustrates the distribution of these variables across the sample. The solid line corresponds to the distribution of male child preference, while the dashed line represents female child preference.

After constructing the data at the household level, we can calculate the fertility preference gap between men and women. Specifically, the fertility preference gap is computed by subtracting the women's fertility preference from the men's fertility preference. As shown in Panel (a) Figure D.3.3, the distribution of the fertility preference gap is highly concentrated around zero, indicating that most couples exhibit similar fertility preferences. Panel (b) of Figure D.3.3 further illustrates the fertility preferences among couples with equal preferences, revealing that the majority of these couples report relatively low desired children.

Figure D.3. 3: Distribution of gap in fertility preferences

a) b)



Note: The panel (a) presents the Kernel Density Estimation (KDE) of gap of fertility preference between couples in household level, with ranging from -10 to 10 in x-axis. The panel (b) presents the KDE of fertility preferences among couples with equal preferences, with ranging from 0 to 10. The y-axis represents the kernel density estimate, which illustrates the distribution of these variables across the sample.

Appendix E: Descriptive statistics of different type of households

Table E.3.1 presents descriptive statistics from the HILDA Survey, highlighting shifts in shared wage and financial control between households before and after 2014. The proportion of male breadwinner households decreased from 46.1% to 43.3% ($p < 0.001$), while female breadwinner households increased from 21.6% to 22.5% ($p < 0.01$). In terms of financial control, male controller households remained stable, while female controller households increased from 6.3% to 8% ($p < 0.001$), alongside a slight decline in equal controller households. These findings suggest a gradual rise in female economic and decision-making power within Australian households during the study period. Notably, equal breadwinner and equal controller households together constitute a significant proportion of the sample. Particularly, the proportion of equal controller households has increased over time, reflecting an ongoing trend toward more egalitarian household arrangements. However, the slight decline in equal breadwinner households may indicate persistent gender disparities in earning capacity or shifts in household income dynamics.

Table E.3. 1: Descriptive Statistics of shared wage and financial control by year

	Year <= 2014		Year > 2014		
Variables	Mean	SD	Mean	SD	Diff
Shared wage					
Male	.461	.498	.433	.495	0.028***
Equal	.323	.468	.342	.474	-0.019***
Female	.216	.412	.225	.418	-0.009**
Financial control					
Male	.196	.397	.19	.393	0.00500
Equal	.741	.438	.73	.444	0.011***
Female	.063	.244	.08	.271	-0.016***

NOTE.—Data are sourced from the HILDA Survey, covering the period from 2005 to 2020. The table reports descriptive statistics on the probability of having at least one child across different household types, categorized by shared wage (male breadwinner, equal breadwinner, and female breadwinner) and financial control (male controller, equal controller, and female controller), for three groups: all households, households with years up to and including 2014, and households with years after 2014. Shared wage is based on annual income from the primary job, calculated as the female's share of the couple's total monthly wage. Financial control is determined by who makes decisions regarding major family financial purchases.

Appendix F: Money management

There are two additional decision-making variables related to finance, such as paying bills, as well as investment and saving, but we consider that they may not fully reflect the concept of financial control, as paying bills may simply reflect routine financial management rather than a true exercise of control, while saving decisions might be influenced by broader financial goals or external factors unrelated to intra-household power dynamics. As Table F.3.1 displayed, similar to the effect of involvement in household chores in previous research (Kan et al., 2019; Erfani and Pilon, 2025), when women engage more in pay bills activity, their bargaining power is diminished, resulting in shorter birth intervals. On the other hand, saving decisions might be influenced by broader financial goals or external factors unrelated to intra-household power dynamics, and consequently, they have an insignificant effect on the hazard rate.

Table F.3. 1: The effect of financial management on the birth interval

	All	First	Second	Third plus
	(1)	(2)	(3)	(4)
Pay bills				
Equal vs. Male	0.007***	0.009	-0.003	0.002
	(0.002)	(0.008)	(0.009)	(0.003)
Female vs. Male	0.012***	0.018**	0.007	-0.003
	(0.002)	(0.008)	(0.010)	(0.003)
Saving and investment				
Equal vs. Male	-0.001	0.007	-0.020**	-0.002
	(0.002)	(0.008)	(0.009)	(0.003)
Female vs. Male	-0.006*	-0.001	-0.010	-0.006
	(0.003)	(0.011)	(0.013)	(0.004)
N	121,478	34,220	23,823	61,160
Controls	Yes	Yes	Yes	Yes
Household & time RE	Yes	Yes	Yes	Yes
Parity & wave FE	Yes	Yes	Yes	Yes

NOTE.—Estimates from discrete-time survival regressions on the birth interval. In survival analysis, the dependent variable is a binary variable indicating whether an event has occurred or not within a time interval. The sample includes 7,075 households with women aged 16-49 and 118,011 observations. The complete set of controls includes average of male and female values for age, education, self-rated health, and fertility preference, as well as marital status, and household income. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Appendix G: Gender attitudes

To examine the impact of gender attitudes on fertility decisions, this study used a module from the survey that measures individuals' attitudes towards parenting and work. The module included three questions. The questions include "*Mothers who don't really need the money shouldn't work*"; and "*It is better for everyone involved if the man earns the money and the woman takes care of the home and children*". Participants are asked to rate their level of agreement on a seven-point scale ranging from "strongly disagree" to "strongly agree". The scores for all four questions were then added together to create a total score for each participant, with higher scores indicating more traditional gender attitudes. For easier interpretation, we normalize the summed gender attitudes score into a range of zero and one.

Item 1 concern women's employment and earning roles, while Item 2 relate to family versus market roles. As Table G.3.1 displayed, these two items measure the different dimension of gender attitude, as their correlation is only 0.52. Moreover, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is 0.68, which falls below the commonly accepted threshold of 0.70 for reliable factor analysis.

Table G.3. 1: Correlation analysis of two gender attitude variables

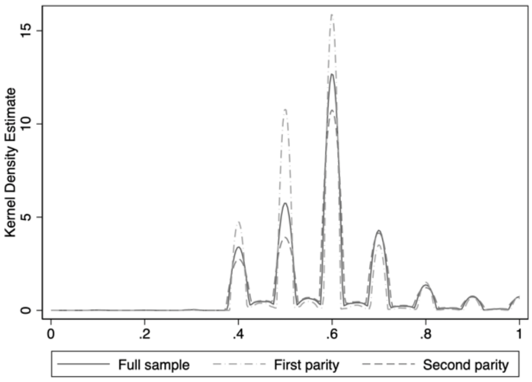
	employment and earning roles	family versus market roles
employment and earning roles	1	
family versus market roles	0.52	1
KMO	**0.68**	

Note: Table presents the correlation analysis of two gender attitudes variables: employment and earning roles, and family versus market roles. Kaiser-Meyer-Olkin (KMO) is a measure of sampling adequacy that indicates the proportion of variance among variables that might be common variance.

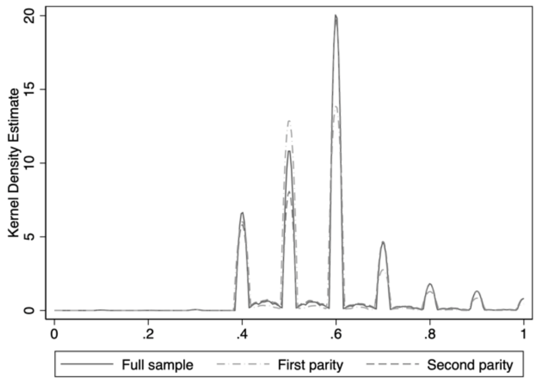
However, Items 1 and 2 exhibit similar distributions, with observations concentrated below 0.7. Based on these distributions (see Figure G.3.1 in Appendix G), we divide the sample into low and high gender-attitude groups using a cut-off value of approximately 0.7.

Figure G.3. 1: Distribution of gender attitudes of Items 1 and 2

a) Item 1



b) Item 2



Note: The Figure presents the Kernel Density Estimation (KDE) of male and female child preferences. The x-axis represents the self-ranking 'gender attitudes', ranging from 0 to 10. The y-axis represents the kernel density estimate, which illustrates the distribution of the variable across the sample.

Table G.3.2 presents the effects of shared wages and financial control on birth intervals among individuals with high gender attitudes. As shown in Columns (1) and (2), in households where men have higher fertility preferences than women, female-controller households exhibit higher hazard rates when women strongly adhere to traditional gender attitudes, despite women having lower fertility preferences than their male partners. This effect is observed regardless of whether high gender attitudes are measured by employment and earning roles (coefficient = 0.048; $p < 0.05$) or by family versus market roles (coefficient = 0.043; $p < 0.1$).

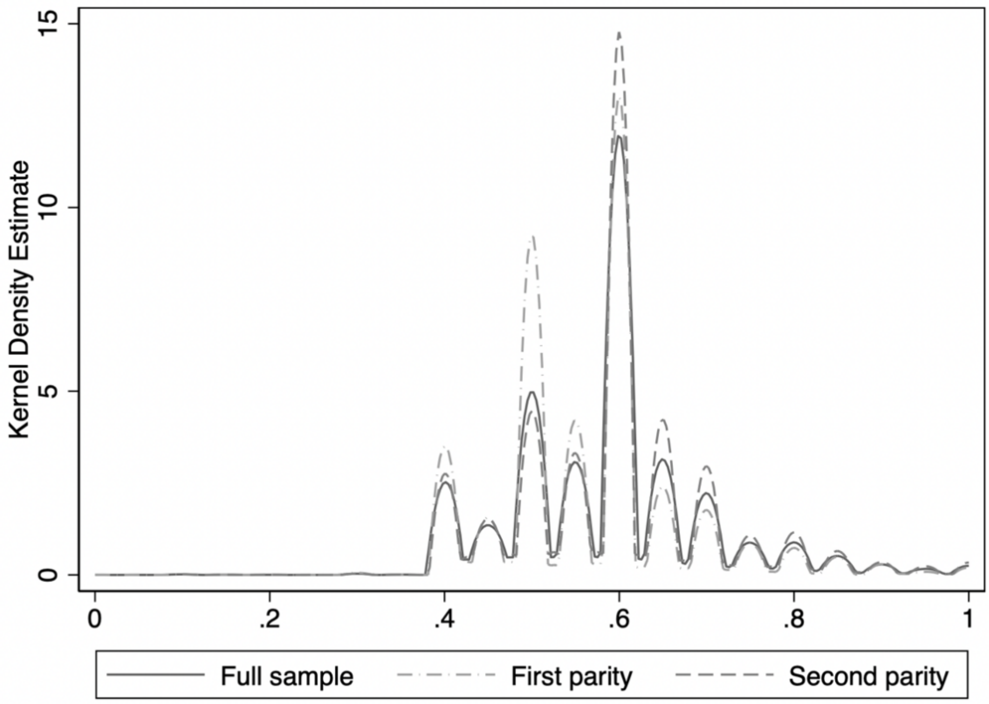
Table G.3. 2: The average marginal effect of shared wage and financial control on the birth interval among women with high gender attitude (men have higher preference than women)

	High employment and earning roles (1)	High family versus market roles (2)
Shared wage		
Equal vs. Male	-0.023*	0.004
	(0.012)	(0.013)
Female vs. Male	-0.052***	-0.013
	(0.016)	(0.017)
Financial control		
Equal vs. Male	0.004	0.014
	(0.012)	(0.013)
Female vs. Male	0.048**	0.043*
	(0.023)	(0.025)
N	3,164	2,940
Controls	YES	YES
Household RE	YES	YES
Parity & duration & wave FE	YES	YES
Money management	YES	YES

NOTE.—Estimates from discrete-time survival regressions on the birth interval. In survival analysis, the dependent variable is a binary variable indicating whether an event has occurred or not within a time interval. As coefficients in non-linear models are not directly comparable, we present average marginal effects instead. The sample includes 5,677 households with women aged 16-49 and 78,252 observations. The complete set of controls includes average of male and female values for age, education, self-rated health, and fertility preference, as well as marital status, and household income. Standard errors are presented in parentheses below the point estimates. *** p < 0.01.** p < 0.05.* p < 0.1.

Figure G.3.2 presents the distribution of women's gender attitudes. Similary, based on this distributions, we divide the sample into low and high gender-attitude groups using a cut-off value of approximately 0.7.

Figure G.3. 2: Distribution of gender attitudes

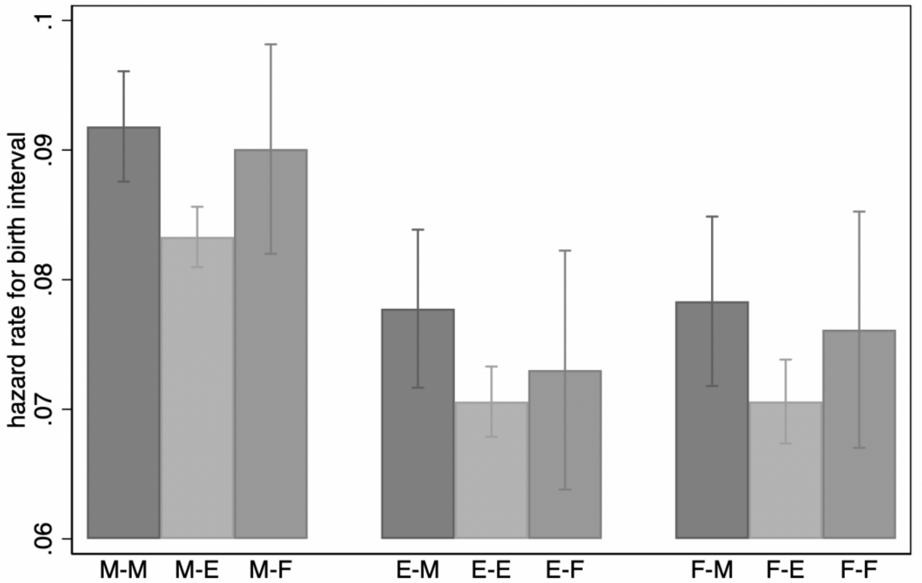


Note: The Figure presents the Kernel Density Estimation (KDE) of male and female child preferences. The x-axis represents the self-ranking 'gender attitudes', ranging from 0 to 10. The y-axis represents the kernel density estimate, which illustrates the distribution of the variable across the sample.

Appendix H: Interaction between shared wage and financial control

The result in Figure H.3.1 reveals that both M-M (male breadwinner and male controller) households have the highest hazard rate of having children, at approximately 0.092. However, the F-F (female breadwinner and female controller) households show a distinctly lower hazard rate, around 0.077.

Figure H.3. 1: The combination effect of shared wage and financial control on the hazard rate



Note: Estimates are from discrete-time survival regressions on the birth interval of the following household types: M-M (male breadwinner; male controller), M-E (male breadwinner; equal controller), M-F (male breadwinner; female controller), E-M (equal breadwinner; male controller), E-E (equal breadwinner; equal controller), E-F (equal breadwinner; female controller), F-M (female breadwinner; male controller), F-E (female breadwinner; equal controller), and F-F (female breadwinner; female controller). The complete set of controls includes average of male and female values for age, education, self-rated health, and fertility preference, as well as marital status, and household income, with confidence intervals set at the 90% level. The y-axis represents the hazard rate.

Appendix I: Dual-earner couples

Table I. 1: The average marginal effect of shared wage and financial control on the birth interval

	Baseline	Dual-earner	Zero wage couple
	(1)	(2)	(3)
Shared wage			
Equal vs. Male	-0.032***	-0.021***	
	(0.002)	(0.002)	
Female vs. Male	-0.009***	-0.023***	
	(0.003)	(0.004)	
Financial control			
Equal vs. Male	-0.007**	-0.006*	-0.004
	(0.003)	(0.003)	(0.007)
Female vs. Male	-0.003	-0.007	0.022*
	(0.005)	(0.006)	(0.012)
N	78,252	55,436	10,719
Controls	Yes	Yes	Yes
Household RE	Yes	Yes	Yes
Parity & duration & wave FE	Yes	Yes	Yes
Money management	Yes	Yes	Yes

NOTE.—Estimates from discrete-time survival regressions on the birth interval. Column (1) includes the full sample of couples, Column (2) restricts the sample to dual-earner couples, and Column (3) restricts the sample to couples in which both partners have zero wage income. In survival analysis, the dependent variable is a binary variable indicating whether an event has occurred or not within a time interval. As coefficients in non-linear models are not directly comparable, we present average marginal effects instead. The sample includes 5,677 households with women aged 16-49 and 78,255 observations. The complete set of controls includes average of male and female values for age, education, self-rated health, and fertility preference, as well as marital status, and household income. Standard errors are presented in parentheses below the point estimates. *** $p < 0.01$.** $p < 0.05$.* $p < 0.1$.

Conclusion

The collective findings of this thesis underscore the persistent and multifaceted nature of gender inequality, cutting across macro-level policies, household dynamics, and organizational practices. By examining diverse national contexts (China, Australia, and the UK) and multiple levels of analysis, the three chapters together reveal common themes about how structural factors continue to constrain women's opportunities in both the labor market and the family sphere. Even as economic and social conditions evolve, gender disparities adapt rather than disappear, indicating that deeply ingrained norms and institutional biases remain resilient. Whether the focus is on fertility planning, labor market participation, or the intersection of both, a unifying conclusion is that gender inequalities may vary over time and place, but they are difficult to overcome, often re-emerging in new forms even in the face of interventions meant to mitigate them.

At the macro level, Chapter 1 demonstrated how a national policy shift in China, the introduction of the two-child policy, though not explicitly aimed at gender issues, inadvertently reinforced existing gender disparities. More specifically, employers, followed by entrenched assumptions about women's familial roles, responded by adjusting their expectations and increasing discrimination against female employees. This contributed to a deepening motherhood wage penalty among certain groups of women, underscoring how policy reforms can produce unintended consequences when filtered through gendered institutional logics. At the meso (firm) level, Chapter 2 revealed that although workplace policies intended to provide flexibility and support, such as flexible working arrangements, they can exacerbate existing structural gender inequalities. Access to flexibility is stratified: men are more likely to benefit from "advantageous" forms of flexibility, such as flexi-time in high-skill roles, whereas women often rely on "constrained" flexibility, such as term-time or low-skill roles. This unequal uptake leads to divergent wage trajectories. For instance, the wage penalty associated with term-time work is substantially higher for women than the penalty men face when opting for flexi-time, highlighting how "well-meaning" working policies can reinforce or exacerbate gender wage gaps when underlying gender norms and occupational segregation remain unaddressed. At the micro (household) level, Chapter 3 illustrated that power dynamics and social norms continue to shape outcomes in ways that disadvantage women's individual preference, particular in fertility behavior, even within the intimate sphere of the family. Traditional intra-household bargaining models fell short in explaining observed fertility behavior, as they often neglected individual-level factors such as career orientation and societal pressures. For example, career-oriented women, even those with greater decision-making power at home, may delay childbearing due to anticipated career penalties, while norm-conforming women may choose early fertility despite having bargaining leverage, reflecting internalized expectations rather than strategic negotiations.

Together, the core contribution of these chapters is to offer a comprehensive account of how structural forces at the macro, meso, and micro levels interact to reproduce and reinforce gender inequality in labor and family fields. A common theme is that women's choices and outcomes are framed – and often limited – by the context in which they operate. Whether through national policy, workplace conditions, or the division of labor within households, women encounter barriers that men typically do not, and these constraints tend to be interdependent and mutually reinforcing. Another recurring insight is the flexibility and adaptability of gender disparities. Even as conditions improve or policies are reformed, inequality often shifts in form rather than being resolved. For example, internalized gender norms continue to influence women's labor market and family-related decisions, limiting their ability to fully take advantage of new opportunities. In the UK case, flexible working arrangements were intended to support work–life balance but ended up reinforcing wage gaps due to gendered patterns of use and persistent assumptions about caregiving. Similarly, China's relaxation of the one-child policy, although not aimed at gender reform, activated latent labor market biases that ultimately reduced women's earning potential. In essence, patriarchal norms and gendered divisions of labor maintain a persistent and pervasive influence, appearing in different forms across the domains explored in these chapters. This cross-cutting conclusion aligns with broader findings in gender inequality research: despite socioeconomic development and ongoing policy efforts, gender gaps tend to evolve rather than disappear. These findings highlight the need for sustained attention and multifaceted strategies to address the structural roots of gender inequality.

Despite its contributions, this thesis has several limitations that must be acknowledged. First, the scope of the analysis is confined to three countries (China, Australia, and the UK), each with its own cultural and institutional particularities. While these cases offer valuable insights, the findings may not be directly generalizable to other contexts, such as countries with different welfare regimes, labor market structures, or gender norms. For example, China's experience with a state-mandated fertility policy is unique, and other nations' policy changes might not yield the same employer responses. Similarly, Australia and the UK have their own social and policy environments; households and firms in other countries might exhibit different dynamics. Extending research to a broader set of countries would be necessary to test the universality of these results. Second, the thesis relies largely on survey-based longitudinal data (e.g., CFPS in China, HILDA in Australia, UK panel data). Such data are rich but come with inherent limitations like self-reported measures, possible measurement errors, and unobserved heterogeneity. Although each study took pains to apply techniques (such as fixed effects, propensity matching, selection models or instrument variable methods) to mitigate these concerns, there remains the possibility that some unmeasured factors (e.g. individual personality traits, precise employer policies, or local cultural variations) could influence the outcomes. In Chapter 3, for instance, proxies for bargaining power might not capture all facets of how couples actually negotiate decisions; unobserved elements like communication quality or extended family influence could also play a role. Third, the time frame of the data analyzed in these studies is a limitation. Most of the data precede the late 2010s (and certainly precede the COVID-19 pandemic of 2020), which means the conclusions were drawn without observing the dramatic shifts in work and family life that have occurred recently. The world has changed since the pandemic – from the normalization of remote work to heightened awareness of work-family balance – and these changes could affect the phenomena studied in this thesis. For example, the rise of remote work may alter how flexible working arrangements impact career outcomes, and pandemic-related disruptions might have shifted household bargaining dynamics in ways not captured by pre-2020 data.

Building on the foregoing insights and acknowledging their limitations, several promising avenues for future research emerge. First, expanding the cross-national scope through comparative analyses of countries with divergent cultural norms and policy frameworks, such as pro-natalist incentives in parts of Europe or evolving fertility regulations in other East Asian contexts, would clarify whether the patterns identified in Chapter 1 for China generalize or reveal notable exceptions. Second, at the firm level, future studies should examine the new forms of flexible-work arrangements that have arisen since the COVID-19 pandemic, particularly remote and hybrid models, to assess their net effects on gender gaps in both labor and family spheres. Finally, within the household domain, investigators should explore how shifting social norms and targeted policies (for example, expanded paternity leave or subsidized childcare) reshape intra-household bargaining power and, consequently, fertility decisions. In particular, further analysis of distinct respondent profiles, such as career-oriented versus norm-conforming women identified in Chapter 3 as deviating from traditional bargaining models, can yield a more nuanced understanding of the interactions among policy, preference, and demographic behavior.

Acknowledgements

The author would like to thank the Melbourne Institute of Applied Economic and Social Research for providing access to the HILDA data set. This paper utilizes unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, conducted by the Australian Government Department of Social Services (DSS). The findings and interpretations presented in this paper are those of the author and should not be attributed to the Australian Government, DSS, or any of DSS's contractors or partners. DOI: 10.26193/YP7MNU, ADA Dataverse, V3.

This research also makes use of data from Understanding Society: the UK Household Longitudinal Study (UKHLS), an initiative funded by the Economic and Social Research Council and various UK Government Departments. The study is led scientifically by the Institute for Social and Economic Research at the University of Essex, with survey delivery by the National Centre for Social Research (NatCen) and Verian (formerly Kantar Public), and data distributed by the UK Data Service. The COVID-19 Study (2020–2021) was funded by the Economic and Social Research Council and the Health Foundation, with serology testing supported by the COVID-19 Longitudinal Health and Wealth – National Core Study. Fieldwork for the web survey was conducted by Ipsos MORI, and for the telephone survey by Kantar.

In addition, this study draws on data from the China Family Panel Studies (CFPS), which is funded by Peking University and the National Natural Science Foundation of China. The CFPS is managed and maintained by the Institute of Social Science Survey at Peking University. The author gratefully acknowledges the institutions and individuals involved in collecting and providing access to these valuable data resources.