MASTERARBEIT | MASTER'S THESIS

Titel | Title

How do extended school days affect maternal labor supply?

verfasst von | submitted by

Paula Reihlen B.A.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 642
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Philosophy and Economics
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Christine Zulehner Privatdoz.

Abstract

Increased availability of public childcare is commonly expected to increase maternal labor supply. Yet, relatively few studies investigate the impact of more readily available public childcare on mothers with *school-aged* children. Therefore, this thesis exploits the quasi-experimental nature of an educational policy in Germany in the early 2000s that shortened the traditional 13-year education to 12 years. As the basic curriculum was left unchanged, this reform effectively led to longer school days for students and thus implicitly provided free childcare to mothers. Identification is based on an extended Difference-in-Differences approach with staggered adoption of treatment, as the reform was implemented over the course of several years in different federal states. Given that even after reunification, full-time employment rates for mothers in East Germany are significantly higher than those in West Germany, my analysis concentrates on West Germany. Using data from the German Socioeconomic Panel (SOEP) from 1995 to 2012, I find no effects on the extensive margins, meaning that West German mothers did not change their labor force status in response to the extended school days. However, on the intensive margins, I estimate small but statistically and economically significant increases in the average weekly working hours of affected mothers several years after the policy's implementation.

Kurzzusammenfassung

Es wird allgemein erwartet, dass ein erweitertes Angebot an öffentlicher Kinderbetreuung das Arbeitsangebot von Müttern erhöht. Dennoch gibt es vergleichsweise wenige Studien, die die Auswirkungen öffentlicher Kinderbetreuung auf Mütter von Kindern im *schulpflichtigen* Alter untersuchen. In dieser Masterarbeit nutze ich daher den quasi-experimentellen Charakter einer bildungspolitischen Reform in Deutschland in den frühen 2000er Jahren, die die traditionell 13-jährige Schulzeit auf 12 Jahre verkürzte. Da das Grundcurriculum unverändert blieb, führte die Reform effektiv zu längeren Schultagen für die Schüler:innen und stellte somit eine kostenlose Kinderbetreuung für die Mütter dar. Die Identifizierung basiert auf einem erweiterten Difference-in-Differences Ansatz mit gestaffelter Einführung des 'Treatments', da die Reform im Laufe mehrerer Jahre in verschiedenen Bundesländern eingeführt wurde. Da die Vollzeiterwerbsquoten von Müttern in Ostdeutschland auch nach der Wiedervereinigung deutlich über denen in Westdeutschland liegen, konzentriert sich meine Analyse auf Westdeutschland. Unter Verwendung von Daten des Sozioökonomischen Panels (SOEP) aus den Jahren 1995 bis 2012, finde ich keine Anhaltspunkte für eine Erhöhung der Erwerbsttätigkeit an den extensiven Rändern, d.h. westdeutsche Mütter haben ihren Erwerbsstatus im Rahmen der G8 Reform und den damit einhergehenden verlängerten Schultagen nicht geändert. Auf den intensiven Rändern finde ich jedoch einen kleinen, aber statistisch sowie ökonomisch signifikanten Anstieg der durchschnittlichen Wochenarbeitszeit von Müttern mehrere Jahre nach der Implementierung der Reform.

Contents

List of Tables

List of Figures

1. Introduction

Over the past decades, significant progress has been made toward achieving gender equality within many Western labor markets. Yet, despite increased labor participation, women still tend to work fewer hours than men. The EU, for instance, estimates that in 2021, 67.7% of women were employed, while the employment rate for men was 78.% European Commission (2022). Many empirical studies have been conducted to investigate the drivers of female labor force participation, but establishing causality is challenging. The decision to enter or exit the labor market is influenced by many different factors which make it difficult to isolate singular effects. For example, a correlation between affordable childcare and maternal labor supply does not necessarily imply that increased childcare availability *causes* higher employment. After all, it could be that in countries with more accessible or less expensive childcare, gender norms constraining female employment are less prevalent.

Consequently, an increasing number of studies rely on quasi-experimental variation induced by different policies to learn about the drivers of female labor participation. These situations are often referred to as *natural experiments* because 'nature' randomizes individuals into treatment and control groups. Similar to Randomized Controlled Trials (RCTs), these studies aim to identify causal effects rather than mere correlations through randomization. Randomization creates statistical independence between treatment and confounders, ensuring that any observed differences in outcomes can be attributed to the treatment itself rather than to the confounders Dunning (2012). In fact, such quasi-experimental approaches have become so popular that some economists argue they have sparked a "credibility revolution" in empirical economics Angrist and Pischke (2010).

In this thesis, I also attempt a quasi-experimental approach to learn about female employment in the context of childcare availability. Since, even after reunification, full-time employment rates for mothers in East Germany are significantly higher than those in West Germany, my analysis focuses on West Germany. More specifically, I exploit the so-called G8 reform implemented in Germany in the early 2000s that shortened the traditional 13-year education to 12 years, thereby reducing the total years students spent in high school from 9 to 8 years. Since the the basic curriculum was left unchanged, students had longer school days to compensate for the one-year loss in total education. Although the reform was primarily an educational policy, I argue that it implicitly increased the availability of free childcare for families through the expansion of the school day. Potentially, this could have affected maternal labor supply because longer school days for children imply more spare time for mothers which may have translated into higher labor participation. Therefore, this thesis investigates the impact of the reform not from the students' perspective but from one group not initially targeted and almost entirely overlooked by the public and academic debates: the student's mothers. The research question, thus, is whether extended school days for high school children have increased the labor supply of affected mothers in West Germany in the early 2000s.

My main empirical model exploits the variation across states and time in the implementation of the G8 reform to estimate labor supply among mothers in an extended Difference-in-Differences (DiD) framework. The core idea behind my approach is that mothers in federal states where the reform was not yet implemented or whose children were enrolled before the first implementation year serve as controls for mothers who were affected by the reform through their children. Thus, similar to an RCT, nature randomizes mothers – depending on their children's school enrollment year and their federal state of residency – intro treatment or control. The DiD framework is extended insofar – as opposed to the canonical setup with only two groups and two time periods – the G8 reform was implemented in 13 out of 16 federal states over the course of seven years, resulting in staggered adoption of treatment.[1] Consequently, the extension of school days affected mothers differently depending on their household's state of residency, the federal state-specific implementation timeline and the Gymnasium enrollment year of their youngest child.

To estimate treatment effects, I use data from the German Socioeconomic Panel (SOEP) from 1995 to 2012 allowing me to link information about individual schooling histories of children with labor market data from their mothers. Ultimately, I find no effects of the extended school days on the *extensive margins*, i.e. mothers did not change their labor force status from 'non-working' to 'working' in response to the policy. However, at the *intensive margins*, I find small positive and statistically significant increases in the average weekly working hours of mothers several years after the policy was implemented when estimating an event study regression model that allows me to capture dynamic treatment effects. Limitations of my research design, as well as the validity of the identification assumptions, are discussed extensively at the end of the thesis. Furthermore, I provide

[1] Exceptions were Saxony, Thuringia, and Rhineland-Palatinate. Whereas in Saxony and Thuringia, the overall instruction time in the academic schooling track was already 8 years since 1949, Rhineland-Palatinate introduced the reform in selected schools only.

robustness checks in the Appendix A.2 to further corroborate my results.

My interest in studying the effects of an increased availability of affordable childcare on female labor supply is driven by several factors. First, women's under-representation in the labor market can lead to increased reliance on others and thus greater vulnerability. Women are much more likely to suffer from old-age poverty due to unequal care and work distributions because "women often have shorter working lives than men and therefore lower incomes from contributory pension programmes" (see Roig et al. (2022)). But increasing female employment also matters for the economy. As many Western countries face labor shortages due to aging populations, women represent an underused resource in the labor market European Commission (2022). Furthermore, lower female labor participation leads to substantial GDP losses. The EU, for instance, estimates that the economic losses due to gender employment gaps amount to approximately €370 billion per year European Commission (2022). Therefore, in this thesis, I am interested in childcare availability as one potential channel through which female (maternal) labor supply might be increased.

The thesis is structured as follows. Section 2 motivates my empirical research from a philosophical point of view and gives an overview of the empirical literature on childcare offerings and maternal labor supply. In Section 3, I introduce the data set I use to estimate treatment effects, the German Socioeconomic Panel (SOEP), and show some general trends in female (maternal) labor supply for the years 2000–20, as well as the differences in employment and work hours by gender and parental status in Germany. In Section 4, I explain the policy background of my research design, the G8 reform. Section 5 provides the empirical analysis of my thesis, where I explain how I constructed my main sample, introduce the regression models, discuss the identification assumptions of my research design and report my findings. In Section 6, I discuss these findings within the broader context of the existing literature, point to some limitations of my research design, and explore some ideas for further research. Section 7 concludes.

2. Related Literature

This chapter reviews the relevant literature for my study, divided into two sections. In Section 2.1, I introduce some main philosophical debates concerning my study to motivate my research question. In Section 2.2, I provide an overview of some of the empirical studies about the effects of the availability of childcare on maternal labor market outcomes.

2.1. Philosophical debates

Although this thesis is mostly empirical, questions about female labor supply and public policy are also highly relevant from a philosophical point of view. Philosophers, especially feminist philosophers, have long been interested in normative questions about the distribution of care work within family structures, as well as factors perpetuating the gendered division of care responsibilities. Most importantly, philosophy has provided reasons to examine what might initially appear as personal or private matters, such as the distribution of care work within households. Accordingly, the distribution of care work is not merely a personal choice but is deeply embedded in societal norms and structures that perpetuate gender inequalities. Thus, in this chapter, I review some important voices in the debate that further motivate why – apart from the economic advantages of higher female employment – economics and public policy should investigate female labor supply.

Fraser (2016), for instance, studies extensively how activities of social reproduction, which she understands as "activities of provisioning, care-giving, and interaction that produce and maintain social bonds" (101) are related to wage labor. To her, reproductive labor constitutes a background condition for economic production because, without cooking, cleaning, or child-rearing, the workforce could not be reproduced. Despite their importance, she argues, reproductive activities are typically relegated to women within the confines of the private or domestic sphere where they become largely invisible (Fraser (2016), 102). Furthermore, since reproductive activities are mostly unpaid, they can create structural relations of inequality and economic dependency.

Philosophers have also questioned whether theories of justice or political action should be applied exclusively to the 'public' sphere (e.g., the state) or also to the 'private' sphere, including the family. To many feminist philosophers, the boundaries between the 'public' and the 'private' are not as clear, criticizing the private-public dichotomy and arguing that the private is much more political than commonly assumed (see, for instance, Satz (2017)). Thereafter, many institutions within the private domain are not 'natural' or 'given' but shaped by laws and societal norms Satz (2017).

Connectedly, philosophers have investigated conflicts of choice or liberty versus (in)equality within family structures. Whilst recognizing the importance of autonomous decision-making, Satz (2017) argues that whenever "the gendered division of labor in the family, even if freely chosen, operates in the context of a background system of injustice", it can give us reasons to reject such arrangements.

Likewise, Nussbaum (2001) has heavily criticized preference-based approaches to social choice. To her, preference-based approaches cannot account for what she calls 'adaptive preferences', i.e., preferences individuals have formed through life-long socialization into their socioeconomic and political circumstances. She cites the case of Vasanti, an Indian woman who 'accepts' the abuse of her husband not only because she is financially dependent on him but because she lacks an understanding of herself "as a bearer of rights and a citizen whose dignity and worth are equal to that of others" (Nussbaum (2001), 69). To her, this example shows that not all preferences should be on par for political purposes and certain preferences (e.g., sexist or racist preferences) should be excluded from the social choice function. Ultimately, she argues that some things should be granted to individuals irrespective of whether they want or 'prefer' them, including bodily integrity and other substantive human goods.

Last, another related debate in philosophy is on the evolution, nature, and functioning of social norms and how they impact choice behavior. Christina Bicchieri, for instance, at the intersection of philosophy and economics, argues that social norms should be understood as the "grammar of society" because they prescribe which behavior is (im)permissible in specific contexts Bicchieri et al. (2023). A similar understanding is shared by Jon Elster who argues that "[a] social norm is an injunction to act or to abstain from acting" (Elster (2015), 348), for instance, to cough in one's hands or greet appropriately. As opposed to legal norms, social norms are not sustained by the law but by emotions (e.g., feelings of guilt or shame in case of norm transgression) or by public disapproval.

One important subset of social norms is gender norms which are defined, according to a UN report, as "the informal rules and shared social expectations that distinguish expected behaviour on the basis of gender" (UN Women (2023), 17). Examples of gender norms are the male breadwinner norm prescribing that men should be the primary or sole earner of the household or the female caregiver norm saying that women ought to shoulder all or most of the household's care responsibilities. Therefore, the UN considers gender norms as "one of the significant factors shaping progress, stasis and backsliding on gender equality" (UN Women (2023), 4).

Empirically, many researchers have tried to elicit different social norms and quantify their impact on labor market outcomes. A comprehensive overview of this debate would go beyond the scope of this thesis. Yet, several interesting studies highlight the significant impact of social (gender) norms on individual behavior. Kleven (2022), for instance, studies how child penalties, i.e., the (negative) effects of parenthood on women, correlate with gender norms. Using data from the General Social Survey (GSS) which asks questions about gender roles in families he constructs a gender progressivity index. He finds that this gender progressivity index varies significantly across US states and significantly influences maternal employment: accordingly, "an increase in the gender progressivity index of one standard deviation reduces the child penalty in annual employment by 24pp and the child penalty in earnings by 36pp" (Kleven (2022), 3).

Relatedly, Cortés et al. (2022) investigate how beliefs about maternal workforce engagement are correlated with social norms. They are interested in the question of how social norms impact people's social appropriateness rankings of maternal work decisions by using hypothetical scenarios they present to survey participants. For instance, they elicit beliefs about whether young mothers should return to work when receiving a lucrative job offer or if wives (instead of their husbands) should stay home in case there is no (external) childcare option. Ultimately, they find that individuals' views on appropriate workforce engagement for women and mothers are heavily influenced by their perceptions of others' opinions.

Last, Bertrand et al. (2015) show that the distribution of the wife's share of the household income significantly decreases just beyond the point where her earnings surpass her husband's, i.e., a wife's share of the household income witnesses a sharp drop where she earns more than her husband. To them, this observation can be best explained by gender norms, which create a reluctance for scenarios in which the wife out-earns her husband. Interestingly, this is also related to different marriage outcomes: according to Bertrand et al. (2015), when a randomly chosen woman's income is expected to surpass that of a randomly chosen man, she is less likely to marry. Furthermore, divorce rates are higher for couples where the women earns more.

2.2. Empirical studies

Several empirical studies have investigated the effects of increased childcare availability on maternal labor supply. Beginning with children below compulsory schooling age in Germany, Bauernschuster and Schlotter (2015) analyze how the introduction of a legal claim to a place in kindergarten for children from the age of three in 1996 affected maternal employment. Using a Difference-in-Differences (DiD) and an Instrumental Variables (IV) approach to corroborate their results, their findings suggest a strong positive effect of the increase in eligibility for kindergarten places on maternal employment: for women with children aged three to four, the kindergarten reform led to an increase in maternal labor supply by about 6 percentage points.

Germany's introduction of the legal claim to a kindergarten place in the mid-90s was followed by two further childcare reforms in 2005 and 2008, the daycare expansion law ("Tagesbetreuungsausbaugesetz") and the law on support for children ("Kinderförderungsgesetz") whose effects Müller and Wrohlich (2020) estimate using data from the German Microcensus. Similar to Bauernschuster and Schlotter (2015), Müller and Wrohlich (2020) also find positive, yet smaller, effects. Accordingly, an increase in childcare places by 1 percentage point leads to a 0.2 percentage point increase in mothers' labor market participation rate. Furthermore, these effects are mainly driven by increases in part-time employment of mothers with medium-level qualifications.

Similar studies on the effects of increased childcare availability for children below primary school entry age were also conducted in many other countries. For Norway, Havnes and Mogstad (2011) analyze the effects of the so-called Kindergarten Act passed in 1975 which considerably expanded subsidized childcare. The authors found no effects of the Kindergarten Act on maternal labor supply, even though it significantly increased federal funding for kindergarten places (from \$35 million in 1975 to \$107 million in 1977). They argue that instead of increasing maternal employment, "universally accessible child care in Norway mostly crowded out informal care arrangements, implying a significant net cost of the child care subsidies" (Havnes and Mogstad (2011), 1464).

Using the National Longitudinal Survey of Children and Youth (NLSCY), Baker et al. (2008) also investigate the effects of increased availability and affordability of childcare by studying a family policy in Quebec, Canada, in the late 1990s. Even though the policy made kindergarten significantly cheaper (\$5.00 per day for 4-year-olds) and more available by offering more places, the authors found no significant increases in maternal employment.

Less evidence exists about the effects of more affordable (or more readily available) childcare on children *after* they have entered school. Using data from the Chilean Socioeconomic Household Survey and the Ministry of Education, Contreras and Sepúlveda (2017) investigate how an expansion of the Chilean school day until the afternoon for children between 8 and 13 years affected maternal labor participation rates. They find significant effects for single mothers but no effects for mothers in relationships (marital and non-marital).

The same reform was also studied by Berthelon et al. (2015), although using different data from Chile's Social Protection Survey (EPS). Contrary to Contreras and Sepúlveda (2017)), Berthelon et al. (2015) find that married mothers respond more strongly to the expansion of the school day. According to the authors, the reform could substantially affect maternal employment: if the policy fully covered all women in their sample, labor force participation would increase by 11.9 percentage points. Furthermore, they find that the reform not only increased employment but also had significant effects on the time women remained in employment.

In the US, Price and Wasserman (2023) analyze seasonal patterns in female employment for the years 1989–2019 and observe that in the summer months (May – July), female working hours dropped consistently by almost 10 % and thus twice as much as the working hours of men. They suggest school closures (and the unavailability of free childcare implicit in these closures) as a "unifying explanation for these [seasonal] patterns" (3) suggesting that increasing summer childcare could (at least partially) remedy the summer drops in female employment. For Germany, to my best knowledge, no study exists that investigates the effects of lengthening the school day on maternal labor supply.

Taken together, the empirical literature presents a mixed picture. Not only does the size of the estimated effects of extended childcare offerings on maternal labor supply vary between countries, but studies also identify differential effects based on which mothers are most affected (e.g., single mothers versus married mothers). Yet, no study has found *negative* effects of more readily available childcare offerings on maternal labor supply (although some studies find no or very small effects). Furthermore, the effects tend to be slightly stronger for younger children, likely because younger children require more care. Additionally, if public childcare crowds out existing private care arrangements (e.g., grandparents or neighbors), the effects on maternal employment may be rendered insignificant (Havnes and Mogstad (2011) see also Bauernschuster and Schlotter (2015)). Last, there are considerably fewer studies about the effects of public care offerings for school-aged children on maternal labor supply, specifically for children above the age of ten.

3. Descriptive Analysis

Before turning to the institutional background of my empirical analysis, this chapter provides a preliminary descriptive analysis of the development of female (maternal) labor supply over the past twenty years in Germany. In Section 3.1, I first describe the data set I use in this thesis. Next, in Section 3.2, I present a descriptive analysis of the development of maternal labor supply in Germany from 2000 to 2020.

3.1. Data

To estimate the relationship between extended school days and maternal labor supply within the context of the G8 reform, I use data from the German Socioeconomic Panel (SOEP) for the years 1995–2012.[1] The German Socioeconomic Panel is a representative study of private households in Germany started in 1984 and conducted annually by the German Institute for Economic Research Berlin (DIW Berlin). Apart from information about households, the SOEP also provides information on all individual members of each household. To ensure the sample represents the German population accurately, the DIW regularly draws refreshment samples and provides design and sampling weights Goebel et al. (2019).

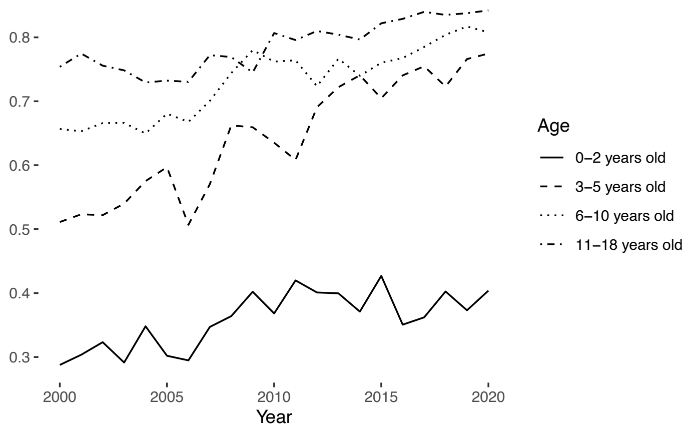
3.2. Development of maternal labor supply in Germany

Figure 1 illustrates maternal labor participation rates over time by the age of a mother's youngest child. As depicted, female labor force participation is lowest when the youngest child is between 0 and 24 months old. As children grow older, particularly in the 3-5 and 6-10 age brackets, there are noticeable increases in maternal labor force participation, potentially indicating a return to work as children enter kindergarten and later elementary school. Furthermore, children in Germany have a legal claim to a place in kindergarten, the so-called 'Kindergartenanspruch', from the age of three until they enter elementary

[1] Since the G8 reform was reversed in many federal states from 2013 onward, I use 2012 as the last survey year (see chapter 4).

school which could explain the large increases in female labor participation when children turn 3.[2] Female employment is highest when children are in high school and thus are between 10 and 18 years old.

Figure 1.: Maternal Labor Force Participation by Age of Youngest Child



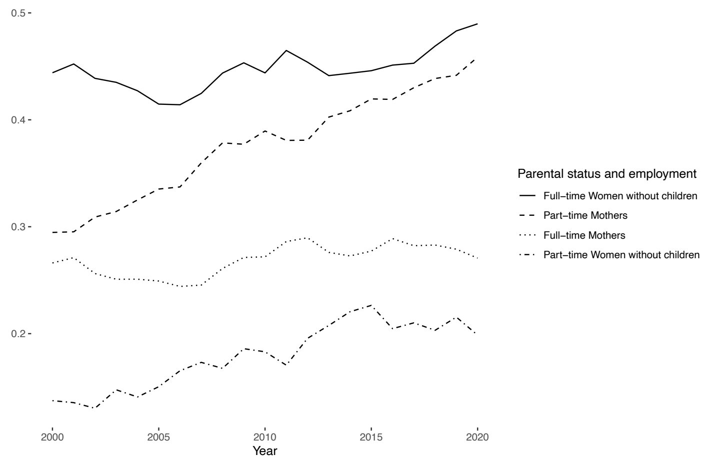
Notes: Maternal labor force participation rates by the age of their youngest child, Germany, based on SOEP data from 2000–2020. The variable labor force participation is based on the annual question on current employment status. To correct for non-random probability sampling designs and attrition, means in labor force participation per group and year are calculated using cross-sectional weights that restore marginal distributions in the population.
Source: SOEP.

Over time, all groups increased labor force participation, with the strongest increase for women with children aged three to five. Altogether, female labor participation rates are relatively high (between 30% and 80%) because the graph depicts all women engaged in paid labor, disregarding how many hours women work and thus the kind of employment they are engaged in (e.g., marginal, part-time, or full-time employment).

Therefore, Figure 2 shows the differences in part-time and full-time employment between women with and without children. As depicted, women without children are almost twice as likely to be engaged in full-time employment compared to mothers. Unsurprisingly, the opposite holds for part-time employment. Mothers are much more likely to work part-time compared to women without children, likely reflecting mothers' additional time constraints due to caregiving responsibilities. Conversely, women without children may be more inclined toward full-time employment because of fewer care obligations. Interestingly,

[2] Since the 1st of August 2013, alongside the legal claim to a place in kindergarten, children also have a legal claim to the so-called early childhood support from the age of one until the age of three (Achtes Sozialgesetzbuch, Kinder- und Jugendhilfegesetz, §24(2)).

Figure 2.: Differences in Part and Full-Time Employment of Women by Parental Status



Notes: Percentage of part-time and full-time working women, in Germany, based on SOEP data from 2000–2020. Women working less than 30 hours per week are considered working part-time. Means are calculated using cross-sectional weights that restore marginal distributions in the population.
Source: SOEP.

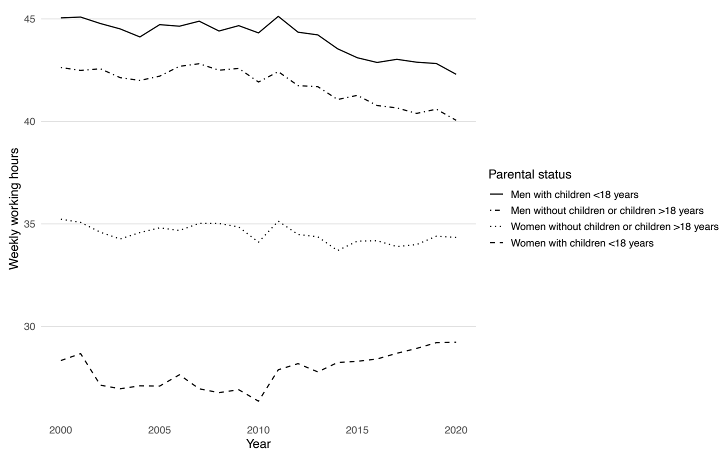
whereas part-time employment rose significantly by about 10 percentage points for both women with and without children, the percentage of women in full-time employment remained relatively constant at 45% (without children) and 25% (with children) between 2000 and 2020.

Transitioning from the discussion of the differences in part-time and full-time employment amongst females, Figure 3 examines the disparity in weekly working hours not only by parental status but also by gender. Notably, regardless of whether men have children or not, they work significantly more hours per week than women. As depicted in Figure 3, men with children dedicate slightly less than 45 hours on average per week to work, while those without children or adult children work around 43 hours. In the literature, this is sometimes called the *fatherhood premium* (as opposed to the *motherhood penalty*), where men work more hours (or earn more) after becoming fathers.[3] Women without children, on the other hand, work around 35 per week whereas women with children work between 20 and 30 hours, depending on the year. Interestingly, the average working hours of men

[3]The phenomenon of men experiencing a positive impact of fatherhood on their career paths and earnings has also been observed in many other countries (see, for instance, Hodges and Budig (2010) for a study conducted in the US). As opposed to the *fatherhood premium*, females often experience setbacks after becoming mothers, for instance, reduced wages or a slower career progression. This *motherhood penalty* can continue to persist even long after children have grown up (see Goldin et al. (2022)).

show a slight decrease of about 2.5 hours per week in the study period, whereas the weekly working hours of women without children remain relatively constant and those with children even show a slight increase.

Figure 3.: Differences in Weekly Working Hours by Gender and Parental Status



Weekly working hours by gender and Parental status, Germany, based on SOEP data from 2000–2020. *Note*: The variable actual weekly working hours is obtained by asking survey participants how many hours they work on average per week (incl. overtime). Means are calculated using cross-sectional weights that restore marginal distributions in the population.
Source: SOEP.

When examining the distributions of weekly working hours for men and women regardless of their parental status, across two exemplary years, 2010 and 2020, the disparities become evident once again. The histograms 4 and 5 show the frequency distribution of weekly working hours, with the x-axis representing the whole range of possible weekly working hours (i.e., 0h per week, 1h per week, ..., 80h per week) and the y-axis indicating the number of individuals falling within each hour range.

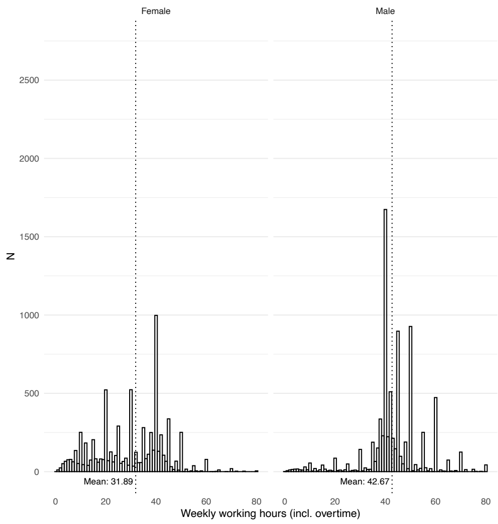
For men, the distributions for 2010 and 2020 peak around 43 and 41 hours, respectively. Furthermore, in both years, the distributions are concentrated around these averages, indicating that the majority of men work hours close to 43 (41) hours with relatively few men working significantly less. The histogram also shows that if men deviate from the average, they are more likely to work overtime instead of part-time.

Interestingly, the frequency distributions of weekly working hours of women also peak at approximately 40 hours in both years. Yet, despite this peak, the average weekly hours for women is at around 30 hours per week and thus significantly lower than those of men.

This difference is attributable to a much larger proportion of women working fewer than 40 hours per week, coupled with a much smaller proportion working overtime. Notably, whereas the average weekly working hours for men slightly decreased between 2010 and 2020, women's average weekly working hours increased by approximately 2.5 hours during these years.

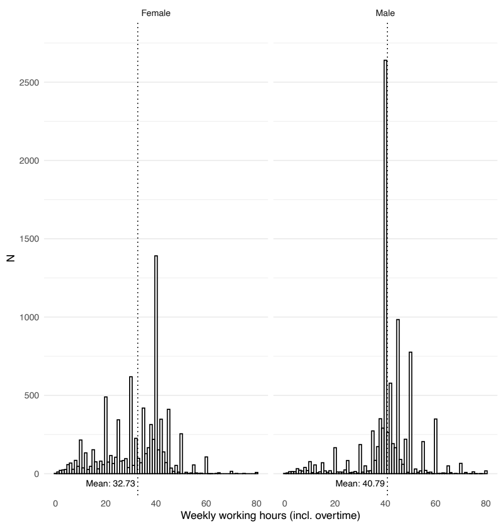
Much more could be analyzed concerning the differences in various labor market variables by gender or parental status. For instance, one could investigate longer time frames, look at country differences, or explore gender/ parent disparities across industries and occupational roles. Yet, these graphs already provide some insights into certain aspects of the German labor market. First, female labor participation rates show large differences depending on the age of the youngest child, with notable increases as children grow older, indicating a potential relation between childcare responsibilities and workforce engagement. Second, the differences in part-time and full-time employment of women with and without children also indicate that mothers exit the labor force or reduce hours in response to the additional care responsibilities. Last, whereas mothers work significantly less than women without children or children above the age of 18, fathers with underage children work even more than men without children or children older than 18 years. These phenomena are known as the *motherhood penalty* and the *fatherhood premium*.

Figure 4.: Distribution of Weekly Working Hours by Gender, 2010



Notes: Weekly working hours by gender (irrespective of parental status), Germany, based on SOEP data from 2010. Means are calculated using cross-sectional weights that restore marginal distributions in the population.

Figure 5.: Distribution of Weekly Working Hours by Gender, 2020



Notes: Weekly working hours by gender (irrespective of parental status), Germany, based on SOEP data from 2020. Means are calculated using cross-sectional weights that restore marginal distributions in the population.

4. Institutional Background: G8 reform

The German educational system is divided into different stages (for a comprehensive overview, see Bundeszentrale für politische Bildung (2013)). Compulsory schooling begins for all children at the age of six when they jointly enter elementary school (*Grundschule*). Elementary school typically comprises grades 1 to 4, except for Brandenburg and Berlin, where elementary school lasts until grade 6. After completion of elementary school, students are separated into different schooling tracks according to a non-binding recommendation primarily based on their academic performance. The three main schooling tracks are *Hauptschule*, *Realschule*, and *Gymnasium*. Some federal states also have comprehensive schools, the so-called *integrierte Gesamtschulen*, which combine different schooling tracks and allow for a later sorting of students into one of the respective tracks.

Whereas in *Hauptschule* and *Realschule*, students receive a certificate allowing them to start vocational training, *Gymnasium* is the only track that prepares students for higher education. Therefore, the *Gymnasium* is sometimes called the academic schooling track because, after successful completion, students earn the *Abitur*, the German university entrance qualification.

In 2021, around 37% of students were in *Gymnasium*, 20% in comprehensive schools, 17% in *Realschule* and 8% in *Hauptschule* KMK (2023). Another noteworthy characteristic of the German school system is that while the Basic Law governs basic educational principles, educational policy is the responsibility of the sixteen federal states. To coordinate joint interests, the educational ministers of the federal states meet regularly in the so-called *Kultusministerkonferenz*, the Standing Conference of the Ministers of Education and Cultural Affairs.

Until 2001, the academic schooling track in all federal states except Thuringia and Saxony lasted until grade 13. However, between 2001 and 2007, 13 out of 16 German states implemented the so-called 'G8 reform', shortening the duration of the academic schooling track from nine to eight years. As a result of this reform, students previously graduating after grade 13 would now earn the Abitur after grade 12. Exceptions were Saxony, Thuringia, and Rhineland-Palatinate. Whereas in Saxony and Thuringia, the overall instruction time in the academic track was already eight years since 1949, Rhineland-Palatinate only implemented a pilot scheme in 19 schools, keeping the previous 9-year system for all other schools. The implementation timeline of each federal state is summarized in Figure 6.

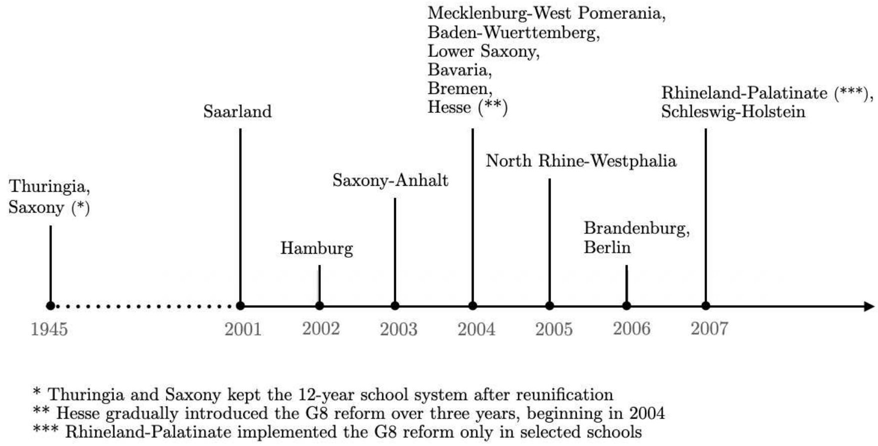


Figure 6.: Implementation Timeline G8 reform

The rationales behind the G8 reform were manyfold (see (Huebener and Marcus (2015), 4-5) and (Dahmann and Anger (2014), 8)). First, Germany sought to align the school leaving age with international standards as German students were comparatively old at the time of high school graduation. Thereby, the reform aimed to increase the competitiveness of German students in international labor markets. Second, by shortening the overall instruction time, German students were meant to transition faster into higher education (and thus the labor market) to relieve the German pension system. Last, the reform was meant to lower overall costs for schools by reducing the duration of schooling by one year.

For students, the G8 reform effectively led to longer school days. Whilst reducing the overall time students spent in school by one year, the instruction time required to earn the higher education entrance qualification was left unchanged. In other words, the same curriculum was taught in one year less such that students had longer school days to compensate for the one-year loss in total education.

The G8 reform was met with criticism from students, parents, and teachers alike (see Bundeszentrale für politische Bildung (2015)). Some raised concerns regarding increased stress and academic pressure on students and feared a negative impact on their leisure time and extracurricular activities. Others argued that the new schooling system would not leave enough time for the development of interdisciplinary skills (*überfachliche Kompetenzen*), such as creativity or tolerance Bundeszentrale für politische Bildung (2015). Parents were especially concerned about the quality of teaching, fearing that the decrease in overall instruction time negatively affected students' learning success. Others questioned whether the reform truly led to earlier transitions into higher education or the labor market, arguing that instead of going to university or starting to work, students would use the additional year for leisure instead. Interestingly, as the Bundeszentrale für politische Bildung (2015) notes, the criticism mainly comes from West Germany, whereas in East Germany, the eight-year Gymnasium is generally accepted.

Until today, the reform remains a topic of heated public discussion, having led some federal states to return to G9 already Bundeszentrale für politische Bildung (2015). Following persistent criticism, in 2014, Lower Saxony was the first state to decide to return to an academic track length of nine years. Bavaria, Lower Saxony, North Rhine-Westphalia, Saarland, and Schleswig-Holstein followed in the years to come. In Hesse, from the school year 2013/14 onwards, and thus only ten years after the reform was implemented, schools could individually decide whether they would keep the 8-year scheme or return to 9 years of schooling Hessisches Kultusministerium (nd).

While debates surrounding the reversal of the reform continue in many federal states, one group, not initially targeted but potentially affected by the reform (and thus its reversal) has received little attention both in public and academic debates: the students' mothers. The channel through which mothers could be affected by the extension of the school day is that extended school days may provide mothers with greater flexibility in balancing work and childcare responsibilities. With children spending more time in school, mothers may have increased capacities to participate in the labor market and work longer hours. Additionally, the provision of implicit childcare through extended school days may reduce the need for costly childcare arrangements, thereby facilitating maternal workforce participation. Therefore, in this thesis, I am interested in whether the G8 reform positively impacted maternal labor supply by implicitly providing free childcare.

To investigate this research question, I exploit the federalism in the G8 implementation, which – as I'll argue – provides a natural experiment to study the effects of increased childcare availability on maternal labor supply. Natural experiments differ from Randomized Controlled Trials (RCTs) in that variations in real-world conditions – such as policy changes implemented at different times or in different locations – create *quasi-random assignment* of individuals into treatment and control groups (see Dunning (2012) and Angrist and Pischke (2010)). The main idea underlying both RCTs and natural experiments is that randomization creates statistical independence between potential confounders and treatment assignment so that any post-treatment differences can be attributed to the treatment, rather than pre-existing group differences Dunning (2012). For instance, in the context of a clinical trial where the effectiveness of a drug is tested and researchers find positive effects on the treated, they can attribute these effects to the drug rather than pre-existing differences in the health or lifestyles of study participants. However, in the social sciences, RCTs often cannot be conducted for ethical, financial, or practical reasons Dunning (2012). Regarding childcare, for instance, it may be unethical to randomly offer some mothers better (or cheaper) childcare, while denying others the same opportunity because it can have profound impacts on later life outcomes, both for the mothers and their children. Therefore, increasingly, researchers rely on natural experiments to learn about causal effects.

In my research context, the idea is that the state-specific implementation timelines randomly assigned mothers into treatment and control groups based on their federal state of residence and the timing of the reform. Intuitively, the idea is that depending on the year a federal state adopted the G8 reform, some children (and their mothers) were affected earlier while others were affected later (or never) by the introduction of the reform and thus the extended school days implicit in them. In other words, the staggered implementation effectively led mothers to be as-if randomly exposed to varying levels of childcare availability: the federal state-specific implementation timelines created exogenous variation in the availability of childcare offerings. Thus, as I'll argue, the G8 reform created a *natural experiment* that I aim to exploit to learn about the effects of extended school days on maternal labor supply.

5. Empirical Analysis

This chapter presents the main empirical analysis of my thesis. In Section 5.1, I begin by explaining how I constructed the main sample. Section 5.2 outlines the identification strategy employed to investigate treatment effects. Finally, in Section 5.4, I discuss my findings, with subsections dedicated to the outcome variables *labor force participation* and *weekly working hours*.

5.1. Main sample construction

To identify the effects of the G8 reform on the labor supply of West German mothers, I use information about the implementation timeline of each federal state combined with data on the household's state of residency and the year a mother's youngest child was enrolled in the academic schooling track. I take the youngest child to be decisive for maternal work decisions as younger children typically require more care. Since neither the year of enrollment into Gymnasium nor primary school is provided by the data set, I identify the year of school entry by considering individual schooling histories and – in case of missings – information about a child's year (and month) of birth.

More specifically, in the household survey, parents are asked what kind of school their child *currently* attends. Since interviews are conducted in the first quarter of the year and high school begins in late summer (typically August or September) (see Goebel et al. (2019) for more detailed information about the SOEP), for all children transitioning from elementary school to Gymnasium in a given year, households report their children to be in elementary school at the time of the interview even though their child enters Gymnasium later that year. For instance, if a household is interviewed in April 2000 and answers that their child currently attends elementary school but transitions to high school later that year, the change in school type is reflected only in the 2001 survey instrument. Therefore, I identify transitions to high school from leads instead of lags.

In the case of incomplete information about a child's transition from elementary school to Gymnasium but complete information about a child's transition from Kindergarten to elementary school, I impute the year of Gymnasium entry based on the year that child enrolls in primary school, taking into account that primary school lasts six not four years in Brandenburg and Berlin. The main advantage of identifying Gymnasium entry years from individual schooling histories is that it can account for deviations from typical schooling paths, such as repeating a grade or early/late enrollment in elementary school.

However, this method has the drawback of losing some mother-child observations when data on individual schooling histories is missing. This can happen when households – for various reasons – do not answer questions about the kind of school their child attends (item non-response). Therefore, to avoid losing mothers for whose children these schooling histories are not given, I impute the year of enrollment in the academic track using the child's birth year and month, along with state-specific school enrollment cut-off dates.[1]

More specifically, in Germany, each federal state determines its own cut-off date by which a child who has reached the age of six must be enrolled in school. Depending on the federal state, this cut-off date is between the 30th of June and the 30th of September.[2] For instance, in Bremen, schooling is compulsory for children the year they turn six if they are born before June 30, whereas the cut-off date in Bavaria is September 30. For children born after the state-specific cut-off date, parents, in consultation with the responsible teachers, can decide whether their child should start school immediately or wait until the following year. These children are called 'Kann-Kinder' because they are not required to start elementary school that year but have the option to do so.

Since there is no official data on how many 'Kann-Kinder' stay in kindergarten for another year or enter elementary school immediately, I assume that children born after the state-specific cut-off date enter school the following year. For instance, for a child born in Bremen after June 30 in the year 2000, the imputed Gymnasium entry year would be 2011, not 2010. Of course, this imputation method is less accurate than using individual schooling histories and can misidentify the exact year of school enrollment, for instance, when parents decide to send their 'Kann-Kind' to school immediately or when students deviate from typical schooling paths.

Therefore, to assess the robustness of my imputations, I also calculate the year of Gymnasium enrollment based on the year and month of birth for the children *with* available individual schooling histories. This enables me to get a rough indication for how many children the imputed and actual school enrollment years coincide. The results are reported in Table 1. Accordingly, for the majority of children (65%), the imputed and imputed values match perfectly. For approximately 31.5%, the imputed values for

[1] Unlike other authors, I do not identify Gymnasium entry years from the year of high school graduation, as this method cannot account for the so-called 'double cohorts' — the simultaneous graduation of the first G8 and the last G9 cohorts.

[2] For the exact cut-off rules of each federal state, see Schraml (2018).

Gymnasium entry years deviate from the actual entry years by only one year. Thus, only for less than 5% of children, the imputed values are off by more than one year. However, to further corroborate my results, I also show that my estimates are robust when only using mother-child observations for which individual schooling histories are given (see Appendix A.2).[3]

Table 1.: Imputed vs. Actual Gymnasium Enrollment Years

Deviation (years)	-6	-5	-4	-3	-2	-1	0	1	2	3	4
# of children	1	1	4	14	67	810	2997	648	70	11	1
% of children	0.02	0.02	0.09	0.3	1.45	17.5	64.81	14.01	1.51	0.24	0.02

Notes: This table displays the number and percentage of children for each deviation in years between the imputed and actual enrollment years (identified through individual schooling histories). Negative values indicate that the imputed Gymnasium entry year is *earlier* than the actual Gymnasium entry year. For example, if the imputed enrollment year is 2000 and the actual year is 2001, the deviation is -1.

Having identified the year of Gymnasium enrollment, I can now assign mothers to treatment and control groups. Treatment assignment of mothers, i.e. whether a mother is affected by the reform through her youngest child, is determined by i) the household's federal state of residency at the time of the survey, ii) the federal state-specific G8 implementation year, and iii) the Gymnasium entry year of the mother's youngest child. In other words, treatment depends on whether a mother's youngest child is young enough to be affected by the new schooling reform, based on the household's state of residency and the year that state implemented the G8 reform.[4]

Since the analysis focuses on West Germany, all observations from East German federal states are dropped from the sample. Furthermore, Rhineland-Palatinate is excluded from the analysis because it implemented the G8 reform during the study period in selected schools only. Last, I drop all observations from Hesse because Hesse implemented the G8 reform over the course of three years to ensure enough university and apprenticeship places were available for students of the double cohorts (Hessisches Kultusministerium

[3] As shown in Table 6 which reports the regression results for the outcome variable *labor force participation*, across most variables, the estimated coefficients are almost identical in size when only using the unimputed values. Furthermore, signs and significance levels are – except for the variable *regional type* which becomes insignificant – identical. Table 7 reports the regression results for the outcome variable *weekly working hours*. Here, the effect sizes vary a bit more than for the outcome variable *labor force participation*; yet, signs and significance levels are the same when using the 'restricted' sample. Thus, my estimates are robust when only using mother-child observations for which individual schooling histories are given.

[4] An exemplary snippet of code for treatment assignment for the year 2004 is given in the Appendix A.1.

(2010), 6). Since I cannot identify which school/ region adopted the reform in a specific year, I cannot reliably determine the treatment status of mothers from Hesse. Thus, the final sample consists of 1872 mothers from Baden-Wuerttemberg, Bavaria, Berlin, Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, Schleswig-Holstein, and Saarland, of whom 1376 are eventually affected by the reform through their children. The descriptive statistics of all mothers and their partners in the sample are presented in Table 2.

Table 2.: Descriptive Statistics of full Sample

	Mean	Minimum	Maximum	Median
Age	42.43	17.00	68.00	43.00
Education (in years)	13.23	7.00	18.00	12.00
Experience (full-time)	8.14	0.00	41.25	7.00
Experience (part-time)	5.62	0.00	45.83	3.92
Net labor income (in €)	1136.07	0.00	80000.00	879.00
Migration background	1.21	1.00	3.00	1.00
Marital status	0.82	0.00	1.00	1.00
Number of children	2.23	1.00	6.00	2.00
Regional type	0.79	0.00	1.00	1.00
Partner's education	13.77	7.00	18.00	13.00
Partner migration background	1.20	1.00	3.00	1.00
Partner's labor force status	0.94	0.00	1.00	1.00
Partner's net labor income (in €)	2902.25	0.00	30000.00	2500.00

Note. This table displays the summary statistics of all mothers in the sample over the entire study period (1995-2012).
Source: SOEP.

As shown, mothers are on average about 42 years old, have slightly more than two children, and earn a net income of about 1136.07 €. Furthermore, they have about 8 years of experience in full-time work and about 5.5 years of experience in part-time work. Last, despite having about the same years of education, their partners earn more than double their income.

Table 3 presents the descriptive statistics for mothers in the sample according to their treatment status. It compares mothers who were never treated with those who eventually received treatment in 2000, i.e. the year before the first federal state, Saarland, implemented the G8 reform. As shown, the two groups exhibit no statistically significant differences in terms of education, full-time work experience, migration background, marital status, number of children in the household, household's regional type, as well as partner's migration background, net labor income, and labor force status. The variables where I find significant differences below the 5%-level are age and part-time work experience. Yet, these differences are relatively unsurprising, as by design, control mothers are mothers whose youngest child was enrolled in Gymnasium before the G8 reform was implemented in their federal state, whereas treatment mothers are those whose children were born and consequently enrolled in the academic track later. Thus, treatment mothers tend to be older and therefore have more part-time work experience.[5] Similar to Bauernschuster and Schlotter (2015) to avoid maturity effects driving my results, I control for age variables. Differences in partner's education are unexpected; however, they are only significant at the 10%-level and economically, very small (partners of treated mothers have around 22 more days of education). All variables reported in Table 2 are control variables in my estimations.

Table 3.: Descriptive Statistics of final Sample

Variable	Control	Treatment	T statistic	P-value
Age	40.64	34.52	15.10	$4.65e^{-44}$
Education	13.10	13.04	-0.44	0.66
Experience (full-time)	7.57	7.18	1.61	0.11
Experience (part-time)	4.62	1.86	8.75	2.72^{-17}
Net labor income (in €)	932.12	850.83	1.44	0.15
Migration background	1.23	1.23	-0.51	0.61
Marital status	0.87	0.85	0.34	0.74
Number of children	2.21	2.17	-0.06	0.96
Regional type	0.84	0.78	1.55	0.12
Partner's education	13.45	13.51	-1.72	0.09
Partner's migration background	1.20	1.21	-0.03	0.98
Partner's labor force status	0.94	0.93	-0.04	0.97
Partner's net labor income (in €)	2463.11	2321.99	1.55	0.12

Note. This table displays the summary statistics of treatment and control mothers in the year 2000, the last year before the first federal state implemented the G8 reform. The means of all variables are calculated using cross-sectional weights that restore marginal distributions in the population.
Source: SOEP.

5.2. Empirical strategy

To examine the effects of extended school days on maternal labor supply, I estimate several models for two outcome variables, *labor force participation* and *weekly working hours*. *Labor force participation* is a binary variable that indicates whether a woman is working or not. *Weekly working hours* is a continuous variable reporting the hours

[5] Bauernschuster and Schlotter (2015), who study the effects of the introduction of a legal claim to a place in kindergarten in Germany, find similar significant differences in their age variables. They also argue that these differences are "by mere construction", i.e. due to the research design.

(incl. overtime) a given woman works on average per week. The first model I estimate by OLS is a simple regression model without any state or time fixed effects to provide a baseline understanding of the relationship between extended school hours and the outcome variables.[6] This model completely ignores the panel structure of the data by 'pooling' data across multiple years, essentially treating the data set as a single cross-sectional dataset:

$$y_{ist} = \alpha + 1(t_{is} \geq t_s^*)\delta + \mathbf{X}_{ist}'\beta + u_{ist} \tag{5.1}$$

where i indexes mother i, t a specific survey year, and s the federal state of residency of mother i at time t. y_{ist} represents the outcome variables *labor force participation* and *weekly working hours*. The main independent variable $1(t_{is} \geq t_s^*)$ is an indicator term that equals 1 if the enrollment year into Gymnasium t_{is} of mother $i's$ youngest child in state s is larger than or equal to the year the G8 reform was implemented in s, t_s^*. $\mathbf{X}_{ist}$ is a vector of control variables for maternal, partner, and household characteristics. More specifically, in all equations, I control for mother $i's$ age, part and full-time work experience, years of education, migration background, marital status, as well as her partner's education, labor force status, and migration background. On the household level, I consider the regional type the household is located at (urban vs. rural) and the number of children in the household.

While pooled OLS offers a straightforward estimation method, it does not account for unobserved heterogeneity between federal states or survey years, potentially leading to biased and inconsistent estimates. The main problem is that for OLS to be unbiased, the correlation between the error term and the regressors must be zero:

$$\text{cov}(\mathbf{X}_{ist}, u_{ist}) = 0$$

However, this condition often fails to hold in panel data settings, where observations may be correlated over time or across groups. To address these issues, I next estimate a one-way fixed effects model that controls for unobserved heterogeneity across the different federal states.[7] By including state fixed effects, this model captures the time-invariant characteristics specific to each federal state that may affect the outcome variables and

[6] My analysis begins with the Pooled OLS, followed by two fixed effects models, and finally progresses to the dynamic TWFE model. This sequence is deliberately chosen to start with the simplest model, despite being more susceptible to bias, in order to gradually introduce the dynamic DiD model, which is the most refined model of this thesis.

[7] Additionally, I conducted a Hausman test to determine the appropriate model specification. The null hypothesis was rejected at the 1%-level, suggesting that fixed effects models are more suitable than random effects models.

are omitted in equation (5.1):

$$y_{ist} = \alpha + \phi_s + 1(t_{is} \geq t_s^*)\delta + \mathbf{X}'_{ist}\beta + u_{ist} \tag{5.2}$$

In this model, again, y_{ist} represents the labor market outcomes for mother i in year t living in federal state s, $1(t_{is} \geq t_s^*)$ is the treatment indicator term and $\mathbf{X}'_{ist}$ is the same vector of controls as before. Note that now, the variable ϕ_s is added to account for time-invariant state-specific effects not included in equation (5.1). Examples of such effects could be state-level labor market or educational policies that potentially affect both the availability of childcare options and mothers' decisions to work; economic and geographic conditions such as industry composition, infrastructure, or urbanization; as well as cultural norms that may be state-specific but constant over time. For instance, the so-called male breadwinner norm dictating that men should be the sole or primary earner in the household may be more prevalent in more conservative federal states such as Bavaria and is likely to be constant across years.

To further account for time fixed effects, I next estimate a so-called two-way fixed effects (TWFE) model that controls for both state and time fixed effects. Time-fixed effects control for potential time-varying confounders that are assumed to be constant across individuals or states. In this setup, time fixed effects could be any macroeconomic conditions that affect all German federal states during the survey period, such as changes in inflation rates, federal labor policies, or general economic fluctuations. The two-way fixed effects model is given by the following equation:

$$y_{ist} = \alpha + \phi_s + \lambda_t + 1(t_{is} \geq t_s^*)\delta + \mathbf{X}'_{ist}\beta + u_{ist} \tag{5.3}$$

which is the same as equation (5.2) except it includes λ_t, the time-fixed effects. In all models (5.1), (5.2), and (5.3), standard errors are clustered at the federal-state level and robust to heteroskedasticity.

The very last model presents the most refined model of this thesis, where I try to understand the dynamics of outcomes while controlling for fixed differences across states and national trends over time. Using multiple survey waves before and after the policy was implemented, I aim to study how potential policy effects develop over time, for instance, whether the effect is immediate or occurs only a few years after G8 was implemented. Furthermore, studying dynamics also allows me to test for pre-trends, i.e. any pre-existing differences in trends that would threaten the parallel trends assumption. The model I estimate is given by the following equation:

$$\begin{aligned} y_{ist} &= \phi_s + \lambda_t + \underbrace{\delta_1 d_{i,1} + \ldots + \delta_{t_0-2} d_{i,t_0-2}}_{\text{Pre-treatment effects}} + \underbrace{\delta_{t_0} d_{i,t_0} + \ldots + \delta_T d_{i,T}}_{\text{Post-treatment effects}} + u_{ist} \\ &= \phi_s + \lambda_t + \sum_{\tau=1}^{t_0-2} \delta_\tau G8_{ist} + \sum_{\tau=t_0}^{T=8} \delta_\tau G8_{ist} + u_{ist} \end{aligned} \tag{5.4}$$

where y_{ist} represent the weekly working hours or the labor force participation of mother i at time t in federal state s, ϕ_s are state-fixed effects and λ_t are time-fixed effects. $G8$ is a dummy variable that equals 1 if mother i has a child affected by the G8 reform in survey year t given their federal state of residency s, and 0 otherwise. In this model, the coefficients δ measure the treatment effects *relative* to period $t_0 - 1$, whereas for:

- $\tau < t_0$, i.e. years *before* implementation of the G8-reform, these are pre-differences
- $\tau > t_0$, i.e. years *after* implementation of the G8-reform, these are post-differences

In other words, each estimate of δ measures the change in outcomes for states where the G8 reform was implemented relative to states where the G8 reform was not implemented in year t, relative to the year before implementation. The endpoint for my estimated event study coefficients is $T = 8$, meaning I analyze 8 years after the implementation of the G8 reform to capture all potential effects from the beginning until the end of high school. If the weekly working hours of women evolved in parallel in treated and non-treated states, I expect the pre-treatment estimates of coefficients δ to be zero.

5.3. Identification assumptions

Before reporting the estimated effects, I now turn to the underlying assumptions of my identification strategy. Identification assumptions are crucial to any analysis because they ensure the validity of the estimated effects. Thus, if met, they provide credibility that the estimated effects are truly attributable to the treatment rather than confounding variables. Determining whether identification assumptions hold is challenging because these assumptions are fundamentally untestable. However, theoretical considerations and empirical evidence can provide support for their validity. In the context of Difference-in-Differences (DiD) analyses, there are two main identification assumptions, the *parallel trends* assumption and the *no anticipation* assumption.

Parallel trends. The primary identification assumption in Difference-in-Differences is the *parallel trends assumption*, according to which, in the absence of treatment, the average outcomes for the treated and control groups would have followed the same trends.

In my setup, this implies that maternal labor supply in both the control and treatment groups would have evolved in parallel if the treatment had never occurred. In other words, for the *parallel trends assumption* to hold, without the G8 reform, maternal labor supply of treatment and control mothers needs to follow the same trend.

Since the SOEP is a panel data set with multiple years before treatment, I can use data from pre-treatment years to provide evidence for the *parallel trends assumption* by testing whether pre-implementation event study coefficients are statistically insignificant. If these coefficients were statistically significant, it would indicate potential pre-existing trends that could bias the estimated treatment effects. However, as shown in the event study plots 7 (*labor force participation*) and 8 (*weekly working hours*), the pre-implementation coefficients are not statistically significant, indicating no differential pre-trends. Furthermore, a potential violation of the *parallel trends assumption* could arise if another policy affected the labor force participation or the weekly working hours of treated and untreated mothers differently. Yet, to my knowledge, no such policy was implemented within the given time frame.

No anticipation. The second identification assumption in Difference-in-Differences designs is the *no anticipation assumption*. This assumption requires that, before the treatment, the treated group is not influenced by the upcoming treatment. In the context of this research, the concern is whether women, in anticipation of the G8 reform and thus the extended school days, entered the labor force or increased their working hours already before the reform was implemented. However, I find it highly unlikely that women systematically changed their employment status or increased their working hours before the reform. This is because, as argued before, the channel through which the reform would impact maternal labor supply depends on the *actual* implementation of extended school days, which reduces the childcare burden only *after* the reform is implemented. Therefore, I find any significant behavioral changes in employment status or working hours before the implementation very improbable, as school days were extended only after the G8 reform took effect.

Randomization. As I explained in Section 4, my research design exploits the differential timing between federal states of the G8 reform as a *natural experiment*, i.e. it treats the reform as a situation in which 'nature' randomizes mothers into treatment and control groups. Intuitively, the idea is that the federal state-specific implementation timeline creates exogenous variation in the availability of child care, effectively creating a treatment and a control group composed of mothers who – on average – share the same characteristics. Thus, the core methodological question of this thesis is whether as-if randomization was successful, i.e. whether the G8 reform truly created a natural experiment akin to a RCT.

Several factors suggest that my setup is in fact a successful natural experiment. Most importantly, as the descriptive statistics according to treatment status provided in Table 3 show is that prior to treatment, no statistically significant differences between the observables were found. The only exception was the 'age'-variables, although, as I argued, these differences can be explained by the setup (see also Bauernschuster and Schlotter (2015)). Furthermore, one could even argue that the G8 reform introduced 'double' randomization because randomization happened via two channels: first, mothers within the *same* federal state are divided based on whether their child is young enough to be affected by the reform; second, mothers whose children are enrolled in Gymnasium in the same survey year are randomized by living in different federal states with different G8-implementation years. In other words, apart from mothers' federal state of residency, mothers are also randomized into treatment and control groups based on the school enrollment year of their youngest child.

A way in which randomization could be violated is through self-selection, for instance, when survey respondents join the treatment group because they anticipate positive effects from the treatment. Consequently, estimated effects can no longer be attributed to the treatment but might be due to pre-existing differences between these groups. In my set-up, however, I find it very implausible that mothers self-select themselves into treatment and control groups. This is because it seems very unlikely that an entire household would relocate to a different federal state to avoid the G8 reform, given that all federal states eventually implemented the reform. Furthermore, moving involves high financial and emotional costs and I find it highly implausible that moving decisions are determined by the educational reform.

All things considered, I therefore think that my setup is a natural experiment, i.e. a setting in which the increased availability of childcare is randomized amongst mothers. First, the event study plots showed no differential trends in the outcome variables before the first treatment year. Second, apart from the 'age'-variables, I found no statistically significant differences when conducting a baseline imbalance test. Third, randomization happens on two levels (individual and federal). Last, I find it very unlikely that mothers self-select into treatment or change their behavior in anticipation of treatment.

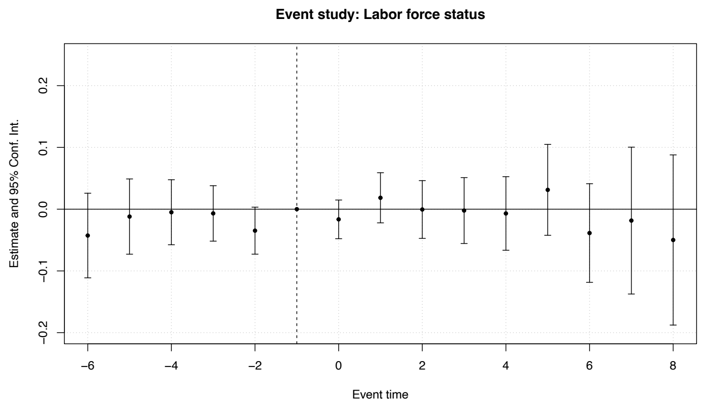
5.4. Findings

The results of my OLS estimates of equations (5.1), (5.2), and (5.3), are provided in Table 4 and Table 5 for the outcome variables *labor force participation* and *weekly working hours*, respectively. The results of the event study regression model (5.4) are given by the event study plots 7 (*labor force participation*) and 8 (*weekly working hours*). For both outcome variables *labor force participation* and *weekly working hours*, I begin by discussing the results of the event study model, as it accounts for the dynamic timing of treatment and provides the most refined model of my thesis.

5.4.1. Labor force participation

The event study plot depicted in Figure 7 illustrates the changes in labor force participation relative to the timing of treatment. Thus, the y-axis represents the estimated effects (coefficients) on the outcome variable of interest (*labor force participation)*, whereas the x-axis represents time relative to the treatment in years.

Figure 7.: Event Study: Labor Force Participation



Note: This figure shows coefficient estimates from equation (5.4) for the outcome variable labor force participation. The coefficients represent the change in outcomes for mothers whose children were affected by the G8 reform relative to those who were not six years before and eight years after the implementation of the G8 reform, as compared with the year immediately prior to implementation.
Source: SOEP.

Before the implementation of the G8 reform, the event study coefficients are statistically insignificant and nearly zero across all pre-implementation years, suggesting that there were no differential trends between the treatment and control groups leading up to the reform. This lack of significant pre-treatment differences supports the validity of the parallel trends assumption (see Section 6). The estimates remain insignificant and close to zero in the years after the reform implementation, indicating that the G8 reform did not have an impact on maternal labor supply on the *extensive* margins: the G8 reform (and thus the extended school days) did not positively impact the likelihood of mothers entering the labor market. Furthermore, as can be seen, standard errors increase as the timing of events moves further away from the first treatment year and fewer available observations are available.

The findings of the event study regression are further corroborated by the results of my OLS estimates of equations (5.1), (5.2), and (5.3), as displayed in Table 4. The first column of Table 4 shows the regression results for the pooled model, the second column the regression results of the one-way fixed effects model where I only control for state fixed effects, and the third column shows the estimates of the two-way fixed effects model where I control for both state and time fixed effects.

Across all models, education and part-time work experience are the strongest positive and statistically significant predictors of maternal labor force participation. In numbers, all else being equal, for every additional year of education (part-time work experience), the likelihood of participating in the workforce increases by 3.8 (4.5) percentage points, suggesting a positive relationship between educational attainment (work experience) and labor force participation.

Conversely, the strongest negative (and statistically significant) predictor of maternal labor force participation is marital status, indicating that being married reduces the likelihood of being in the labor force by about 12.5 percentage points. Moreover, the number of children is also negatively associated with maternal work engagement. Age, modeled through linear and quadratic terms, shows that the positive effect age has on labor force status diminishes as individuals become older, although only marginally significant from an economic point of view. Apart from partner's education, partner characteristics such as labor force status or migration background do not have a statistically significant effect on maternal labor force participation. Notably, the parameter estimates are quite consistent in all three models, with a tendency for effects to become slightly smaller in the two-way fixed effects model. The overall fit of all three models, as indicated by the R-squared values, is for all models at around 0.33, suggesting that around 30% of the variation in the outcome variable can be explained by the models.

Regarding the main variable of interest, the G8 reform, the regression results present a mixed picture. According to the pooled and one-way fixed effects models, the reform has a statistically significant and positive effect on maternal labor force participation of around 0.036, meaning that the G8 reform (and thus the extended school days implicit in the reform) increases the likelihood of being in the labor force on average by about 3.6 percentage points. However, this effect disappears when accounting for both time and state-fixed effects in the two-way fixed effects model. Considered together and given the susceptibility of Pooled OLS to bias and inconsistency, as well as the likely presence of state-fixed effects, the findings suggest that extended school days at high school age have no significant effect on maternal labor force participation.

Thus, taking into account both the findings of the fixed effects regression, as well as the event study regression, the introduction of the G8 reform and the consequent extension of school days appear to have had no noticeable impact on the *extensive* margins: mothers did not significantly increase their labor force participation in response to the reform.

Table 4.: Regression Results: Labor Force Participation

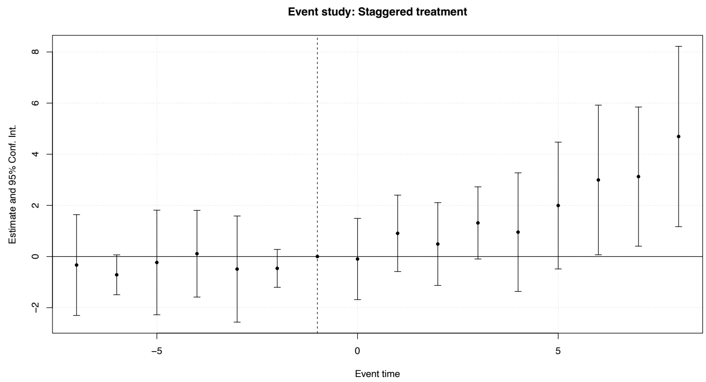
	Dependent variable:		
		Labor Force Participation	
	Pooled OLS	One-Way Fixed Effects	Two-Way Fixed Effects
	(1)	(2)	(3)
Education	0.038***	0.037***	0.037***
	(0.003)	(0.003)	(0.003)
Experience (Full-Time)	0.019***	0.018***	0.020***
	(0.003)	(0.005)	(0.004)
Experience (Part-Time)	0.045***	0.045***	0.045***
	(0.002)	(0.0004)	(0.0003)
Experience (Full-Time) Squared	0.0004***	0.0004**	0.0003**
	(0.0001)	(0.0001)	(0.0001)
Age	0.063***	0.063***	0.058***
	(0.010)	(0.009)	(0.009)
Age Squared	−0.001***	−0.001***	−0.001***
	(0.0001)	(0.0001)	(0.0001)
Migration Background	0.014	0.017	0.014
	(0.020)	(0.034)	(0.034)
G8 Reform	0.036**	0.037***	0.003
	(0.015)	(0.011)	(0.011)
Regional Type	−0.028	−0.027***	−0.022**
	(0.019)	(0.008)	(0.009)
Marital Status	−0.125***	−0.125***	−0.121***
	(0.027)	(0.012)	(0.014)
Number of Children	−0.024**	−0.025***	−0.023***
	(0.009)	(0.007)	(0.007)
Partner's Labor Force Status	−0.041	−0.040	−0.038
	(0.027)	(0.048)	(0.047)
Partner's Education	−0.008***	−0.008***	−0.009***
	(0.003)	(0.002)	(0.002)
Partner's Migration Background	−0.031	−0.035	−0.036
	(0.022)	(0.026)	(0.026)
Constant	−0.846***		
	(0.228)		
Observations	10,397	10,397	10,397
R^2	0.339	0.337	0.317
Adjusted R^2	0.338	0.336	0.314
F Statistic	380.034*** (df = 14; 10382)	376.722*** (df = 14; 10374)	343.081*** (df = 14; 10357)

Notes: This table displays the estimates of equations (5.1)-(5.3) using SOEP data for the years 1995-2012. Standard errors, reported in parentheses, are clustered at the state level and robust to heteroscedasticity. Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

5.4.2. Weekly working hours

The estimates of the event study regression model (equation 5.4) are presented in Figure 8. Again, the x-axis represents the years relative to treatment and the y-axis the estimated coefficients. As shown in Figure 8, before the implementation of the G8 reform, weekly working hours trended similarly across treatment and control groups: the pre-implementation event study coefficients are not statistically significant and are almost zero for all pre-implementation years, again indicating no prior differences between the groups. However, following the first year of the reform's implementation, a small positive trend in weekly working hours emerges among mothers affected by the G8 reform compared to control mothers.

Figure 8.: Event Study: Weekly Working Hours



Note: This figure shows coefficient estimates from equation (5.4) for the outcome variable weekly working hours. The coefficients represent the change in outcomes for mothers whose children were affected by the G8 reform relative to those who were not seven years before and 8 years after the implementation of the G8 reform, as compared with the year immediately prior to reform implementation. *Source*: SOEP.

Although the effect sizes are relatively small during the first five years after implementation, treated mothers work approximately one additional hour per week compared to control mothers although these effects are only statistically significant at the 10%-level and include effects that are not meaningful from an economic point of view. In years 6, 7, and 8, I find the first statistically significant effects at the 5%-level of the G8 reform on weekly working hours, with increases of about 2.5 to 3.5 additional hours per week. Yet the confidence intervals are relatively large and – on the lower bounds – include positive effects that are very small from an economic point of view. Overall, this suggests that if the reform had an impact on the weekly working hours of affected mothers, these effects are relatively small and emerge only several years after the implementation of the reform.

The estimation results of equations (5.1), (5.2), and (5.3) for the outcome variable *weekly working hours* are presented in Table 5.[8] Accordingly, education consistently shows a positive and significant effect on weekly working hours in all three models, with coefficients of around 1.2, indicating that one year of education is associated with slightly more than one additional work hour per week. Full-time and part-time work experience also show positive impacts, with part-time experience being particularly strong and highly significant. In contrast, the strongest negative predictors are marital status and partner's labor force status, indicating that married women and women whose partner is in the labor force work on average around 4.5 to 4.9 hours less per week than unmarried women or women whose partners do not work. Interestingly, the number of children has no significant impact on weekly working hours in all models.

For all variables except age, regional type, and mother's migration background, the estimates across all models are very similar. However, for the main variable of interest, the impact of the G8 reform on *weekly working hours*, the parameter estimates show relatively big differences in size across the models, with coefficients ranging from -1.49 to -2.57. Surprisingly, the negative signs indicate that the G8 reform is associated with a *reduction* (although relatively small in size) in weekly working hours, particularly in the two-way fixed effects model. At first sight, this seems to suggest that the reform may have led to a decrease in the number of hours worked by mothers, contrary to what I expected and what I found in the event study regression. However, it should be noted that models (5.1), (5.2), and (5.3) do not consider the staggered adoption of the reform across different federal states and the specific effects relative to the implementation year. This I'll discuss extensively in the next Section 6, where I investigate the heterogeneity in the results and discuss my findings from a policy perspective.

[8] The sample size is slightly smaller than for outcome variable *labor force participation* because of missing values in *weekly working hours*.

Table 5.: Regression Results: Weekly Working Hours

	Dependent variable:		
		Weekly Working Hours	
	Pooled OLS	One-Way Fixed Effects	Two-Way Fixed Effects
	(1)	(2)	(3)
Education	1.276***	1.202***	1.216***
	(0.139)	(0.151)	(0.150)
Experience (Full-Time)	0.192**	0.181***	0.176***
	(0.078)	(0.045)	(0.046)
Experience (Part-Time)	0.859***	0.835***	0.832***
	(0.135)	(0.140)	(0.148)
Experience (Full-Time) Squared	0.001	0.0001	0.001
	(0.005)	(0.005)	(0.005)
Age	−1.009***	−0.978***	−0.608***
	(0.351)	(0.214)	(0.213)
Age Squared	0.008*	0.008***	0.003
	(0.004)	(0.003)	(0.003)
Migration Background	1.710**	1.905***	1.953***
	(0.824)	(0.572)	(0.612)
G8 Reform	−1.496**	−1.201**	−2.572***
	(0.618)	(0.608)	(0.642)
Regional Type	0.490	0.141	0.236
	(0.866)	(1.086)	(1.069)
Marital Status	−4.926***	−4.586***	−4.531***
	(1.182)	(1.025)	(0.986)
Number of Children	−0.070	0.029	0.019
	(0.430)	(0.461)	(0.455)
Partner's Labor Force Status	−4.578***	−4.657***	−4.603***
	(0.992)	(0.940)	(0.949)
Partner's Education	−0.261**	−0.242**	−0.221*
	(0.129)	(0.109)	(0.117)
Partner's Migration Background	1.590*	1.636**	1.644**
	(0.920)	(0.666)	(0.645)
Constant	35.818***		
	(7.616)		
Observations	6,889	6,889	6,889
R^2	0.224	0.201	0.204
Adjusted R^2	0.222	0.199	0.199
F Statistic	141.521*** (df = 14; 6874)	123.735*** (df = 14; 6866)	125.268*** (df = 14; 6849)

Notes: This table displays the coefficient estimates of equations (5.1)-(5.3) using SOEP data for the years 1995-2012. Standard errors, reported in parentheses, are clustered at the state level and robust to heteroscedasticity. Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

6. Discussion

The regression tables above showed mixed results. While the G8 reform did not impact maternal labor supply at the *extensive margins*, i.e., the G8 reform did not affect the likelihood of mothers *entering* the labor market, I found differential effects on the *intensive margins*. Whereas the pooled and fixed effects models show small negative effects of the reform on weekly working hours, the event study regression shows small positive, yet delayed, effects. Thus, in this chapter, I discuss and evaluate the reliability of these findings within the broader context of the existing literature. In Section 6.1, I address the heterogeneity in outcomes for the variable *weekly working hours* and discuss effect sizes. In Section 6.2, I turn to the policy implications of my findings and how they relate to the findings of other researchers. Last, in Section 6.3, I consider some main limitations of my research design and explore some ideas for further research.

6.1. Treatment effects

Both fixed effects and event study regressions show similar results for the outcome variable *labor force participation*: the childcare subsidy implicit in the lengthening of the school day through the G8 reform appears to have no positive impact on maternal decisions to enter the workforce. I find two explanations for this absence of effects most plausible. First, the extension of school days might simply be too small to significantly influence decisions regarding entering the labor force. For many mothers, the slight increase in available child care may not have offered sufficient additional capacities to seek employment. Second, it could be that mothers who are not already back in the labor force when their youngest child reaches high school might simply prefer not to (re-)enter the labor force. Several reasons could contribute to this decision, for instance, the fulfillment of other personal or family responsibilities, sufficient income from their partner, or the lack of suitable job opportunities that align with their needs and skills.

Regarding the outcome variable *weekly working hours*, I found negative effects in the fixed effects models and positive effects in the event study regression model. The first explanation for this heterogeneity refers to the inherent differences between the models

I estimated. Whereas in the fixed effects models, I controlled for federal state and/ or time-fixed effects and used indicator terms that equaled 1 if the enrollment year into Gymnasium t_{is} of mother $i's$ youngest child in state s is larger than or equal to the year the G8 reform was implemented in s, t_s^*. In the event study regression, I also included fixed effects; yet, I estimated treatment effects *relative* to the first period before implementation, t_{-1}, which was omitted from the analysis.

Another explanation for the observed heterogeneity in findings could be negative weighting issues associated with staggered treatment adoption. Very recently, some researchers in econometric methodology have cast doubt on the practice of extrapolating the two-way fixed effects estimator (hereafter TWFE) from the standard two-group/two-period settings to situations where treatment is staggered, i.e. treatment is not given to all units at once but at different periods in time. The problem, as authors like De Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), and Sun and Abraham (2021) have shown is that extrapolating the TWFE estimator from the canonical two-by-two setup to staggered adoption relies on relatively strong assumptions that may not be met in all setups. Furthermore, in scenarios where treatments are adopted at different times across units, fixed effects models can produce biased estimates.

Intuitively, the problem is that in a staggered roll-out setting the TWFE estimator is a weighted average of all possible 2x2 Difference-in-Differences estimators where the weights depend on the sample composition and the treatment timings Roth et al. (2023). However, these authors argue that not all these possible 2x2 comparisons are justified, insofar as "TWFE regressions make both "clean" comparisons between treated and not-yet-treated units as well as "forbidden" comparisons between units who are both already-treated" (Roth et al. (2023), 2219).

Roth et al. (2023) discuss the example of the Affordable Care Act (ACA), which significantly expanded health insurance coverage for two groups, one early and one late-treated group. Accordingly, the TWFE-DiD estimator is a weighted average of all two-by-two comparisons: early vs. untreated, late vs. untreated, early vs. late, *before* late is treated; and early vs. late, *after* late is treated. Whereas the first three comparisons are unproblematic, in the last scenario, units that have received the treatment early serve as controls for those that are treated late. This comparison, however, would only be valid when assuming parallel trends of the *treated* outcome which Sun and Abraham (2021) show is a very strong assumption. In my setup, for instance, this would imply that even after treatment, the labor supply of now-treated and control mothers evolved in parallel. Potentially, this negative weighting distorts the average treatment effect, resulting in both positive and negative signs in the models I estimated.

Therefore, the most refined model in my thesis is the event study regression model, which measures the change in average weekly working hours for mothers with children affected by the G8 reform relative to mothers not affected by the reform, starting from the year immediately before the implementation, t_{-1}. Interestingly, as Figure 8 shows, the first significant effects on weekly working hours appear six years after implementation, although the confidence intervals are relatively large and include economically insignificant effects. The effects become more pronounced in years seven and eight, corresponding to the period when children enter the so-called *gymnasiale Oberstufe*, which comprises grades 11 and 12.

One possible explanation for this delay in observed effects could be related to gender norms. As discussed in Section 2.1, women are still often expected to carry the bigger share of care responsibilities; yet, this expectation (as well as the care needs of children) may decrease as children become older and thus are near the end of school. Consequently, women may increase working hours as children become older. However, as described above, the event study model estimates treatment effects relative to control mothers who are – in case of successful randomization – similarly affected by these norms and aging effects. Therefore, I think that these factors do not drive the observed effects.

Instead, I think a much more plausible explanation for the delayed effects is that many federal states decided to spread the school hours of the former 13^{th} school year unevenly to avoid imposing excessively long days on younger students (see Bundeszentrale für politische Bildung (2015)). Consequently, in many federal states, students in lower grades continued to have school days similar to those in the former G9 system, whereas students in the *gymnasiale Oberstufe* had significantly more hours than before. This uneven spreading of the additional hours necessary to compensate for the one-year loss in total education likely contributed to the observed delay in effects: since many federal states spread the additional hours to the *gymnasiale Oberstufe* which starts six years after students enter *Gymnasium*, only then many students experienced substantial increases in school hours. This timing aligns with the point at which I first found significant effects.

6.2. Policy implications

What can be taken away from the findings and how do they relate to other studies conducted in Germany? Generally, my findings are quite consistent with the existing research in Germany on the effects of increased childcare availability on maternal labor supply in Germany (see Section 2.2).

For instance, although Müller and Wrohlich (2020) identified positive and statistically significant effects on maternal labor supply on the extensive margins resulting from the daycare expansion law and the law on support for children, these effects are very small: they found that a 1 percentage point increase in childcare places leads to only a 0.2 percentage point rise in mothers' labor market participation rate.

Another important study in Germany was conducted by Bauernschuster and Schlotter (2015) who examined the effects of a policy introducing a legal claim to a place in kindergarten for children aged three to four. They discovered that this policy led to a notable increase in maternal labor supply by about 6 percentage points. Thus they found relatively large effects even on the extensive margins. However, this difference in findings can likely be explained by the fact that the children Bauernschuster and Schlotter (2015) studied were much younger. Unlike my study, which focuses on children above the age of ten, Bauernschuster and Schlotter (2015) looked at children at kindergarten age. As younger children require more intensive care, more readily available childcare offerings for this age group are likely to have a bigger effect on a mother's ability to participate in the labor force. Furthermore, the legal claim to a place in kindergarten provided significantly more relief in terms of hours compared to the extended school days, which could also explain why their effects are more pronounced.

One of the few studies that investigated the impact of childcare on maternal labor supply was the study of Contreras and Sepúlveda (2017) who looked at the effects of the expansion of the Chilean school day until the afternoon for children between 8 and 13 years. They found a "5 percent increase in labor participation and employment rates of single mothers with eligible children (between 8 and 13 years old) with no younger children" (Contreras and Sepúlveda (2017), 747). Again, one main difference between this study and mine is that children were much younger and the reform was more extensive (previously school in Chile lasted until 1 p.m., whereas after the reform, it lasted until 4 p.m.). Yet, they did not find any effects on married mothers or families with younger children.

From a broader perspective, the findings of my thesis and related studies discussed above and in Section 2.2 show that increases in childcare availability do not necessarily translate into substantially higher female labor force participation at the *extensive margins*. What seems to matter most is the age of the children affected by various policies: more (or cheaper) childcare offerings are likely to have bigger effects on female labor force participation the younger the children are. Unfortunately, less evidence exists at the *intensive margins*, i.e., whether women increase their working hours (as opposed to entering the labor force altogether) with more available childcare. One possible explanation could be that it is much easier to inquire about employment status compared to detailed information on actual weekly working hours in surveys. Yet, as I have tried to show is that extending childcare availability, e.g. through longer school days, can still have small positive effects on the weekly working hours for mothers already employed.

In conclusion, increasing childcare offerings – especially for older children – are likely not to have huge effects on the likelihood of mothers entering the labor force, but may still lead to some increases in work hours. Ultimately, thus, achieving gender equality in the labor market cannot be accomplished solely through expanding childcare options to help women balance care responsibilities and paid work. Instead, gender equality in the labor market requires more diverse policies that also address other structural and societal factors, such as workplace policies or gender norms to support women's equal participation in the labor market and close existing gender employment gaps.

6.3. Limitations and further research

While these policy implications highlight that many different policies are necessary to support gender equality in the labor market, it is important to mention some further limitations of my study and consider areas for further research before I conclude. The main limitation of my research design is that I could not use mother-child observations from two relatively big federal states (fifth and sixth largest federal states as of 2022), Rhineland Palatinate and Hesse. This is because Rhineland Palatinate did not universally implement the G8 reform but started a pilot scheme in selected schools only and Hesse implemented the reform over three years. Furthermore, as discussed extensively in Section 5, the individual schooling histories were incomplete or missing for some children. Therefore, I used imputed school enrollment years for some children which could bias the estimated effects. Yet, as argued in Section 5 when studying the accuracy of imputed Gymnasium enrollment years and actual Gymnasium enrollment years for children for whom schooling histories are given, imputed and actual values match for a large majority of children. Furthermore, in the Appendix A.2, I showed that my estimates are robust when using only mother-child observations for whom the individual schooling histories are given.

As for further research, studies could investigate how the increased availability of childcare offerings affects different subgroups of mothers, such as single versus married mothers, and those with varying levels of qualifications or education. Furthermore, it would be interesting to see if there are similar policies in other countries with different labor markets (e.g., developing countries or countries with higher/ lower female employment) to gain a more comprehensive understanding of how childcare policies influence maternal labor participation across diverse settings. Lastly, it could be interesting to investigate how cultural attitudes toward gender roles and caregiving responsibilities interact with childcare policies.

7. Conclusion

In this thesis, I studied how the extension of school days for children above the age of ten affects maternal labor supply. To estimate treatment effects, I took advantage of the federalism in educational policy in Germany exploiting the variation induced by the differential timing in the implementation of the G8 reform. At the core of my thesis lies the idea that mothers who live in federal states where the reform was not yet implemented or whose children were enrolled in the academic schooling track before the reform was implemented serve as the control group for treated mothers. Thus, as I argued, the reform effectively created a *natural experiment*.

To estimate treatment effects, I utilized an extended Difference-in-Differences (DiD) approach using data from the German Socioeconomic Panel (SOEP) for the years 1995-2012. At the *extensive margins*, my estimates suggest that the extended school days did not affect maternal labor supply. In other words, women who were not in the labor force before the implementation of the G8 reform did not change their labor force status by entering the labor market in response to the reform. Yet, at the *intensive margins*, reducing childcare burdens through the extended school hours may have small positive, although delayed, effects of about 4 additional hours per week, although the confidence intervals included effects that were not significant from an economic point of view. Most plausibly, this delay in effects can be explained by the fact that many federal states decided to spread the additional hours necessary to compensate for the one-year loss in total education to later grades.

To corroborate my results, I conducted a baseline imbalance test which indicated no differences between treatment and control mothers. Furthermore, my pre-implementation event study coefficients were not statistically significant. Last, I showed that estimated treatment effects do not change much in size when only using mother-child observations with complete schooling histories. Thus, they provide support for my baseline results, where I also use data from mothers for whose children I imputed the year of Gymnasium enrollment. Yet, the study also has several limitations. First, some observations were lost due to incomplete data and the implementation schedules by Rhineland-Palatinate and Hesse. Second, TWFE estimation in staggered adoption settings can suffer from negative weighting issues; which I offered as an explanation for the heterogeneity in estimated treatment effects.

From a policy perspective, also taking into account my descriptive analysis and previous literature, women remain underrepresented in the labor market, especially when children are below compulsory schooling age. Thus, increasing childcare availability appears to be most effective when children are younger. Yet, many other factors apart from childcare availability also influence maternal labor market participation. A relatively new research field, right at the intersection of philosophy and economics, is the study of how social and Gender norms relating to the division of domestic care responsibilities and the societal acceptability of maternal work impact work decisions. As some researchers have argued, offering better childcare options or providing families with financial incentives to share caregiving responsibilities more equally are important steps, but they alone may not suffice, as societal norms can also significantly impact work decisions.

These norms can hinder a mother's career progression, for instance, when dictating that mothers should carry (bigger parts of) a household's care responsibilities. Thus, a relatively new research field, right at the intersection of philosophy and economics, is emerging which studies how Gender norms relating to the division of domestic care responsibilities and the societal acceptability of maternal work impact work decisions. From these studies, it becomes more and more evident that Gender equality in the labor market cannot solely be achieved through more readily available childcare options but must also challenge deeply ingrained gender norms.

Bibliography

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, 24(2):3–30.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116(4):709–745.

Bauernschuster, S. and Schlotter, M. (2015). Public child care and mothers' labor supply—Evidence from two quasi-experiments. *Journal of Public Economics*, 123:1–16.

Berthelon, M., Kruger, D. I., and Oyarzun, M. A. (2015). The effects of longer school days on mothers' labor force participation. *IZA Discussion Paper Series*.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.

Bicchieri, C., Muldoon, R., and Sontuoso, A. (2023). Social Norms. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Stanford University, Winter 2023 edition.

Bundeszentrale für politische Bildung (2013). Das Bildungssystem in Deutschland. https://www.bpb.de/themen/bildung/dossier-bildung/163283/das-bildungssystem-in-deutschland/. Accessed April 2024.

Bundeszentrale für politische Bildung (2015). Der Streit um "G8": Kürzere Schulzeit, mehr Stress, weniger Bildung? Positionen und Befunde zur Schulzeitdebatte. https://www.bpb.de/themen/bildung/dossier-bildung/211182/der-streit-um-g8-kuerzere-schulzeit-mehr-stress-weniger-bildung/. Accessed June 2024.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Contreras, D. and Sepúlveda, P. (2017). Effect of lengthening the school day on mother's labor supply. *The World Bank Economic Review*, 31(3):747–766.

Cortés, P., Koşar, G., Pan, J., and Zafar, B. (2022). Should mothers work? How perceptions of the social norm affect individual attitudes toward work in the US. Technical report, National Bureau of Economic Research.

Dahmann, S. C. and Anger, S. (2014). The impact of education on personality: Evidence from a German high school reform. *IZA Discussion Paper Series*.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press.

Elster, J. (2015). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge University Press.

European Commission (2022). Women's situation in the labour market. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/women-labour-market-work-life-balance/womens-situation-labour-market_en#gender-segregation-in-the-labour-market. Accessed April 2024.

Fraser, N. (2016). Contradictions of Capital and Care. *New Left Review*.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, 239(2):345–360.

Goldin, C., Kerr, S. P., and Olivetti, C. (2022). When the kids grow up: Women's employment and earnings across the family cycle. Technical report, National Bureau of Economic Research.

Havnes, T. and Mogstad, M. (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, 95(11-12):1455–1465.

Hessisches Kultusministerium (2010). Fragen und Antworten zu G8. https://www.steb-wiesbaden.de/wp-content/uploads/G8-Brosch%C3%BCre.pdf.

Hessisches Kultusministerium (n.d.). Änderung Schulgesetz: Wahlfreiheit zwischen G8 und G9. https://kultus.hessen.de/schulsystem/schulformen-und-bildungsgaenge/gymnasium/g8-und-g9. Accessed: July 2024.

Hodges, M. J. and Budig, M. J. (2010). Who gets the daddy bonus? Organizational hegemonic masculinity and the impact of fatherhood on earnings. *Gender & Society*, 24(6):717–745.

Huebener, M. and Marcus, J. (2015). Moving up a gear: The impact of compressing instructional time into fewer years of schooling. Technical report, DIW.

Kleven, H. (2022). The geography of child penalties and gender norms: Evidence from the United States. Technical report, National Bureau of Economic Research.

KMK (2023). Education in Germany 2023. https://www.kmk.org/fileadmin/Dateien/pdf/Dokumentation/en_2023.pdf. Accessed April 2024.

Müller, K. and Wrohlich, K. (2020). Does subsidized care for toddlers increase maternal labor supply? Evidence from a large-scale expansion of early childcare. *Labour Economics*, 62:101776.

Nussbaum, M. C. (2001). Symposium on Amartya Sen's philosophy: 5 adaptive preferences and women's options. *Economics and Philosophy*, 17(1):67–88.

Price, B. M. and Wasserman, M. (2023). The summer drop in female employment. Technical report, National Bureau of Economic Research.

Roig, M., Maruichi, D., et al. (2022). Old-age poverty has a woman's face. *United Nations Policy Brief no°142*.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

Satz, D. (2017). Feminist Perspectives on Reproduction and the Family. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2017 edition.

Schraml, P. (2018). Die Stichtagsregelung in den Bundesländern. https://www.bildungsserver.de/innovationsportal/bildungplusartikel.html?artid=1098. Accessed June 2024.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

UN Women (2023). Social Norms, Gender and Development: A Review of Research and Practice.

A. Appendix

A.1. R Code for Treatment Assignment

The following code shows treatment assignment for mothers whose state of residency is in Bavaria, Lower Saxony, Baden-Wuerttemberg, Bremen, Hesse, and Mecklenburg-Western Pomerania, where in 2004, the G8 reform was implemented. Based on the individual identifier, *pid*, and the survey year, *syear*, the code performs the following command. For all survey years before 2004, no woman was 'treated' as the reform was not yet implemented. For all years after 2004, two conditions apply:

- If a mother's youngest child entered the academic schooling track before 2004 (and thus before the implementation of the G8 reform), the mother is in the control group for all survey years (*never treated group*).
- If a mother's youngest child is enrolled in Gymnasium after 2004, the mother is 'treated' but only in survey years after her child's school enrollment year. For instance, when the household is located in Bavaria where the G8 reform was implemented in 2004 but her youngest child enters Gymnasium in 2006, the mother is treated only from the years 2006 onward.

The following Code A.1 shows how I assigned mothers to treatment and control groups for one exemplary year, 2004, where Bavaria ('by'), Lower Saxony ('ni'), Baden-Wuerttemberg ('bw'), Bremen ('hb'), Hesse ('he') and Mecklenburg West-Pomerania ('mvp') implemented the G8 reform.

```
# Treatment for federal states implementation year 2004
data_main <- data_main %>%
  group_by(pid, syear) %>%
  mutate(
    treatment_status_2004 = case_when(
      bula %in% c("by", "ni", "bw", "hb", "he", "mvp") & syear < 2004 ~ 0,
      bula %in% c("by", "ni", "bw", "hb", "he", "mvp") & syear >= 2004 & youngest_child_eyeargym < 2004 ~ 0,
```

```r
      bula %in% c("by", "ni", "bw", "hb", "he", "mvp") & syear >= 2004 & youngest_child_eyeargym >= 2004 ~
        case_when(
          syear >= youngest_child_eyeargym ~ 1,
          TRUE ~ 0
        )
    )
  ) %>%
  ungroup()
```

Listing A.1: Treatment Assignment in R

A.2. Robustness

Tables 6 and 7 present the regression results for the variables *labor force participation* and *weekly working hours*, only using mothers for whom the Gymnasium entry year of their youngest child could be identified through individual schooling histories. Thus, they exclude mothers for whose children the Gymnasium entry year was imputed. Unexpectedly, the number of observations decreases because fewer mothers can be used in the analysis when only considering children with complete schooling histories. As can be seen, the results are robust when restricting the sample to mother-child pairs for whom individual schooling histories are given, especially for the outcome variable *labor force participation*: the effects have the same sign for all variables and effect sizes are very similar when comparing to the estimates using the 'full' sample. The regression table for the outcome variable *weekly working hours* shows a bit more variation in effect sizes; yet, the signs of the effects and their significance levels are the same in both specifications.

Table 6.: Regression Results: Labor Force Participation (unimputed)

	Dependent variable:		
		Labor Force Participation	
	Pooled OLS	One-Way Fixed Effects	Two-Way Fixed Effects
	(1)	(2)	(3)
Education	0.039***	0.038***	0.038***
	(0.004)	(0.004)	(0.004)
Experience (Full-Time)	0.023***	0.023***	0.025***
	(0.004)	(0.005)	(0.005)
Experience (Part-Time)	0.048***	0.049***	0.048***
	(0.003)	(0.001)	(0.001)
Experience (Full-Time) Squared	0.0003**	0.0003*	0.0003
	(0.0001)	(0.0002)	(0.0002)
Age	0.058***	0.058***	0.053***
	(0.013)	(0.010)	(0.010)
Age Squared	−0.001***	−0.001***	−0.001***
	(0.0002)	(0.0001)	(0.0001)
Migration Background	0.005	0.009	0.007
	(0.021)	(0.040)	(0.041)
G8 Reform	0.046***	0.048***	0.011
	(0.017)	(0.014)	(0.016)
Regional Type	−0.027	−0.017	−0.012
	(0.022)	(0.011)	(0.013)
Marital Status	−0.132***	−0.130***	−0.126***
	(0.029)	(0.024)	(0.026)
Number of Children	−0.026**	−0.026**	−0.024**
	(0.011)	(0.010)	(0.010)
Partner's Labor Force Status	−0.061**	−0.059	−0.055
	(0.029)	(0.050)	(0.050)
Partner's Education	−0.011***	−0.011***	−0.011***
	(0.003)	(0.003)	(0.003)
Partner's Migration Background	−0.021	−0.024	−0.026
	(0.024)	(0.027)	(0.027)
Constant	−0.690**		
	(0.285)		
Observations	8,146	8,146	8,146
R^2	0.349	0.346	0.324
Adjusted R^2	0.348	0.345	0.320
F Statistic	311.040*** (df = 14; 8131)	307.514*** (df = 14; 8123)	277.194*** (df = 14; 8106)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7.: Regression Results: Weekly Working Hours (unimputed)

	Dependent variable:		
		Weekly Working Hours	
	Pooled OLS	One-Way Fixed Effects	Two-Way Fixed Effects
	(1)	(2)	(3)
Education	1.230***	1.132***	1.142***
	(0.167)	(0.150)	(0.147)
Experience (Full-Time)	0.214**	0.207***	0.202***
	(0.094)	(0.072)	(0.074)
Experience (Part-Time)	0.890***	0.857***	0.849***
	(0.165)	(0.124)	(0.131)
Experience (Full-Time) Squared	0.002	0.002	0.003
	(0.006)	(0.004)	(0.005)
Age	−1.363***	−1.350***	−0.942***
	(0.397)	(0.267)	(0.238)
Age Squared	0.012**	0.012***	0.007**
	(0.005)	(0.003)	(0.003)
Migration Background	0.756	0.981*	1.055*
	(0.850)	(0.562)	(0.595)
G8 Reform	−0.686	−0.400	−1.355**
	(0.698)	(0.446)	(0.619)
Regional Type	1.240	0.775	0.806
	(0.976)	(1.272)	(1.225)
Marital Status	−5.229***	−4.859***	−4.879***
	(1.235)	(1.208)	(1.194)
Number of Children	−0.315	−0.150	−0.156
	(0.480)	(0.457)	(0.453)
Partner's Labor Force Status	−5.134***	−5.212***	−5.094***
	(1.149)	(0.965)	(0.991)
Partner's Education	−0.429***	−0.399***	−0.380***
	(0.148)	(0.105)	(0.116)
Partner's Migration Background	2.236**	2.185**	2.223**
	(1.072)	(0.993)	(0.966)
Constant	47.031***		
	(8.487)		
Observations	5,220	5,220	5,220
R^2	0.237	0.212	0.212
Adjusted R^2	0.235	0.209	0.206
F Statistic	115.767*** (df = 14; 5205)	99.873*** (df = 14; 5197)	99.710*** (df = 14; 5180)
Note:			*p<0.1; **p<0.05; ***p<0.01