This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Essays in Applied Micro-Econometrics

Brian Jagusch

THE UNIVERSITY of EDINBURGH

Doctor of Philosophy
School of Economics
The University of Edinburgh
2023
Declaration of Own Work

This thesis has been composed and written solely by myself and is the result of my own work unless clearly indicated and referenced. The work in this thesis has not been submitted for any other degree or professional qualification.

Brian Jagusch
Acknowledgements

The dictionary defines an adventure as a risky or unexpected undertaking, and my journey of pursuing a PhD has been exactly that. From embarking on every research idea with high hopes until ending up in unforeseen territory when concluding research projects, I am honoured to have had two outstanding supervisors, brilliant colleagues and an amazing partner by my side, navigating me through the currents of research.

I am deeply grateful to my supervisory team, Dr Andreas Steinhauer and Dr Steven Dieterle, for their invaluable advice and support. I want to thank all members of the School of Economics for attending my presentations and giving me constructive feedback to improve my work. In particular, I would like to thank my personal tutor, Dr Ina Taneva, for encouraging me to accept an offer to intern in economic consulting, which has broadened my horizons tremendously.

I would not have been able to complete this adventure without my fantastic colleagues and office mates Stuart Breslin and Kavisha Gupta, as well as fellow PhD students David Mesa-Ruiz and Adolfo Fuentes Werlinger, who helped me countless times figure out my research problems and be a sounding board for my ideas.

Amalie has been an understanding, kind, and loving partner. Thank you for always having my back, especially during challenging times throughout this journey. I would also like to thank my parents, Petra and Michael, for always supporting me, helping me reach my goals, and inspiring me to dream big. Onward to new adventures!
Abstract of Thesis

This thesis consists of three chapters using micro-econometric techniques to evaluate German public policy and policy reforms regarding their impact on individuals and the labour market. Although the selection of topics might seem broad, the general theme linking my results is an interest in informing policymakers about the efficacy of reforms frequently discussed in the public debate in Germany in recent years.

In chapter 1, I use administrative labour market data to estimate the effect of compressing instructional time in school on labour market outcomes for young adults. Leveraging the staggered introduction of the so-called G8 reform, I compare students on the German academic school track who graduated under a compressed curriculum to those who did not. Using a robust difference-in-differences strategy, I find that the reform reduced labour market earnings by 14% over the sample mean six years after graduating from high school. The decline is persistent and primarily driven by changes to the curriculum inducing higher weekly workloads, while age effects play a minor role. A range of alternative specifications and placebo tests confirm my main findings.

In chapter 2, I use German survey data to investigate whether married couples shift the timing of marital breakup to reduce income tax liabilities. In Germany, married couples are eligible for income tax rebates if they are married for at least one day in the calendar year. This creates an incentive to delay the breakup date into the next calendar year. I show that a large proportion of married couples report a
breakup in the first quarter of a year, especially in January, between 1984 and 2017. I find that a EUR 1,000 increase in the tax rebate is associated with an increase in the probability of a breakup occurring in the first quarter by 2.9 percentage points, which is consistent with spouses postponing their separation date in response to tax deductions. I find no differences in the probability of divorce regardless of the timing of the breakup.

In chapter 3, I estimate the effect of childcare provision on labour market outcomes of mothers with young children. Exploiting variation in the intensity of local childcare expansions, I find that mothers’ employment is positively associated with childcare expansions in the short run, but not in the long run. Using a novel triple DiD approach, I show that the effects of additional childcare places are heterogeneous depending on initial rates of childcare provision. As the effects of childcare on mothers’ outcomes differ between the short and the long run, vary by calendar year when fixing years since childbirth and are heterogeneous across initial levels of childcare, previous estimates derived from two-way fixed effect models could be biased.
Lay Summary

Evaluations of how public policy interventions affect behaviour are not only an important part of economic research but also inform policymakers about how to design future interventions. Although the selection of topics might seem broad, the general theme linking my results is an interest in informing policymakers about the efficacy of reforms frequently discussed in the public debate in Germany in recent years.

In the first chapter, I investigate how a reform to the German schooling system impacted the labour market outcomes of graduates who graduated under the new policy. The reform reduced the number of years a pupil spends on the academic track. The German population is the second-oldest by median age in Europe. As the aim of the policy is to increase payments to the social security system, it is important to understand how the reform impacted labour market outcomes. I find that the reform reduced labour market earnings by 14% over the sample mean six years after graduating from high school. This reduction persists for many years after graduating high school and cohorts that graduate many years after the reform was introduced. I show that most of the effect can be attributed to changes in the school curriculum changing the timing of when pupils learn certain subjects in school. The results in my paper are suggestive that the reform reduced the amount of knowledge and skills pupils learn at school. Therefore, policymakers face a trade-off between pupils graduating and entering the labour market earlier and lower labour earnings and human capital.
In my second chapter, I examine the relationship between a rebate in income tax liabilities and the decision of when to separate in Germany. Married couples can file their income taxes jointly, which reduces their overall tax burden. Importantly, married couples can only file for joint taxation if they are married for at least one day in the calendar year. Therefore, if a couple decides to end their marriage, they might want to wait until the next calendar year to start the divorce process. Besides the government missing out on tax revenue due to this rule, understanding the timing decision of marital breakups is important to economic welfare as a delay might increase the risk of domestic violence between partners. I show that between 1984 and 2017, a large proportion of married couples report a breakup in the first quarter of a year, especially in January. I find that a EUR 1,000 increase in the tax rebate is associated with an increase in the probability of a breakup occurring in the first quarter by 2.9 percentage points. I find no differences in the probability of divorce regardless of when a couple breaks up.

My third chapter studies the effect of childcare provision on labour market outcomes of mothers with young children. In the early 2000s, the German federal government provided large funds to finance expansions in the local supply of childcare for children between the ages of one and three. Federal funds trickle through to the local level at different speeds, leading to different intensities in childcare expansions. Using these differences across areas, I show mothers’ employment is positively associated with childcare expansions in the short run, but not in the long run. The effect also varies by how high the initial rates of childcare provision were. Importantly, my results show that the effect of childcare expansions on mothers differs along different dimensions, which might lead to biased estimates if not accounted for.
1.6.2 Robustness Checks ................................................. 24
1.6.3 Effects of the G8 Reform in the Longer Run .................. 27
1.7 How Does the G8 Reform Impact Labour Earnings? ............ 29
  1.7.1 Mechanism ....................................................... 29
  1.7.2 Channels ......................................................... 33
1.8 Conclusion ......................................................... 36
  References - Chapter 1 ............................................... 39
  Figures - Chapter 1 .................................................. 44
  Tables - Chapter 1 .................................................... 53
  Appendix 1A: Additional Figures and Tables ....................... 56

2 Income Taxes and the Timing of Marital Dissolution: Evidence from Germany ........................................ 65
  2.1 Introduction ....................................................... 65
  2.2 Income Tax Incentives and Marital Decisions ................. 69
  2.3 The German Income Tax System and Incentives on Marital Timing Decisions ........................................... 70
  2.4 Data and Descriptive Statistics .................................... 72
    2.4.1 Sample Selection ............................................. 72
    2.4.2 Calculating the Marriage Subsidy ............................ 73
    2.4.3 Descriptive Statistics ....................................... 75
  2.5 Empirical Strategy ................................................. 75
  2.6 Results ............................................................ 78
    2.6.1 Baseline Results .............................................. 78
    2.6.2 Robustness Checks .......................................... 80
    2.6.3 Characteristics of Delayers ................................ 83
  2.7 Are Delayers Less Likely to Divorce? .......................... 85
    2.7.1 Divorce Laws in Germany .................................... 85
    2.7.2 Probability of Divorce for Delayers and Non-Delayers .... 86
  2.8 Discussion and Conclusion ........................................ 88
3 Who cares? The Role of Childcare in Reducing Labour Market Inequalities
   3.1 Introduction .............................................. 114
   3.2 Empirical Studies on the Effect of Childcare on Mothers’ Labour Market Outcomes .............................................. 117
   3.3 Mothers and Childcare in Germany .............................................. 120
      3.3.1 Labour Market Participation of Mothers .............................................. 120
      3.3.2 Childcare Reform and Expansions .............................................. 121
   3.4 Data .............................................. 122
      3.4.1 Data on Labour Market Outcomes and Childcare .............................................. 122
      3.4.2 Sample Selection and Descriptive Statistics .............................................. 125
   3.5 Empirical Framework .............................................. 126
      3.5.1 Generalised Difference-in-Differences .............................................. 126
      3.5.2 Identification .............................................. 127
   3.6 Results .............................................. 129
      3.6.1 Baseline Results .............................................. 129
      3.6.2 Long-run Effects of Childcare on Labour Market Outcomes .............................................. 132
      3.6.3 Heterogeneous Effects by Levels of Childcare Coverage .............................................. 133
   3.7 Conclusion and Discussion .............................................. 137
   References - Chapter 3 .............................................. 140
   Figures - Chapter 3 .............................................. 144
   Tables - Chapter 3 .............................................. 148
   Appendix 3A: Additional Figures and Tables .............................................. 152
## List of Figures

1.1 The German School System ................................................. 44
1.2 Mathematics Competencies by Grade ................................. 45
1.3 Staggered Introduction of the *Gymnasium* in 8 Years (G8) Reform ... 46
1.4 Share of Pupils Finishing *Gymnasium* after 9 and 8 Years .......... 47
1.5 Effect of the G8 Reform on Yearly Labour Earnings ................. 48
1.6 Effect of the G8 Reform on Labour Market Outcomes ............... 49
1.7 Effect of the G8 Reform in the Longer Run .......................... 50
1.8 Effect of Curriculum Changes ........................................... 51
1.9 Effect of Teachers’ Attention Allocation .............................. 52
1A.1 Mathematics Competencies by Grade Translated to English ........ 56
1A.2 Observed and Predicted Year of School Entry ..................... 57
1A.3 Test Scores and Cohort Size around the Reform .................. 58
1A.4 Grade Progression ...................................................... 59
1A.5 Heterogeneity of the Effect of the G8 Reform in the Longer Run, Only University Graduates ................................. 60
1A.6 Heterogeneity of the Effect of the G8 Reform in the Longer Run, Only Non-university Graduates ............................... 61
2.1 Marriage Subsidy for Exemplary Within-Marriage Distributions of Taxable Household Income ................................. 94
2.2 Distribution of Month of Separation by Marital Status ............ 95
2.3 Marriage Subsidy by Quarter of Separation .......................... 96
2.4 Distribution of Month of Divorce by Country ........................ 97
2B.1 Declaration of Permanently Living Separately ................. 106
2B.2 Survivor and Cumulative Hazard Function Estimates by Quarter of Separation ................................................. 107
2B.3 Event-Study Results of Overall Life Satisfaction ................. 108
2B.4 Event-Study Results of Labour Earnings .......................... 109
2B.5 Event-Study Results of Reported Hours Worked ................. 110

3.1 Childcare Enrolment Rates of Under Three-year-olds in West Germany Between 2002 and 2017 ........................................ 144
3.2 Trends in Labour Market Outcomes Around the Years of Expansion in Childcare by Counties with Above and Below-median Growth .... 145
3.3 Types of Childcare Expansions ..................................... 146
3.4 Trends in Labour Market Outcomes Before and After Years of Expansions in Childcare by Expansion Type ........................ 147
3A.1 Robustness Check - Dynamic Treatment Effects of Childcare Expansions on Labour Market Outcomes ...................... 152
## List of Tables

1.1 Descriptive Statistics .................................................. 53
1.2 Baseline Results - Effect of G8 Reform on Labour Earnings and Employment .............................................................. 54
1.3 Effect of the G8 Reform on Labour Market Outcomes ............ 55
1A.1 Cut-off Dates for School Entry After 1997 ......................... 62
1A.2 Placebo Results - Effect of G8 Reform on Unaffected Tracks .... 63
1A.3 Effect of the G8 Reform on Number of Graduates on the Academic Track 64
2.1 Descriptive Statistics .................................................. 98
2.2 Baseline Results - The Incentive of the Marriage Subsidy on Postponing Separation to the First Quarter of a Year .................. 99
2.3 Complier-Characteristics Ratios for High and Low Marriage Subsidies 100
2.4 Probability of Divorce - Cox Proportional Hazard Model ........ 101
2B.1 Placebo Results - The Incentive of the Hypothetical Marriage Subsidy of Unmarried Couples on Postponing Separation Until the First Quarter of a Year ................................................................. 111
2B.2 Robustness Check - Alternative Specifications ................... 112
2B.3 Robustness Check - Bandwidth Analysis .......................... 113
3.1 Descriptive Statistics .................................................. 148
3.2 Baseline Results - Generalised Difference-in-Differences Specification 149
3.3 Long-run Results - Generalised Difference-in-Differences Specification 150
3.4 Difference-in-Difference-in-Differences Results - Effect Heterogeneity Across Initial Levels of Childcare Coverage .......................... 151
3A.1 Additional Results - Generalised Difference-in-Differences Specification Using Full-time Care Coverage . . . . . . . . . . . . . . . . . . . . . . . . . . . 153
Chapter 1

The Effect of Compressed Schooling on Early Career Outcomes

1.1 Introduction

Germany has the second-oldest population by median age in Europe, placing financial strain on the social security system and the labour market. In the early 2000s, policymakers searched for ways to take pressure off the pay-as-you-go pensions system and turned to the younger generations for solutions. German graduates were among the oldest compared to their peers internationally. Therefore, the German Conference of Ministers of Education (Kultusministerkonferenz) agreed to reduce the years a high school student spends on the academic track from nine to eight years. The goal was to allow graduates to accumulate human capital faster, enter the labour market earlier, and increase the pool of young, skilled workers (Huebener and Marcus, 2015). Notably, the reform aimed to sustain the level of human capital upon graduating by keeping the curriculum content unchanged.

How does compressing education affect labour market outcomes? One strand of literature investigates the changes to school structuring choices where learning content changes and find that reducing the age at labour market entry, fewer
years for degrees, and reduced compulsory schooling have negative consequences on labour market outcomes (Card, 1999; Bedard and Dhuey, 2012; Morin, 2013; Krashinsky, 2014). In contrast, the literature investigating the effect of compressing education is small. Pischke (2007) finds that compressing two school years into one to shift the beginning of the school year to fall in West Germany in the 1960s, did not reduce learning outcomes, earnings or employment of affected students. However, little is known regarding the consequences for labour market outcomes of the recent reform reducing the length of the academic track from nine to eight years.

My paper contributes to the literature on how compressing instructional time at school affects labour market outcomes in several ways. Firstly, my identification strategy leverages the staggered implementation of the schooling reform across states for identification, whereas previous studies relied on never-treated units and a single date of implementation. Secondly, using administrative data from social security records, I can not only measure how the reform affects individual-level outcomes but also firm characteristics at which post-reform graduates work. Thirdly, I link my results on labour market outcomes to findings on educational outcomes by describing a mechanism of lower human capital accumulation for post-reform graduates. Finally, my results are robust to treatment effect heterogeneity as I use estimators recently proposed to circumvent shortcomings in difference-in-difference settings with staggered treatment timing.

I show that the reform negatively affects labour earnings and employment when measured six years after graduation when tertiary education is usually completed. I estimate that labour earnings are EUR 2,840 lower due to the reform, equivalent to 14.4% of the average labour income in the sample. The results are larger in magnitude than the estimates in Munteanu and Zhang (2021) investigating a similar reform of reducing years of schooling but modifying the curriculum, resulting in a stable learning intensity in Canada in 1999. They find that graduating one year
earlier with one year less instruction time reduced earnings by 5% to 10%. In line with Marcus and Zambre (2019), I do not find any effect on the probability of having obtained a college degree six years after graduating high school. At the firm level, I show that post-reform graduates not only work for firms that pay lower wages but also have less well-paid positions at these firms. The negative effect persists for later cohorts and beyond the six years after high school graduation for a given cohort.

Previous research has shown that the German "Gymnasium in 8 Years" reform negatively affected final school marks (Büttner and Thomsen, 2015; Huebener and Marcus, 2017), delayed entry to university (Büttner and Thomsen, 2015; Meyer et al., 2019) and led to more irregular study progress at university (Marcus and Zambre, 2019). My findings are consistent with a mechanism that operates through lower human capital accumulation of affected graduates and perhaps partially through signalling on the apprenticeship level. A theoretical model of job search and investment in multidimensional skills (Sanders and Taber, 2012) predicts that the reform's effects should be more negative for outcomes of university graduates than on non-university graduates due to fewer accumulated cognitive skills. My estimates show that the reform is associated with a larger reduction in earnings for university graduates than non-university graduates, supporting a mechanism related to human capital. The reform reduced time spent in school from nine to eight years without changing the scope of the curriculum, which increased weekly workloads and exposed students to more complex content earlier. I show that changes to the curriculum are the primary driver of the large and negative effects, whereas age effects play a minor role. I do not find evidence for teachers allocating more attention to pupils of cohorts graduating after eight years at the expense of pre-reform cohorts.

My results suggest that policymakers face a trade-off between earlier entry into the labour market and lower earnings due to less human capital accumulation. The reform likely increases cash flow within the social security system by graduates entering the labour market one year earlier. However, if the lifetime earnings of post-reform
graduates are persistently lower, then the net benefit of this policy change is uncertain.

The rest of the paper is structured as follows. Section 2 summarises key findings of previous literature on compressing schooling. Section 3 describes the German educational system and the policy change from 2001. Section 4 describes the data and explains the construction of my treatment variable. Section 5 presents the analytical approach. Section 6 presents the estimated effects of the reform on labour market outcomes and robustness checks. Section 7 discusses potential mechanisms before section 8 concludes.

1.2 Effects of Compressing School

My paper is closely related to the literature examining the effect of the "Gymnasium in 8 Years" ("G8") reform on pupils’ post-schooling outcomes. Overall, the literature finds that the reform negatively affects students’ academic performance. Büttner and Thomsen (2015) investigate the link between learning intensity and educational outcomes and treat the reform as a natural experiment. They gather primary data by surveying the class of 2007 at 12 generic secondary schools representative of other public schools in Saxony-Anhalt. Using variation in learning intensity induced by the reform across cohorts, they find a significant negative effect on standardised scores for mathematics but not for German. Final exam grades in mathematics decreased between 0.62 and 0.86 points (pre-reform average: 7.8, from 0 being worst to 15 being best). University enrolment rates for women are significantly lower for the years immediately after the introduction. The effect of enrolment is transitory, implying a delay of university entry rather than a reduction. The reform also changed the choices made after graduating from secondary school. Meyer and Thomsen (2016) find that men are less likely to enrol in educational and social sciences at university. Women are less likely to study natural sciences, mathematics and STEM subjects. My paper contributes to the literature by providing results on how compressing instructional time at school affects labour market outcomes. I link my results on labour market
outcomes to findings on educational outcomes by describing the mechanism of lower human capital accumulation of G8 graduates.

Similar to my paper, some studies leverage the staggered implementation of the G8 reform across states. Huebener and Marcus (2017) find a negative effect on overall final marks but no effect on graduation rates, although grade repetition rates among G8 cohorts are slightly higher. Meyer et al. (2019) use nationwide data on high school graduates and find that besides delaying entry into post-secondary education, students are more likely to participate in other activities following graduation, such as taking a break after high school, volunteering or going abroad. They point out that universities might also mechanically lack sufficient capacities to teach a double cohort of high school graduates, leading to some delayed enrolment. However, as admission cut-off grades do not tighten, they conclude that limited capacities are unlikely to explain delayed enrolment. Marcus and Zambre (2019) show that the study progress of cohorts who finished the academic track in eight years is lower due to drop-out, change in major and formal interruptions at university. Before the reform, over 80% of students progressed regularly. After the reform, they estimate that the share of irregularly progressing students slightly increased by 2.6 percentage points. While these papers exploit the staggered implementation of the reform for identification, my results are robust to treatment effect heterogeneity as I use estimators recently proposed to circumvent shortcomings in difference-in-difference settings with staggered treatment timing.

This paper also relates to the small literature on compressing education on labour market outcomes. Pischke (2007) investigates a reform in the West German school system where students living in some states were subject to two short school years between 1966 and 1967 to shift the beginning of the school year from spring to fall. Using states that opted for one long school year between 1966 and 1967 as a control group, he leverages variation across school entry cohorts, federal states
and secondary school tracks to identify the effect of curriculum compression due to the short school year on various outcomes. Surprisingly, the short school years did not reduce the earnings or employment of affected students. The author concludes that although the overall length of schooling was slightly reduced without altering the curriculum or the highest grade completed, human capital accumulation was not much affected.

In 1999, Canada introduced a reform that shortened the time spent in high school from five to four years and reduced the material taught in schools. Studying the double cohort, Krashinsky (2009) finds that students who graduated one year earlier from high school received ten percent lower wages than earlier cohorts one year after graduation. Munteanu and Zhang (2021) estimate a five to ten percent wage penalty for graduates of the new system conditional on education measured at age 22 to 24. The Canadian reform also reduced students’ academic performance at university, though estimated effects are small and vary by major (Morin, 2013; Krashinsky, 2014). In contrast to the Canadian reform, the German reform kept the content of the curriculum constant. Therefore, I investigate how compressing educational material into fewer years affects labour market outcomes.

1.3 Institutional Background

1.3.1 Tracking in the German School System

In Germany, schooling systems can vary across states. Nevertheless, pupils progress through similar stages during their schooling career. Figure 1.1 illustrates these stages.

The first stage is early childhood education. Children between the ages of three and six attend kindergarten. Enrolment rates are high, especially for children aged five and six, although attendance is not mandatory (Bauernschuster and Schlotter, 2015). Kindergarten provides early socialisation and allows children to prepare for
primary school in a play-based manner. Most children enter primary school in the year they turn six. Depending on the exact day and month of birth, some children may enter primary school at the age of five or seven. Primary schools teach a comprehensive curriculum, including German and English as a foreign language, natural science, social studies, and mathematics. Pupils attend primary school for four years. By the end of fourth year, the transcript of marks includes a recommendation for which secondary school track to attend next.

There are three main secondary school tracks pupils can attend based on academic performance. One potential track is Hauptschule which offers a curriculum focusing on acquiring vocational skills and preparing students for vocational school or entry-level jobs. Pupils attend Hauptschule for five years and finish with a secondary degree. Some states offer stronger students to complete an additional 10th grade for further qualification. This school track is not part of the school system in states formerly part of the German Democratic Republic. Some states have integrated this school track into German comprehensive schools, Gesamtschule, in recent years.

Realschule teaches a broader curriculum compared to Hauptschule. Pupils study theoretical and practical subjects to prepare for either vocational training or higher education. Pupils graduate from Realschule with an upper secondary degree after six years. Strong students can apply to continue on the academic track to attain the university entrance qualification.

Gymnasium is the academic track. The curriculum aims to prepare students to study at university level. It provides a range of subjects that include advanced courses in science, foreign languages and social sciences. Graduates of this track go on to vocational school or enrol at university. Pupils spent nine years on this track before graduating with the university entrance qualification, Abitur. Exceptions are states formerly part of the German Democratic Republic, where pupils graduate after eight years. During the last three years, pupils prepare to take final exams. Graduates of the Realschule who wish to continue school can enter the academic
track at this stage. The final GPA of the university entrance qualification depends on grades attained during the last two years and the results in final exams.

Besides Hauptschule, Realschule and Gymnasium, German comprehensive schools, Gesamtschule, also confer any degree awarded at these other school types. German comprehensive schools exist besides the tracking system and are always part of the recommendations given by primary schools. Some of these comprehensive schools differentiate pupils by performance, sorting them into more or less advanced classes, and some teach all students together. Pupils spend six years at Gesamtschule and can continue at the same school to prepare for final exams to attain the university entrance qualification. Besides these four main school types, there are several hybrid schools and alternative schools that account for a small fraction of school enrolments.

1.3.2 G8 Reform in the German Schooling System

In 2001, the Kultusministerkonferenz (hereafter: KMK), the assembly of ministers of education of the German states, agreed to reduce the years spent on the academic track from nine to eight. Commonly, the reform is referred to as the Gymnasium in 8 Years or "G8" reform, as opposed to Gymnasium in 9 Years or "G9". Compared to their peers internationally, German school graduates were substantially older, so the rationale for the reform was to reduce the age of high-skilled school graduates (Huebener and Marcus, 2015). Policymakers wanted to increase the pool of skilled workers available to the labour market, counteracting the adverse effects of demographic change on the German pay-as-you-go social security systems.

The reform was designed not to compromise the quality of education. Policymakers kept the scope of the curriculum the same, thus compressing the Gymnasium curriculum by one year rather than reducing a pupil's exposure to learning content by one year. The dashed rectangle illustrates this in figure 1.1. Pupils spend eight years on the academic track after the reform instead of nine years. Notably, the curriculum
compression affected only the first six years of Gymnasium. The curriculum of the last three years, when pupils prepare to take final exams, remained unaffected by the reform. Thus, the reform compressed the curriculum of the first six years into five years.

The KMK implemented the curriculum compression in two ways. First, the number of minimum contact hours required before graduating from Gymnasium did not change. The average number of contact hours on the academic track summed until attaining the university entrance qualification remained unchanged at 265.\(^1\) A pupil graduating with the university entrance certificate before the reform had, on average, 29.4 weekly lesson hours over nine school years. After the reform, pupils had, on average, 33.1 weekly lesson hours over eight years (Homuth, 2017).\(^2\) By keeping the average number of contact hours constant, the G8 reform did not impact the overall teaching workload for schools. However, during the transition period from G9 to G8, the increase in weekly instruction hours meant an increased demand for teaching hours due to G8 and G9 cohorts attending the academic track simultaneously. The increase in weekly instructions for pupils had only a limited effect on teachers’ working hours. Using the microcensus for the years 2002 until 2012, Huebener et al. (2017) show that during the transition period, reported working hours for teachers increased by 0.23 hours per week for each new G8 cohort starting on the academic track. Additionally, their estimates indicate that the share of younger and older teachers increased suggesting that older teachers stayed in service for

---

\(^1\)Out of these 265 hours, 260 hours need to be attributed to certain subjects, while 5 hours can be general studies (e.g. project work, focus hours). Many schools supplied more contact hours than the minimum requirement, but less so after the reform (Kühn et al., 2013).

\(^2\)This model is known as the 5+3 model. Pupils attend the academic track for five (instead of six) years before being admitted to the last three years to work towards the university entrance certificate. Most states (Bavaria, Bremen, Hesse, Lower Saxony, North Rhine-Westphalia, the Saarland, Schleswig-Holstein and Thuringia) implemented the reform this way (Homuth, 2017). Some states (Baden-Württemberg, Hamburg, Mecklenburg-West Pomerania, Saxony and Saxony-Anhalt) implemented the reform in a 6+2 model. Pupils attend the academic track for six years as usual before being admitted to the last two years to work towards the university entrance certificate. In these states, grade 10 serves a double function as the last year in which grades do not count towards the final GPA and as an introductory phase for students from Realschule. In Brandenburg and Berlin, primary school finishes after six years. The reform was implemented as a 3+3 model in Brandenburg, while Berlin chose a 4+2 model.
longer and more younger teachers were hired. Overall, they conclude that the effect of the G8 reform on the teacher body was limited.

Second, the education ministries of the federal states accommodated the compression of instructional time by redesigning the curriculum for starting cohorts affected by the reform. While the teaching content in each subject remained unchanged, the redesigned curriculum altered the timing of when pupils are supposed to learn certain subject matters. Figure 1.2 provides an example of the mathematics core curriculum from the largest state by population, North Rhine-Westphalia (Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen, 2004; Ministerium für Schule, und Weiterbildung des Landes Nordrhein-Westfalen, 2007). Panel (a) of figure 1.2 shows an overview of the core curriculum for mathematics as set by the state before the reform. Panel (b) shows the overview of the core curriculum for mathematics as set by the state after the reform. Figure 1A.1 in the appendix shows the same core curriculum translated from German to English. The evaluation categories in the column headers remained the same before and after the reform. Before the reform, the highest grade in the overview was grade 10, while the highest grade in the outline was grade 9 after the reform. Overall, the types of competencies stayed the same within each category. However, pupils were expected to have acquired these competencies at different grades before and after the reform. In the Reasoning/Communicating column, the skill to extract information from authentic texts, such as newspapers, was a grade 9/10 competency for pre-reform cohorts. For post-reform cohorts, this is a grade 7/8 competency. Other skill blocks, such as presenting, verifying and evaluating problem-solving approaches, were divided into multiple competencies in the post-reform curriculum. Sometimes, the wording changed to draw attention to specific competencies explicitly, for example, in the Arithmetic/Algebra category. The pre-reform overview mentioned natural numbers, while the post-reform overview refers to integers. Although integers were not explicitly mentioned in the pre-reform overview, working knowledge of integers is required to comprehend techniques taught in later grades.
In summary, the 2001 schooling reform reduced the years spent on the academic track from nine to eight, commonly referred to as the G8 reform. The reform changed pupils’ education experience in three ways. Firstly, pupils graduate with the university entrance certificate one year earlier, reducing their age at test taking and when deciding on post-secondary education. Secondly, as the cumulative minimum required contact hours remained unchanged, the reform increased pupils’ weekly workload. Thirdly, the redesigned curriculum exposed students to complex content at earlier grades.

1.4 Data Sources

1.4.1 Sample of Integrated Labour Market Biographies - SIAB

I use the Sample of Integrated Labour Market Biographies (hereafter: SIAB), which consists of a 2% random sample based on the Integrated Employment Biographies (IEB) of the Institute for Employment Research (IAB) at the German Federal Employment Agency (Bundesagentur für Arbeit) in Nuremberg (Frodermann et al., 2021). It provides a worker’s employment status and precise individual-level data on wages and other characteristics. Included in the data are individuals who are employed (i.e. subject to social security, marginal part-time employment), receive benefits in accordance with the German Social Code III or II, are registered as job-seeking or planning on participating in programs of active labour market policies. Information on civil servants, self-employed, and individuals out of the labour force is not included.

A feature of the IEB is that individuals are only included if they have been recorded as employed or registered unemployed\(^3\) at least once. I impute missing spells for all individuals between their first appearance in the SIAB and 2019, the last available

\(^3\)For readability, I use the terms "employed or registered unemployed" to refer to any labour force status that would result in a registered spell in the IEB. These include individuals who are employed (i.e. subject to social security, marginal part-time employment), receive benefits in accordance with the German Social Code III or II, are registered as job-seeking or planning on participating in programs of active labour market policies.
calendar year. For example, if an individual is recorded as marginally employed at a young age (e.g. helping out in a shop on the weekend), then they would be included in my sample, even if they are not employed or registered unemployed when I measure their labour market outcomes for a fixed number of years after graduation. However, if an individual has never been employed before measuring their outcomes, then they would not appear in my sample. Therefore, my estimates should be interpreted as conditional on appearing in my sample based on the SIAB. Using the SOEP, I find that 90% of graduates from the academic track were either employed or registered unemployed at least once before the age of 24. Reassuringly, my estimates based on the SOEP also suggest that the G8 reform did not affect this share.

**Construction of Cohort Indicator**

While the SIAB includes information on the highest school leaving degree, it does not contain detailed information about individuals’ schooling background, such as the years of schooling or the state where a school was attended. Therefore, I construct a treatment indicator for whether each individual was part of a G8 or G9 cohort using the state of residence when the first employment spell is recorded, the type of high school leaving certificate and the month and year of birth. The procedure follows three steps.

Firstly, I define the state where an individual attended school as the state where the individual lived at the time the first employment spell is observed in the SIAB. Munteanu and Zhang (2021) show that the observed location of the first job and the location of attaining a high school certificate are highly correlated in Canada. I use the German Socio-Economic Panel (Goebel et al., 2019; hereafter: SOEP) to validate this correlation in the German context. In the SOEP, individuals are asked about where they live and where they earned their school leaving degree. I find a

\[4\]

My baseline estimate for the effect of the G8 reform on labour market outcomes is EUR 2,844.41. Suppose the effect of the G8 reform on all individuals who were not employed or registered unemployed at least once before the age of 24 was 0. Then, I would overestimate the effect by 11%, or EUR 284.41 in yearly earnings.
correlation coefficient between the state of residency at first employment and the state of receiving the high school leaving certificate of 0.96, similar to Munteanu and Zhang (2021).

Secondly, I assign each individual to an entry cohort. Children enter primary school in the year they turn six if their birthday is before a state-specific cut-off date. Before 1997, the cut-off was June 30th in all states. Since 1997, each state has set the cut-off date individually. Table 1A.1 in the appendix provides an overview of the changes in cut-off dates by state and year used to infer the year of school start. I use the exact month and year of birth and the state of residence when the first employment spell is recorded to impute the year of school start. I validate this step by comparing the imputed and the reported year of school start using the SOEP. Figure 1A.2 in the appendix shows that my imputation procedure can reliably predict the year of school start.

Thirdly, I assign each individual to have completed high school under the G8 or G9 system based on the observed state of residency at first appearance and the imputed year of school entry.

1.4.2 Sample Selection and Descriptive Statistics

For my main sample, I only keep individuals who graduated high school with a university entrance certificate, as observed in the SIAB. A caveat of the data is that I cannot distinguish between individuals who earned their university entrance certificate after attending the academic track for the entire length, entered after graduating from Realschule or received their certificate from a comprehensive school. I exclude all individuals who received their university entrance certificate in Hesse, Rhineland-Palatinate, Saxony and Thuringia. Hesse introduced the reform gradually over three years, so I cannot cleanly assign individuals to either a G8 or G9 cohort. Rhineland-Palatinate introduced G8 only at 19 Gymnasiums, so some pupils have graduated after either eight or nine years in a given year. Therefore, I cannot distinguish between G8 and G9 graduates from this state. After the reunification,
Saxony and Thuringia never adopted the standard of the former West German states of a nine-year Gymnasium. Thus, I assign all individuals in these states as having graduated under the G8 system. Since these states are always-treated, I drop individuals who graduated in these states from the sample. After aggregating the data to the federal state-school entry cohort level, I trim the data to include entry cohorts between 1989 and 2000.

Table 1.1 shows summary statistics for the primary labour market outcomes of interest measured six years after high school graduation. All monetary values are denoted in 2015 euro. I impute missing spells for each individual between their first entry to the labour market as recorded in the SIAB and the last observed calendar year, 2019. Earnings and daily wage are recorded as zero for these missing spells, and the employed dummy is set to zero. Throughout the analysis, earnings data is reported in levels rather than logs to keep individuals who are not employed included in the estimation sample. Similarly, I impute the experience measure using the value of the last observed employment spell. Tenure, the mean wage at a firm and the relative distance to the firm’s mean wage are only recorded for employed individuals and set to missing otherwise. The relative distance to the firm’s mean wage is measured as the difference between the mean wage at the firm and an individual’s daily wage divided by the mean wage at the firm. Additionally, the mean wage at a firm and the relative distance to the firm’s mean wage exclude firms with less than ten employees. This restriction prevents the mean wage at the firm from being mechanically lowered if the wage of one individual is reduced due to the reform. The relative distance to the firm’s mean wage excludes individuals who are recorded as employed but have zero earnings in a given year (e.g. marginally employed workers). The sample includes a total of 71,637 individual-level observations, with some variables only recorded for employed individuals, such that the number of observations varies with the outcome.
1.5 Analytical Approach

1.5.1 Staggered Introduction of the G8 Reform

Education is subject to state legislation. Figure 1.3 illustrates the staggered introduction of the G8 reform across states in different years.

By 2016, all states in my sample had at least one cohort graduating after eight years at Gymnasium. Figure 1.4 shows the share of G9 and G8 pupils on the academic track across all grades and states in my sample. The first G8 cohort entered 5th grade of the academic track in Saxony-Anhalt in 2001. As more states adopted the reform and more G8 cohorts entered the academic track in early-adopter states, the share of G9 pupils decreased until all pupils on the academic track were G8 after 2016. The first G8 entry cohort graduated simultaneously with the last G9 entry cohort in each state. I will refer to these two entry cohorts as the double graduation cohort.

1.5.2 Treatment Effects of Interest

Let \( s = 1, 2, ..., S \) denote federal states and \( c = 1, 2, ..., \Gamma \) school entry cohorts. Define \( D_{s,c} \) as a binary dummy variable equalling one if school entry cohort \( c \) in state \( s \) attended school under the G8 system and zero otherwise. Let \( \theta_{s,c}(1) \) and \( \theta_{s,c}(0) \) denote the potential outcome of interest in state \( s \) for school entry cohort \( c \) having attended school under the G8 system (with treatment) and not having attended school under the G8 system (without treatment), respectively.

Each state only has one school system in place at the time each cohort starts school. Thus, the observed outcome for cohort \( c \) in state \( s \) is

\[
\theta_{s,c} = D_{s,c}\theta_{s,c}(1) + (1 - D_{s,c})\theta_{s,c}(0). \tag{1.1}
\]

I define the causal effect of the G8 reform for cohort \( c \) in state \( s \) as the difference between the observed outcome under the G8 system (with treatment) and the
unobserved potential outcome for cohort $c$ in state $s$ under the G9 system (without treatment):

$$\alpha_{s,c} \equiv \theta_{s,c}(1) - \theta_{s,c}(0). \quad (1.2)$$

Following Marcus and Sant’Anna (2021) and Callaway and Sant’Anna (2021), I express these individual causal treatment effects as group-time average treatment effects, $ATT(g, c)$. Define $G_g$ as a dummy variable equalling one if a state introduced the G8 system for cohorts starting in year $g$. Hereafter, I drop the state index $s$ to keep the notation concise. Then, the average treatment effect of the G8 reform on school entry cohort $c$ across all states first introducing the G8 reform for school entry cohort $g$ can be written as:

$$ATT(g, c) \equiv E[\theta_c(1)|G_g = 1] - E[\theta_c(0)|G_g = 1]. \quad (1.3)$$

As each state can only be observed having one schooling system in place, I need to make the following parallel trend assumption that for all $g, t, c = 2, \ldots, \Gamma$ such that $c \geq g, t \geq c$,

$$E[\theta_c(0) - \theta_{c-1}(0)|G_g = 1] = E[\theta_c(0) - \theta_{c-1}(0)|D_t = 0]. \quad (1.4)$$

The assumption imposes that, for treated states in the absence of treatment, the expectation of the outcome of interest would have followed the same path as for all not-yet-treated states by entry cohort $c$. By invoking this parallel trend assumption, I restrict the observed pre-treatment trend across groups to be the same.$^5$

Although this type of parallel trend assumption based on not-yet-treated groups is stronger than a parallel trend assumption based on a never-treated group (Marcus and Sant’Anna, 2021), there is no state that did not introduce the G8 system at

---

$^5$This parallel trend assumption is similar to PTA 7 in Marcus and Sant’Anna (2021). They show that by using not-yet-treated units as a control group, equation 1.4 does not restrict pre-trends for periods before the first unit is treated or pre-trends for the earliest treatment group. The assumption does restrict observed pre-treatment trends across groups.
all. Rhineland-Palatinate implemented the reform as a pilot project in some schools. However, I cannot distinguish between individuals who were part of the project and those who were not. If Rhineland-Palatinate was used as a never-treated unit, some treatment effects would be contained in the observed outcomes used for calculating counterfactuals.

Imposing equation 1.4 and when pre-treatment covariates play no part in identification, Callaway and Sant’Anna (2021) show that, for $c \geq g$, $ATT(g, c)$ is non-parametrically identified by

$$ATT_{ny}(g, c) = E[\theta_c - \theta_{g-1}|G_g = 1] - E[\theta_c - \theta_{g-1}|D_c = 0].$$  

(1.5)

One concern Callaway and Sant’Anna (2021) point out is that early-treated groups are exposed to different economic environments than those treated later in the sample. As my sample spans 12 years, potential non-parallel pre-trends due to different exposure to shocks should be considered when interpreting my results.

### 1.5.3 Difference-in-Differences in Staggered Reforms: Two-way Fixed Effect Regression and Robust Estimators

Empirical research has traditionally employed two-way fixed effect regression models to estimate an overall average treatment effect on the treated. Leveraging variation in the timing of the reform to identify the causal effect of the G8 reform on states that adopted the G8 system, I can estimate a static two-way fixed effect (hereafter: TWFE) regression model of the form:

$$\theta_{s,c} = \kappa_s + \tau_c + \beta_{TWFE}D_{s,c} + u_{s,c}. \quad (1.6)$$

$D_{s,c}$ indicates treatment status for each school entry cohort $c$ in state $s$. $\theta_{s,c}$ is the outcome of interest. $\kappa_s$ and $\tau_c$ are state and school entry cohort-fixed effects, respectively. $u_{s,c}$ is an idiosyncratic error term.

In a specification like this, $\beta_{TWFE}$ is the parameter of interest. Goodman-Bacon
(2021) shows in his decomposition analysis that $\beta_{TWFE}$ is a variance-weighted combination of each group $g$’s respective canonical difference-in-differences estimator, where the weights depend on panel length and the size of each group $g$. Recent literature shows that $\beta_{TWFE}$ only recovers a sensible weighted average of treatment effects under treatment effect homogeneity across time and across groups (Chaisemartin and d’Haultfœuille, 2020; Goodman-Bacon, 2021; Borusyak et al., 2023).

However, as treatment effects are potentially heterogeneous across time, I follow the empirical literature and estimate a dynamic TWFE specification of the form:

$$\theta_{s,c} = \kappa_s + \tau_c + \sum_{e=-5}^{5} \beta_e 1[e = c - g_s] + v_{s,c}. \tag{1.7}$$

In this dynamic TWFE model, the parameters of interest are coefficients $\beta_{-5}$ to $\beta_5$ associated with dummies for $e$ time periods relative to treatment. $g_s$ is the first cohort in state $s$ that graduated under the G8 system and $v_{s,c}$ is an idiosyncratic error term.

Similarly to the criticism of the static TWFE model, the parameters of interest in this dynamic specification, $\beta_e$, might suffer from contamination of treatment effects from other periods if treatment effects are heterogeneous across treatment groups $g$ (Chaisemartin and d’Haultfœuille, 2020; Sun and Abraham, 2021).

As both the static and the dynamic TWFE specification require strong assumptions regarding treatment effect heterogeneity to recover a sensible weighted average of treatment effects, I also present estimates based on robust estimators as proposed by Chaisemartin and d’Haultfœuille (2020), Callaway and Sant’Anna (2021), Sun and Abraham (2021) and Borusyak et al. (2023).
1.6 Results

1.6.1 Baseline Results

Baseline Estimates

Table 1.2 reports estimation results of $\beta_{TWFE}$ in equation 1.6 on yearly labour earnings in column 1 and the probability of employment in column 2 measured six years after high school graduation. All monetary values are denoted in 2015 euro. The reform has reduced yearly labour earnings by EUR 2,844. The estimated effect is statistically significant on all conventional levels. In context, this is a reduction of more than 14% of the average wage of those employed in my sample. Similarly, the G8 reform reduced the probability of being employed six years after high school graduation by 2.5 percentage points. The effect is precisely estimated and statistically significant.

Table 1.3 presents effect estimates of the G8 reform on individual-level and firm-level outcomes. In all specifications, the point estimates are negative. Columns 1 and 2 show that the reform did not affect tenure at the firm or experience. This result is expected as I fix the time since leaving high school by measuring outcomes six years after graduation. The estimate in column 3 suggests that the reform does not affect the probability of having attained a college degree six years after graduating high school. This result aligns with findings by Büttner and Thomsen (2015) and Meyer et al. (2019), who estimate a temporary negative effect on university enrolment for G8 graduates, suggesting a delay of university entry rather than a reduction. The reform reduced the daily wage of graduates by EUR 7.72, which is precisely estimated. This coefficient mirrors the estimated effect on yearly labour earnings and suggests that the reduction in earnings stems from lower wages rather than fewer days worked.

Columns 5 and 6 report results for firm-level outcomes as the dependent variable. Column 5 shows that G8 graduates work for firms that pay a EUR 6.01 lower average daily wage. The estimate is statistically significant on all conventional levels. Indicative of an individual’s income rank within a firm relative to its coworkers, column 6 shows
the estimated effect on the relative distance between a graduate’s daily wage and the average daily wage at the firm, divided by the average daily wage at the firm. The estimate suggests that the relative distance to the firm’s mean wage is about 2.5 percentage points more negative for G8 graduates. These results imply that G8 graduates not only work for firms that pay lower wages but also have less well-paid positions at these firms. When interpreting my results, conditioning on individuals having graduated from high school with a university entrance certificate requires my coefficients to have a conditional-on-positives meaning.

**Event-study Estimates**

Treatment effects of the G8 reform are likely heterogeneous across cohorts. Therefore, I estimate the dynamic version of the TWFE specification as specified in equation 1.7. Although the parallel trend assumption can never be tested for post-treatment periods, event-study plots allow me to test whether parallel trends hold in pre-treatment periods. However, regression models, such as equation 1.7, can suffer from bias when treatment effects are heterogeneous across time and treatment groups (Sun and Abraham, 2021). Thus, I not only show results from an even-study version of the TWFE specification estimated using OLS but also estimators that are robust against treatment effect heterogeneity (Chaisemartin and d’Haultfoeuille, 2020; Sun and Abraham, 2021; Callaway and Sant’Anna, 2021; Borusyak et al., 2023). The overarching theme of these estimators is to prevent the use of early-treated units as a control group for later-treated units which can lead to bias under treatment effect heterogeneity. Callaway and Sant’Anna (2021) propose an estimator that calculates narrow group-time average treatment effects of the treated based on weighted differences in means. Group-time average treatment effects of the treated can be aggregated to reflect different parameters of interest, such as treatment effects over time for each timing group or the overall treatment effect of the treated. Sun and Abraham (2021) present an estimator similar to Callaway and Sant’Anna (2021) but use only the last treatment group as the control group, rather than not-yet-treated
groups. A key feature of the estimator proposed by Chaisemartin and d'Haultfoeuille (2020) is that it can accommodate treatment switching on and off. As the G8 reform was introduced in a staggered manner, treatment does not switch off for any units in the sample.\textsuperscript{6} Their main estimator is a weighted average of comparisons between the evolution of the mean outcome over two periods for units whose treatment switches on and those units remaining untreated. Borusyak et al. (2023) derive an imputation-based estimator which predicts the expected potential outcome for treated units via OLS using only untreated observations before calculating the difference between observed and imputed outcomes. Parameters of interest are calculated as weighted averages of these differences. The dummy for the school entry cohort one year before the double graduation cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort.

Figure 1.5 presents event-study results for yearly labour earnings measured six years after high school graduation. Vertical lines represent 95 percent confidence intervals. The point estimates for cohorts that finished school under the G9 system are close to zero and not statistically significant for virtually all estimators. This indicates that the parallel trend assumption holds for pre-treatment cohorts. The absence of a pre-trend also corroborates my treatment assignment procedure. Falsely assigning individuals who graduated under the G8 system to pre-treatment cohorts would result in pre- and post-treatment coefficients being biased. The absence of a significant difference between pre-treatment coefficients close and far to the double graduation cohort provides evidence that my procedure correctly assigns individuals to pre-treatment cohorts.

Labour earnings are substantially lower for cohorts that graduated under the G8 system. For the first G8 cohort, the estimated negative effect on yearly earnings is between EUR 2,300 and EUR 3,350, depending on the estimator used. This coefficient likely provides the most transparent effect estimate of the G8 reform. As some states have recently reverted back to a G9 system, this estimator might be a sensible choice for future research investigating switching from G9 to G8 to G9.  

\textsuperscript{6}As some states have recently reverted back to a G9 system, this estimator might be a sensible choice for future research investigating switching from G9 to G8 to G9.
the last G9 cohort and the first G8 cohort graduated in the same state in the same calendar year, both cohorts face similar conditions when applying for university or apprenticeships and in the labour market later on. The grey-shaded area in the event-study plots in figure 1.5 illustrates this. While the last G9 cohort is unaffected by the reform, the estimated effect for the first G8 cohort is similar to other G8 cohorts, though slightly smaller.\textsuperscript{7,8} Assuming a simple model of labour supply and demand, the double graduation cohort would represent a right shift in the labour supply curve, leading to lower wages for all graduates of the double graduation cohort in competitive labour markets. However, the event-study plots in figure 1.5 show that the last G9 cohort is unaffected by the reform. What does that imply for the labour market and potential general equilibrium effects? Suppose that firms can distinguish G8 and G9 graduates.\textsuperscript{9} If firms offered different jobs to G8 and G9 graduates without affecting each others’ productivity in each job, then firms might be able to offer different wages to G8 and G9 graduates. Based on the results in table 1.3, G8 graduates appear to have less well-paid positions, suggesting that firms hire G8 and G9 graduates for different jobs. As the effect on the probability of being employed six years after graduation is small, especially for the double graduation cohort, firms also seem to be able to absorb the number of graduates from the double graduation cohort.

My results only allow for speculation about the reform’s consequences on product markets. A supply shock of this kind could lead to higher production if firms are able to scale up their production to employ additional workers rather than replacing retired workers or hiring in anticipation of an ageing workforce. If firms scale up and

\textsuperscript{7}The estimate on the first G8 cohort likely presents a lower bound for the absolute effect of the G8 reform on labour earnings for this cohort. Falsely assigning individuals who graduated under the G9 system to G8 cohorts would result in flat pre-trends and a coefficient for the first treatment cohort that is biased towards zero.

\textsuperscript{8}I cannot distinguish between individuals who received a university entrance certificate after attending the academic track for the entire length, entered after graduating from Realschule or earned their certificate at a comprehensive school. Therefore, I likely incorrectly classify some individuals as having graduated under the G8 system based on their state of residency at their first employment spell and imputed school entry. These measurement errors due to incorrect classifications lead to an attenuation bias in my estimates. Thus, my estimates should be interpreted as a lower bound.

\textsuperscript{9}I explore this idea further below and establish that firms could only recognise a G8 or G9 graduate based on accompanying application documents, such as CVs or motivation letters, rather than the degree certificate. Whether this is a strong assumption thus depends on the understanding of a firm’s hiring process.
increase output then prices in the output market will be lower due to the reform. The magnitude of such effects would depend on the size of the supply shock. The number of graduates of the double graduation cohort in any given state only represents a small share of workers. Over the sample period when labour market outcomes are measured, the number of employed individuals in Germany was around 40 million. In contrast, the average cohort size of graduates on the academic track in any state was between 60,000 and 80,000, representing 2.4% to 3.2% of the total labour force in an average state. Additionally, as the reform was implemented in response to an ageing population, firms have either been facing labour shortages already or are expecting shortages in the future. Therefore, the effect of a double graduation cohort, given the size of the total pool of workers, is likely small. Thus, as marginal costs for firms were only minimally affected, supply in product markets likely remained constant. The small relative size of the pool of graduates compared to the overall demand is likely affecting prices minimally in equilibrium. However, in case there are general equilibrium effects such that labour market outcomes of G9 cohorts are negatively affected by a supply shock to the overall labour market, then my estimates would be biased towards zero. This downward bias of the absolute effect size arises from cohorts being used to infer counterfactual outcomes, although they are negatively affected by the reform. In this circumstance, my results can be seen as a lower bound of the absolute effect size.

The event-study plots indicate that treatment effects are also slightly dynamic. The negative effect of the G8 reform is slightly increasing for cohorts that have graduated more recently. For the fifth G8 cohort, estimates range between EUR 3,900 and just over EUR 6,000. The estimated coefficients are similar across the used estimators, including TWFE-OLS results.

Figure 1.6 displays event-study plots for the effect of the G8 reform on additional labour market outcomes of interest. Regardless of the outcome, the coefficients for G9 cohorts, including the G9 double graduation cohort, are close to zero and statistically insignificant. The dynamics of the effect on employment are more pronounced
compared to yearly labour earnings. While the first two G8 cohorts are between 1 and 2 percentage points less likely to be employed, the estimated effect for the fifth G8 cohort is more than five percentage points. Experience and tenure exhibit no systematic dynamics. Point estimates of the effect for more recent cohorts indicate higher experience and tenure, though the coefficients are imprecisely measured. The estimated effects on the probability of having obtained a college degree do not exhibit any dynamics and are statistically insignificant for nearly all estimators. This result is in line with previous research. Once students are admitted to university, Marcus and Zambre (2019) show that G8 graduates are slightly more likely to progress irregularly through their studies. However, the estimated effect is only about 2.6 percentage points, which is small compared to over 80% of students progressing regularly. Therefore, a delay in graduation from university is unlikely to drive my results on labour earnings. The event-study plot for daily wage follows the dynamics in figure 1.5. The finding that G8 graduates work for firms that pay, on average, less is mainly driven by the first two G8 cohorts. Graduates of more recent cohorts also tend to work for firms that pay lower daily wages on average, but the results are overall imprecisely measured. The negative effect on the relative distance between the average and individual daily wage at the firm is strongest for the second G8 cohort, and the effect appears to fade out for more recent cohorts.

1.6.2 Robustness Checks

Unaffected Tracks

While the G8 reform changed the education experience of pupils attending the academic track, pupils on other tracks did not experience such a change. Thus, I can test the robustness of my main results by repeating the analysis on a sample of individuals who did not attend the academic track. For this placebo falsification, I select a sample of individuals with a school leaving degree no higher than a secondary
degree (*Hauptschule*) or no degree. According to the SIAB, these graduates have a very different trajectory than individuals with a university entrance certificate. Therefore, these two groups are unlikely to be substitutes in the labour market.

Keeping all treatment indicators unchanged and using the same difference-in-differences design as in equation 1.6, table 1A.2 in the appendix presents TWFE-OLS results from this placebo exercise for outcomes measured six years after high school graduation. The event time coefficients of post-treatment dummies are statistically insignificant on all conventional levels for all relevant outcomes, both individual- and firm-level. As such, these estimates provide evidence against the hypothesis of cyclical or other unobservable factors driving my main results.

**Timing of the G8 Reform**

Although figure 1.5 suggests that the parallel trend assumption holds for pre-treatment periods, parallel trends for post-treatment periods might be violated if both the timing of the reform and the outcome variable were correlated with other factors. States commonly introduced the reform when pupils entered 5th grade at *Gymnasium*. However, I measure the outcome variables of interest many years after the introduction of the reform. Nevertheless, a potential confounding factor could be that policymakers wait for an academically strong cohort to implement the reform. In the appendix, figure 1A.3 probes this hypothesis by plotting standardised test scores after partialling-out state effects for reading, maths and science. Reassuringly, none of the discontinuity coefficients is statistically significant. Additionally, if policymakers wanted to probe the workings of the reform, they would be more likely to do so with a smaller cohort. Using administrative data on enrolled pupils on the academic track from the German federal statistical office (Statistisches Bundesamt (Destatis), 2023), I measure cohort size in the last year of primary school before pupils are sorted into tracks. As with test scores, the discontinuity coefficient is statistically insignificant and contextually small.

10The education variable in the SIAB does not clearly distinguish between an intermediate secondary degree (*Realschule*) and a secondary degree (*Hauptschule*). The *Realschule* track takes one year longer to complete than the *Hauptschule* track. Therefore, the most transparent placebo falsification test involves individuals with a school-leaving degree no higher than a secondary degree.
I find no documentation about the timing of the reform being in response to outcomes. Marcus and Zambre (2019) conjecture that Saxony-Anhalt and Mecklenburg-Vorpommern were early adopters of the reform because of their familiarity with an eight-year Gymnasium. Similarly, Saarland, the third state implementing the reform, was eager to switch to the G8 system due to its proximity to France, where pupils graduate after eight years. While five states introduced the reform with the cohort starting school in 2000, North Rhine-Westphalia might have held off introducing the reform by one year, being the largest state by population.

**Compositional Changes**

If the reform changed who attends the academic track, then the parallel trend assumption would be violated as I condition on individuals to have a recorded university entrance certificate in the data. I address this concern by plotting different measures of grade progression using the same administrative data on enrolled pupils on the academic track (Statistisches Bundesamt (Destatis), 2023) as before. Panel (a) in figure 1A.4 in the appendix presents year-on-year repetition rates until high school graduation for G8 and G9 cohorts. Repetition rates are slightly higher for G8 cohorts, especially in later years, which is in line with previous research (Huebener and Marcus, 2015; Huebener and Marcus, 2017). The dip in drop-out rates in the third year before high school graduation under both systems is explained by the influx of students having completed the Realschule track and entering the academic track to obtain the university entrance certificate. Panel (b) of figure 1A.4 displays a grade progression index until high school graduation. As cohort sizes are indexed at grade 7, the grade progression of G9 cohorts appears mechanically lower in years further away from graduation when averaged over school grades. The index for G8 cohorts reaches a comparable level to G9 cohorts in years closer to graduation.

Common to both plots is that my measure of grade repetition cannot distinguish between track switchers, repeaters and, although unlikely, school drop-outs. Using a similar identification strategy as my paper, Marcus and Zambre (2019) discuss three
potential reasons for compositional changes that could pose a threat to identification. Firstly, parents or students could strategically move between states to avoid or select into the G8 system. Using the SOEP, I show that the correlation between the state where primary education was completed and the state where secondary education was commenced is 0.986 for pupils on the academic track. This correlation is unaffected when either G8 cohorts only or G9 cohorts only are considered. Secondly, students could avoid the academic track altogether and attend a lower track instead. If students still aim to attain the university entrance certificate, then the only viable option would be the Realschule track as a transfer to the academic track after completing the Hauptschule track is not possible, as shown in figure 1.1. Thirdly, pupils could attend comprehensive school types offering a 13-year path to the university entrance certificate. Another concern could be that pupils on the Realschule track would have continued to complete the last three years of the academic track to attain the university entrance certificate under the G9 system but not under the G8 system. However, as most of the compressing occurs within the first five years on the Gymnasium track, the influence on the decision to continue on the academic is likely small. All these reasons can cause a problem if they lead to fewer graduations from the academic track. However, table 1A.3 in the appendix shows that the effect of the G8 reform on the number of graduates from the academic track is unaffected. Similarly, Huebener et al. (2017) do not find any evidence for compositional changes caused by the reform.

1.6.3 Effects of the G8 Reform in the Longer Run

So far, I presented estimates for the effect of the G8 reform on labour market outcomes measured at a fixed number of years after graduating high school. Some states introduced the G8 reform early enough, so I can trace its effects on labour market outcomes for more years after graduation. For example, Saxony-Anhalt

---

11Grewenig (2021) shows that switches from lower school tracks to higher school tracks are uncommon between grades 5 to 9. Therefore, the margin for strategically starting on a lower track and transferring to a higher track is unlikely to drive my results.
introduced the reform with the cohort entering the schooling system in 1995. My dataset includes observations up to 2019, so I can trace out the causal effect of the G8 reform for up to 11 years after high school graduation for this cohort. At the same time, other school entry cohorts are only observed for a limited number of years due to the recency of the reform. Thus, when estimating the effect of the reform, say ten years after graduation, I can limit my sample only to include entry cohorts between 1989 and 1996.

I study the longer-run effects of the G8 reform by estimating a variation of equation 1.6:

\[
\theta^k_{s,c(k)} = \kappa_s + \tau_c(k) + \beta^k_{c(k)} D_{s,c(k)} + u_{s,c(k)},
\]

where \(\theta^k_{s,c(k)}\) is the labour market outcome of interest measured at \(k\) years after graduating high school. My notation \(c(k)\) captures that the inclusion of entry cohort \(c\) depends on how many years after graduation the outcome of interest is measured. \(D_{s,c(k)}\) indicates treatment status for each included school entry cohort \(c(k)\) in state \(s\). \(\kappa_s\) and \(\tau_c(k)\) are fixed effects for state and included school entry cohort, respectively. \(u_{s,c(k)}\) is an idiosyncratic error term.

Figure 1.7 presents the estimates of \(\beta^k_{c(k)}\) for different subsamples of entry cohorts and shows that the G8 reform significantly lowered labour earnings of G8 graduates even ten years after graduating from high school. The figure also illustrates that the estimated effect becomes more negative the more cohorts are included to estimate the effect \(k\) years after graduation. This suggests that the effect on labour market earnings between three to six years after graduating is increasingly negative for more recent cohorts. The estimates for daily wage as the outcome mirror the effect on yearly labour earnings. Regarding firm-level outcomes, G8 graduates work for firms that pay, on average, EUR 6 to EUR 9 less a day in the longer run. For nearly all

---

12 A G8 cohort entering school in 1995 graduated in 2007, whereas a G9 cohort graduated in 2008. The latest year included in my dataset is 2019. Therefore, I can measure labour earnings up to 11 years after graduating high school while preserving a balanced panel in state and school entry cohort space.
years after graduation, G8 graduates also receive wages that are further away from the firm’s mean wage compared to had they graduated under the G9 system.

These results indicate that the large and negative effects are not short-lived. On the contrary, negative effects on labour earnings are becoming larger over time. Given the size of the estimates, the reform could reduce the lifetime earnings of G8 graduates despite one more year of potential labour market participation.

1.7 How Does the G8 Reform Impact Labour Earnings?

The G8 reform compressed instructional time from nine to eight years, resulting in pupils having an increased weekly workload and being exposed to complex content earlier under the new curriculum. The reform also lowered the average age at high school graduation by one year. Additionally, not only pupils had to adjust to the new curriculum, but also teachers. Recent research has established that the G8 reform reduced the overall GPA and grades in final exams of G8 pupils (Büttner and Thomsen, 2015; Huebener and Marcus, 2017). As lower grades can either be due to noise in the educational signal or an indicator of lower human capital, I discuss potential links between lower grades and adulthood outcomes in the German context. Moreover, I will provide estimates for the importance of three main channels that affected labour market outcomes through the described mechanism: 1) curriculum changes with higher workloads and reformed timing, 2) lower age at test taking, and 3) allocation of teachers’ attention.

1.7.1 Mechanism

Before entering the labour market, high school graduates have the choice of attending university, starting an apprenticeship, taking a break after leaving school, or a combination of those. After finishing high school, graduates of the first and second G8 cohorts are more likely to participate in other activities following graduation, such
as taking a break, volunteering after high school or staying abroad (Meyer et al., 2019). Büttner and Thomsen (2015) show that these activities delay enrolment at university or apprenticeship rather than replacing them. Additionally, in figure 1.6, I show that even for these cohorts, the probability of having obtained a college degree six years after graduating high school is not affected, indicating that this is not the margin driving my results.

Thus, a potential mechanism of how grades at school affect adult outcomes is via changing outcomes at post-secondary stages and the subsequent entry into the labour market. I discuss two potential ways grades might have affected these outcomes: grades as a noisier signal after the reform and lower grades as a measure of human capital. Overall, my findings are consistent with a mechanism operating through lower human capital accumulation.

**Signalling**

The G8 reform might depress labour market outcomes through crowding out university admission if grades are a more noisy signal after the reform conditional on academic ability. For example, Büttner and Thomsen (2015) find that final exam grades in mathematics decreased between 0.62 and 0.86 points (G9 average: 7.8, from 0 being worst to 15 being best). Investigating the link between student test scores and adulthood outcomes in Mexico, Peña (2017) argues that if allocation mechanisms for opportunities in higher education are based on test scores, then lower test scores make it more difficult to get into rigorous study programmes at renowned universities. In the German case, admission to high-demand study programmes is based on a GPA ranking rather than a fixed cut-off GPA, implying that any cut-off for admission is determined in retrospect. Therefore, when competition for scarce university slots was fierce for the double graduation cohort, the first G8 cohort should incur the largest negative effects as a G8 graduate would be denied admission to a competitive study programme in favour of a G9 graduate conditional on academic talent. However, figure 1.5 shows that the effect of the reform on labour earnings becomes more negative
for more recent cohorts, providing evidence against this explanation. If admission to competitive programmes at university was the main mechanism behind my results, then the share of restricted study programmes should also, at least weakly, increase. Marcus and Zambre (2019) show that there was no tightening in retrospectively set cut-offs due to the reform, even considering the double graduation cohort. Additionally, Meyer and Thomsen (2016) find that the choice of study programme was hardly affected by the reform. Men tend to study fewer subjects in the educational and social sciences, though the effect is transitory. Women are less likely to enrol in STEM subjects but more likely to study medical sciences. Regarding final grades attained at the undergraduate level and completion rates, Meyer and Thomsen (2018) do not find any effect of the G8 reform. Thus, a mechanism involving G8 graduates being crowded out into less competitive programmes at university or changes in the preference for study subjects seems unlikely to explain my results.

Instead of attending university, graduates of the academic track can start an apprenticeship. When the double graduation cohort finished high school, twice the amount of graduates entered the apprenticeship market. When receiving applications from graduates, employers cannot directly distinguish whether an applicant completed the academic track under the G8 or the G9 system based on the certificate.\footnote{Upon request, the \textit{KMK} explained that the university entrance certificates do not indicate whether the academic track was completed within eight or nine years. Employers could use an applicant's date of birth and calculate the expected year of school entry. Using the calculated year of school entry and comparing it with the year when the G8 system was introduced in the state of high school graduation could give some indirect indication. The \textit{KMK} also explained that information about foreign languages on the back side of the document could potentially indicate under which system the applicant completed the academic track.} Thus, only accompanying documents, such as CVs or motivation letters, would show whether the academic track was attended for eight or nine years. If firms take into account that the distribution of grades might change due to the reform without affecting human capital, then this should not drive my results. If firms simply compare grades of otherwise identical applicants, then G8 graduates might be initially offered lower pay at firms that pay, on average, lower wages due to more noisy grades. These effects might be persistent and could lower earnings in the long run (Wachter, 2020;
Bardhi et al., 2020), even if firms learn about the "true" underlying human capital and start offering higher wages. With my dataset, it is not feasible to quantify the proportion of firms that take the G8 reform into account when making hiring decisions. One way of probing the importance of this mechanism is to use linked survey data asking firms about their hiring process.

**Human Capital Accumulation**

What could drive large and negative effects on labour market outcomes is a reduction in human capital accumulation due to the G8 reform. Suppose a very general framework of job search and investment in multidimensional skills as in Sanders and Taber (2012). Let human capital be a combination of cognitive skills and various non-cognitive skills. If the reform reduced cognitive and non-cognitive skills accumulated during high school, human capital accumulation through attending university or completing an internship might also be negatively affected. University education likely increases cognitive skills rather than non-cognitive skills, whereas apprenticeships are more focused on non-cognitive skills. Thus, G8 graduates who attend university likely accumulate fewer cognitive skills than non-university graduates. Although cognitive skills have large effects on wages for university graduates and non-university graduates (Yamaguchi et al., 2018), the reform should have larger negative effects on university graduates than on non-university graduates, as the returns for cognitive skills are much higher compared to non-cognitive skills (Lise and Postel-Vinay, 2020).

Exploring this mechanism, I estimate equation 1.8 separately for university and non-university graduates. Results are presented in figures 1A.5 and 1A.6 in the appendix for university and non-university graduates, respectively. The estimated coefficients show a strong negative relationship between the G8 reform and yearly labour earnings in the long run for university graduates. The estimates become more negative for more years since graduating high school. The reform is associated with lower yearly labour earnings measured six years after graduation of university graduates by more than EUR 5,240 (overall TWFE estimates: 2,844.41; SE: 314.61).
The coefficients for non-university graduates are overall smaller in magnitude. The point estimates suggest that the reform is associated with lower yearly labour earnings measured six years after graduating by about EUR 1,880 for non-university graduates. These findings align with the notion that university graduates accumulate cognitive skills at a slower pace compared to the counterfactual, providing additional evidence related to a human capital mechanism. One potential confounding factor correlated with outcomes and the reform is age. If age is an important component of accumulating human capital or earnings, then the estimates might partially reflect an age effect. When discussing the relative importance of channels through which the reform affected outcomes, I control for age effects by estimating my main specification only on a sample of G8 and G9 graduates with the same calendar age. My results are slightly smaller in magnitude but qualitatively identical.

Overall, my findings are consistent with a mechanism that operates through lower human capital accumulation of G8 graduates and perhaps partially through signalling on the apprenticeship level. A theoretical model of job search and investment in multidimensional skills predicts a more negative effect of the reform on university graduates than on non-university graduates due to fewer accumulated cognitive skills, which is supported by empirical evidence. A negative effect from signalling emerges from firms not taking into account that the distribution of grades could change due to the reform without affecting human capital. If graduates initially work for lower wages at low-wage firms, the negative effect on wages might be persistent in the long run. However, quantifying the importance of this mechanism requires additional data, for example, linked survey data asking firms about their hiring process.

### 1.7.2 Channels

In this section, I provide evidence on the importance of three main channels through which the G8 reform affected labour market outcomes via the described mechanism: 1) curriculum changes with higher workloads and reformed timing, 2) lower age at test taking, and 3) allocation of teachers’ attention.
Higher Workload and Reformed Timing

A key building block of the reform was to increase the weekly workload of pupils and change the timing of when subject matter was first introduced in class. At the same time, students who were exposed to the new curriculum also graduated a year earlier. Therefore, disentangling this change-of-curriculum effect from an age effect requires comparing G8 and G9 cohorts while holding the calendar age fixed when working towards the final GPA and taking final exams. Exploiting the birth date cut-offs for determining school start, I compare outcomes measured six years after graduation of G8 and G9 graduates with the same calendar age when completing the academic track.\textsuperscript{14} Although the calendar age of G8 and G9 graduates are the same, their age relative to their classmates is different. G9 graduates are relatively young to their classmates, while G8 graduates are relatively old. Recent empirical research suggests that the effect of relative age on test scores is small (Black et al., 2011; Peña, 2017; Cascio and Schanzenbach, 2016). Peña (2017) also estimates the effect of relative age on earnings, finding small effects.

Figure 1.8 shows the results of estimating the curriculum change effect of the G8 reform on labour market outcomes measured six years after graduation while keeping the calendar age constant. The effect is about two-thirds the size compared to the estimated effect on yearly labour earnings in the baseline results. Point estimates reach between EUR 2,200 and over EUR 4,000, depending on the estimator. The coefficients are less precisely measured for some cohorts compared to the main specification. The main reason for this is likely the reduced sample size of underlying observations for each state-school entry cohort cell. Similarly, the curriculum change effect on daily wage compares to about two-thirds of the measured effect in the main results. Confidence intervals around the estimated effect become wide for more recent cohorts.

\textsuperscript{14}Despite G8 and G9 graduates having the same calendar age when completing the academic track, their average age difference is still about six months.
**Lower Age at Test Taking**

The age effect arises as the reform lowered the average age by one year when working towards the final GPA and taking final exams at the end of high school. Past research by Black et al. (2011), Schneeweis and Zweimüller (2014) and Peña and Duckworth (2018) show that calendar age has a positive effect on test scores given an individual's relative age to their classmates. The reduction in test scores due to lower age at test taking likely links back to the production function of human capital, in which age might be an important factor. Thus, this age effect likely drives part of the negative effect on pupils’ test scores and labour earnings under the G8 system. However, as students who graduated a year earlier were also exposed to the new curriculum, disentangling the age effect would require a shift in pupils’ age at each grade without a change in the curriculum. As the estimated effects related to curriculum changes are about two-thirds in size compared to the estimated event-study coefficients in figure 1.6, the age effects are likely smaller in magnitude.

**Allocation of Teachers’ Attention**

The reform changed not only the structure of the academic track for students but also for teachers. As the reform changed pupils’ weekly workload, timing and requirements of grade-specific competencies and years spent on the academic track, teachers needed to update materials and revise teaching styles. As shown in figure 1.2, pupils could be two grades lower than their G9 peers when learning the same concepts. Consequently, the reform put additional strain on teachers due to having to make adjustments to accommodate the changes in content structure. One concern could be that teachers allocate more time and effort in preparing for lessons under the G8 system and away from G9 cohorts. If this reallocation of "attention" in favour of G8 cohorts happened at the expense of G9 cohorts, my estimates could be biased. If cohorts graduating right before the double graduation cohorts were also affected by the reform as teachers pay more attention to G8 lessons, these cohorts would not be a valid control group for the main specification. In this case, the control group
would also be on a treatment trajectory, meaning treatment effects would be used for imputing the counterfactual for the treatment group. As Goodman-Bacon (2021) shows, using treatment trajectories to calculate counterfactuals under a parallel trend assumption leads to biased estimates. Thus, this exercise can also be seen as a test of the potential spillover effects of the G8 reform on G9 cohorts.

I estimate the effect of teachers teaching G8 and G9 classes in the year by shifting the treatment year for each state such that the last untreated cohort is the G9 cohort that has never attended school with any G8 pupil. The first treatment cohort was in 13th grade when the first G8 cohort entered the academic track, while the fifth cohort after the first treatment cohort was in 8th grade when the first G8 cohort entered the academic track. Thus, later G9 cohorts were more exposed to teachers potentially allocating more attention to G8 cohorts than earlier G9 cohorts.

Figure 1.9 shows results from estimating a regression specification as in equation 1.7 using a selection of estimators robust to treatment effect heterogeneity. Across all outcomes, the estimated effect of being taught by teachers who simultaneously instructed G8 and G9 cohorts on pupils graduating under the G9 system is not systematically different from zero. I do not find any effect on labour market outcomes, even for cohorts that spend most of their time on the academic track together with pupils on the G8 track. Estimates are similar across the selected estimators. These results present evidence that teachers do not allocate more attention or resources to G8 cohorts at the expense of G9 cohorts. However, my findings cannot rule out that teachers are allocating more attention to G8 cohorts than G9 cohorts but at the expense of their own leisure.15

1.8 Conclusion

In this paper, I examine the effects of increasing efficiency in the education system on labour market outcomes using administrative data from Germany.

15The education worker's union has long opposed the G8 reform due to higher learning intensity and longer days harming the quality of education. After the G8 system was introduced, representatives of the union have since advocated for a return to the G9 system.
In the early 2000s, the German Conference of Ministers of Education (Kultusministerkonferenz) agreed to reduce the years a student spends on the academic track, Gymnasium, from nine to eight years. The G8 reform was intended to allow graduates to accumulate human capital faster and enter the labour market earlier. German high school graduates were among the oldest compared to their peers internationally around the late 1990s. The reform aimed to preserve the level of accumulated human capital upon graduating by not changing the scope of the curriculum.

Using the SIAB, an administrative data set derived from social security records, I find that the G8 reform reduces the yearly labour earnings of G8 graduates by over EUR 2,840 six years after graduation when tertiary education is usually completed. The effect is not transitory and persists for students who have graduated up to 4 years after the double graduation cohort. In line with previous research, I do not find that G8 graduates are less likely to have a college degree. Post-reform graduates work for firms that pay lower wages and have less well-paid positions at these firms. As states introduced the reform at different times, I verify that the estimated effects are robust to treatment effect heterogeneity by using heterogeneity-robust estimators recently proposed in the literature. A range of alternative specifications and placebo tests confirm the robustness of my main findings. Taking a longer-run perspective, I show that the negative effects of the G8 reform are increasing over time and are measurable even ten years after graduating from high school.

Recent research has established that the G8 reform reduced the overall GPA and final grades of G8 pupils. Lower grades can either be due to noise in the educational signal or an indicator of lower human capital accumulation. My findings are consistent with a mechanism that operates through lower human capital accumulation of G8 graduates and perhaps partially through signalling on the apprenticeship level. A theoretical model of job search and investment in multidimensional skills predicts a more negative effect of the reform on university graduates when compared to non-university graduates due to fewer accumulated cognitive skills, which is sup-
ported by my empirical findings. I provide evidence on the importance of three main channels through which the G8 reform affected labour market outcomes via the described mechanism: 1) curriculum changes with higher workloads and reformed timing, 2) lower age at test taking, and 3) allocation of teachers’ attention. The large and negative effects are primarily driven by changes to the curriculum and only to a smaller extent by an age effect. I do not find any evidence for teachers’ allocating more attention to pupils of G8 cohorts at the expense of G9 cohorts.

My results paint a negative picture of the G8 reform and its effects on labour market outcomes. Even though the reform likely increases cash flow within the social security system by graduates entering the labour market one year earlier, the lifetime earnings of post-reform graduates might be persistently lower. Then, the net benefit of a policy change like the German reform might be negative. Disentangling the relevance of each mechanism showed that the reform led to worse labour market outcomes due to lower human capital accumulation caused by compressing the curriculum. At the time of writing, some states have already reverted to reintroducing the G9 system in academic-track high schools. However, as German society is ageing rapidly, this is likely not the last reform to the education system. Policymakers can use these results to inform the design of future reforms.
References - Chapter 1


The data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.


**Figures - Chapter 1**

**Figure 1.1: The German School System**

Note: The figure illustrates the German school system. Children enter the German school system by starting to attend *Kindergarten* at age 3. Children usually enter primary school in the year they turn six. Primary schools teach German, English as a foreign language, mathematics and sciences. When finishing fourth grade, pupils receive a recommendation from their primary school for one of the secondary school tracks. There are three main school tracks. *Hauptschule* focuses on acquiring vocational skills and preparing students for vocational school or entry-level jobs. *Hauptschule* finishes after ninth grade with a secondary degree. This school track does not exist in some states, while some have started to integrate this track into comprehensive schools in recent years. At *Realschule*, pupils learn a broader range of subjects. This track prepares students for vocational training or higher education by teaching theoretical and practical skills. The most demanding track is the academic track, *Gymnasium*. *Gymnasium* is split into two levels. Level 1 is defined as grades 5 to 10 and prepares for grades 11 to 13 at level 2. Marks attained in the last three years count towards the final GPA of the university entrance certificate. Before the reform, pupils spent nine years on the academic track. After the reform, pupils spend eight years on the academic track. The difference is marked by the red, shaded area. The reform reduced the number of years spent on level 1 from six to five years. The reform was designed not to compromise the quality of education. Additionally, *Gesamtschulen* are comprehensive schools offering all degrees obtained on other tracks. Arrows represent progressions and transitions along and across tracks.
Note: The figure shows an example core curriculum of mathematics from the largest state by population, North Rhine-Westphalia (Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen, 2004; Ministerium für Schule, und Weiterbildung des Landes Nordrhein-Westfalen, 2007). The G8 reform in the German school system resulted in changes to the timing and requirements of specific competencies within evaluation categories. Panel (a) shows an overview of the core curriculum for mathematics as set by the state before the reform. Panel (b) shows the overview of the core curriculum for mathematics as set by the state after the reform. The evaluation categories in the column headers remained the same before and after the reform. Before the reform, the highest grade in the overview was grade 10, while the highest grade in the outline was grade 9 after the reform. Overall, the types of competencies stayed the same within each category. However, pupils were expected to have acquired these competencies at different grades before and after the reform. In the Reasoning/Communicating column, the skill to extract information from authentic texts, such as newspapers, was a grade 9/10 competency for pre-reform cohorts. For post-reform cohorts, this is a grade 7/8 competency. Other skill blocks, such as presenting, verifying and evaluating problem-solving approaches, were divided into multiple competencies in the post-reform curriculum. Sometimes, the wording changed to draw attention to specific competencies explicitly, for example, in the Arithmetic/Algebra category. The pre-reform overview mentioned natural numbers, while the post-reform overview refers to integers. Although integers were not explicitly mentioned in the pre-reform overview, working knowledge of integers is required to comprehend techniques taught in later grades.
Figure 1.3: Staggered Introduction of the *Gymnasium* in 8 Years (G8) Reform

*Note:* The figure shows the staggered introduction of the G8 reform. The transition timing from nine years of *Gymnasium* to eight years varied across German states as education policy is subject to state legislation. The first G8 cohort graduated alongside the last G9 cohort in each state. I refer to these two entry cohorts as the double graduation cohort. Saxony and Thuringia already had an eight-year *Gymnasium* system at the time of the reform as a legacy of the former German Democratic Republic. Hesse and Rhineland-Palatinate had unique implementations, with Hesse introducing the reform gradually over three years and Rhineland-Palatinate implementing G8 to a specific number of *Gymnasiums* only.
Figure 1.4: Share of Pupils Finishing Gymnasium after 9 and 8 Years

Note: The figure shows the share of G9 (blue, solid) and G8 (red, dashed) pupils on the academic track for states in my sample across all grades over time. The first G8 cohort entered 5th grade of the academic track in Saxony-Anhalt in 2001. As more states adopted the reform and more G8 cohorts entered the academic track of early-adopter states, the share of G9 pupils on the academic track decreased until all pupils were taught under the G8 system in all states after 2016. The first G8 entry cohort graduated simultaneously with the last G9 entry cohort in each state. I refer to these two entry cohorts as the double graduation cohort. The figure is based on administrative data on enrolled pupils on the academic track from the German federal statistical office provided by Statistisches Bundesamt (Destatis) (2023).
Note: The figure presents event-study plots with estimates from five different estimators for yearly labour earnings as the outcome: TWFE specification in equation 1.7 using OLS (red, circles); Callaway and Sant’Anna (2021) (blue, triangles); Borusyak et al. (2023) (green, diamonds); Chaisemartin and d’Haultfoeuille (2020) (orange, squares); and Sun and Abraham (2021) (black, crosses). The grey, shaded area represents the double graduation cohort. The outcome variable is yearly labour earnings measured six years after graduation. The time variable is the year of school entry for a cohort, and the treatment group indicator is the year of school entry for the first cohort graduating under the G8 system. The dummy for the school entry cohort one year before the double graduation cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort. All monetary values are denoted in 2015 euro. Bars represent 95 percent confidence intervals. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Figure 1.6: Effect of the G8 Reform on Labour Market Outcomes

<table>
<thead>
<tr>
<th>(a) Employment</th>
<th>(b) Experience</th>
<th>(c) Tenure</th>
<th>(d) College Degree</th>
<th>(e) Daily Wage</th>
<th>(f) Mean Wage at Firm</th>
<th>(g) Relative Distance to Mean Wage at Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The figure presents event-study plots with estimates from five different estimators for multiple labour market outcomes: TWFE specification in equation 1.7 using OLS (red, circles); Callaway and Sant’Anna (2021) (blue, triangles); Borusyak et al. (2023) (green, diamonds); Chaisemartin and d’Haultfoeuille (2020) (orange, squares); and Sun and Abraham (2021) (black, crosses). The grey, shaded area represents the double graduation cohort. The outcome variables are measured six years after graduation. The time variable is the year of school entry for a cohort, and the treatment group indicator is the year of school entry for the first cohort graduating under the G8 system. The dummy for the school entry cohort one year before the double graduation cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort. All monetary values are denoted in 2015 euro. Bars represent 95 percent confidence intervals. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Figure 1.7: Effect of the G8 Reform in the Longer Run

(a) Yearly Labour Earnings
(b) Employment
(c) Experience
(d) Tenure
(e) Daily Wage
(f) Mean Wage at Firm
(g) Relative Distance to Mean Wage at Firm

Note: The figure shows the effects of the G8 reform on labour market outcomes in different years after graduating high school. Some school entry cohorts are only observed for a limited number of years due to the recency of the reform. I show results for estimating equation 1.8 via OLS on subsamples of school entry cohorts: cohorts 1989 to 2000 for up to six years after graduation (circles; X refers to the results in column 1 of table 1.2); cohorts 1989 to 1999 for up to 7 years after graduation (crosses); cohorts 1989 to 1998 for up to 8 years after graduation (squares); cohorts 1989 to 1997 for up to 9 years after graduation (diamonds); cohorts 1989 to 1996 for up to 10 years after graduation (triangles). All specifications include state-fixed effects, school entry cohort-fixed effects and a treatment indicator for having graduated under the G8 system on the right-hand side. All monetary values are denoted in 2015 euro. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Figure 1.8: Effect of Curriculum Changes

(a) Yearly Labour Earnings
(b) Employment
(c) Experience
(d) Tenure
(e) College Degree
(f) Daily Wage
(g) Mean Wage at Firm
(h) Relative Distance to Mean Wage at Firm

Note: The figure presents event-study plots of the estimated curriculum change effect on multiple labour market outcomes using five different estimators: TWFE specification in equation 1.7 using OLS (red, circles); Callaway and Sant’Anna (2021) (blue, triangles); Borusyak et al. (2023) (green, diamonds); Chaisemartin and d’Haultfoeuille (2020) (orange, squares); and Sun and Abraham (2021) (black, crosses). The grey, shaded area represents the double graduation cohort. The outcome variables are measured six years after graduation, and all individuals are 25 years old. The time variable is the year of school entry for a cohort, and the treatment group indicator is the year of school entry for the first cohort graduating under the G8 system. The dummy for the school entry cohort one year before the double graduation cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort. All monetary values are denoted in 2015 euro. Bars represent 95 percent confidence intervals. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Figure 1.9: Effect of Teachers’ Attention Allocation

The figure presents event-study plots of the estimated effect of teachers’ attention allocation on multiple labour market outcomes using five different estimators: TWFE specification in equation 1.7 using OLS (red, circles); Callaway and Sant’Anna (2021) (blue, triangles); Borusyak et al. (2023) (green, diamonds); Chaisemartin and d’Haultfoeuille (2020) (orange, squares); and Sun and Abraham (2021) (black, crosses). The grey, shaded area represents the double graduation cohort. The outcome variables are measured six years after graduation. The time variable is the year of school entry for a cohort, and the treatment group indicator is the year of school entry for the first G9 cohort that first overlapped with a G8 cohort. The dummy for the school entry cohort one year before the first treatment cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort. All monetary values are denoted in 2015 euro. Bars represent 95 percent confidence intervals. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.

Note: The figure presents event-study plots of the estimated effect of teachers’ attention allocation on multiple labour market outcomes using five different estimators: TWFE specification in equation 1.7 using OLS (red, circles); Callaway and Sant’Anna (2021) (blue, triangles); Borusyak et al. (2023) (green, diamonds); Chaisemartin and d’Haultfoeuille (2020) (orange, squares); and Sun and Abraham (2021) (black, crosses). The grey, shaded area represents the double graduation cohort. The outcome variables are measured six years after graduation. The time variable is the year of school entry for a cohort, and the treatment group indicator is the year of school entry for the first G9 cohort that first overlapped with a G8 cohort. The dummy for the school entry cohort one year before the first treatment cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort. All monetary values are denoted in 2015 euro. Bars represent 95 percent confidence intervals. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
### Table 1.1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Individual-Level Outcome</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Obs</th>
<th>Cohorts</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yearly Labour Earnings</td>
<td>15,364.87</td>
<td>16,693.75</td>
<td>71,637</td>
<td>1989-2000</td>
<td>12</td>
</tr>
<tr>
<td>Daily Wage</td>
<td>44.28</td>
<td>46.46</td>
<td>71,637</td>
<td>1989-2000</td>
<td>12</td>
</tr>
<tr>
<td>Employed</td>
<td>0.718</td>
<td>0.450</td>
<td>71,637</td>
<td>1989-2000</td>
<td>12</td>
</tr>
<tr>
<td>College Degree</td>
<td>0.481</td>
<td>0.500</td>
<td>71,637</td>
<td>1989-2000</td>
<td>12</td>
</tr>
<tr>
<td>Age</td>
<td>25.41</td>
<td>0.58</td>
<td>71,637</td>
<td>1989-2000</td>
<td>12</td>
</tr>
<tr>
<td>Experience</td>
<td>1,817.97</td>
<td>1,237.72</td>
<td>71,637</td>
<td>1989-2000</td>
<td>12</td>
</tr>
<tr>
<td>Tenure</td>
<td>839.44</td>
<td>653.61</td>
<td>51,786</td>
<td>1989-2000</td>
<td>12</td>
</tr>
</tbody>
</table>

| Firm-Level Outcome                           |         |         |         |            |        |
| Mean Wage at Firm                            | 124.73  | 49.10   | 39,275  | 1989-2000  | 12     |
| Relative Distance to Firm Mean Wage          | -.444   | .358    | 39,245  | 1989-2000  | 12     |

| Observations by age                          |         |         |         |            |        |
| Age = 24                                     | 24      | 0       | 3,471   |            |        |
| Age = 25                                     | 25      | 0       | 35,093  |            |        |
| Age = 26                                     | 26      | 0       | 33,073  |            |        |

**Note:** The table presents summary statistics for the main labour market outcomes from the SIAB measured six years after graduation. Obs refers to the number of individual-level observations. Depending on the outcome, the number of observations varies. Mean Wage at Firm and Relative Distance to Firm Mean Wage exclude firms with less than ten employees. Relative Distance to Firm Mean Wage is measured as the difference between the mean wage at the firm and an individual's daily wage divided by the mean wage at the firm. Relative Distance to Firm Mean Wage excludes individuals that are recorded as employed but have 0 earnings in a given year (e.g. marginally employed workers). Cohorts refers to the school entry cohorts included in the sample. States is the number of states included in the sample. The states of Hesse, Rhineland-Palatinate, Saxony and Thuringia are excluded. Hesse introduced the reform gradually over three years. Rhineland-Palatinate implementing G8 to a specific number of Gymnasiums only. In Saxony and Thuringia, the academic track has always taken eight years to complete. All monetary values are denoted in 2015 euro.
Table 1.2: Baseline Results - Effect of G8 Reform on Labour Earnings and Employment

<table>
<thead>
<tr>
<th></th>
<th>Yearly Labour Earnings (1)</th>
<th>Employment (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of G8 Reform</td>
<td>-2,844.41</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(314.61; p: 0.02)</td>
<td>(0.008; p: 0.04)</td>
</tr>
<tr>
<td>State fixed effects</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cohort fixed effects</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Observations</td>
<td>71,637</td>
<td>71,637</td>
</tr>
<tr>
<td>State-Cohort Cells</td>
<td>144</td>
<td>144</td>
</tr>
</tbody>
</table>

Note: The table shows the estimated effect of the G8 reform on yearly labour earnings and employment ($\beta_{TWFE}$ from equation 1.6) using OLS. All outcomes are measured six years after graduating high school. Column 1 shows results from a regression specification with yearly labour earnings as the dependent variable. Column 2 shows results from a regression specification with employment as the dependent variable. Both specifications include state-fixed effects, school entry cohort-fixed effects and a treatment indicator for having graduated under the G8 system on the right-hand side. All monetary values are denoted in 2015 euro. Standard errors in parentheses are clustered at the state level. $p$-values are presented based on wild cluster bootstrapping (1,000 replications, Mammen weights, testing under $H_0$). All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Table 1.3: Effect of the G8 Reform on Labour Market Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Individual-Level Outcomes</th>
<th>Firm-Level Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experience (1)</td>
<td>Tenure (2)</td>
</tr>
<tr>
<td>Effect of G8 Reform TWFE</td>
<td>-26.01 (27.99; p: 0.71)</td>
<td>-20.92 (22.77; p: 0.44)</td>
</tr>
<tr>
<td>State fixed effects ✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cohort fixed effects ✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Observations 71,637</td>
<td>51,786</td>
<td>71,637</td>
</tr>
<tr>
<td>State-Cohort Cells 144</td>
<td>144</td>
<td>144</td>
</tr>
</tbody>
</table>

Note: The table shows the estimated effect of the G8 reform on labour market outcomes measured six years after graduation. Each column refers to a different outcome of interest. Depending on the outcome, the number of observations varies. Mean Wage at Firm and Relative Distance to Firm Mean Wage exclude firms with less than ten employees. Relative Distance to Firm Mean Wage excludes individuals that are recorded as employed but have 0 earnings in a given year (e.g. marginally employed workers). All specifications include state-fixed effects, school entry cohort-fixed effects and an indicator for completing the academic track under the G8 system. All monetary values are denoted in 2015 euro. Standard errors in parentheses are clustered at the state level, p-values are presented based on wild cluster bootstrapping (1,000 replications, Mammen weights, testing under $H_0$). All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Appendix 1A: Additional Figures and Tables

Figure 1A.1: Mathematics Competencies by Grade Translated to English

(a) Gymnasium in 9 year (G9)  
(b) Gymnasium in 8 year (G8)

<table>
<thead>
<tr>
<th>Competence</th>
<th>5/6</th>
<th>7/8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number sense and operations</td>
<td>Integers (addition and multiplication)</td>
<td>Decimals</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>Solving problems</td>
<td>Solving problems</td>
</tr>
<tr>
<td>Reasoning</td>
<td>Reasoning</td>
<td>Reasoning</td>
</tr>
<tr>
<td>Communication</td>
<td>Communicating</td>
<td>Communicating</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competence</th>
<th>5/6</th>
<th>7/8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functions</td>
<td>Linear functions</td>
<td>Quadratic functions</td>
</tr>
<tr>
<td>Geometry</td>
<td>Analytical geometry</td>
<td>Analytical geometry</td>
</tr>
<tr>
<td>Stochastics</td>
<td>Probability and statistics</td>
<td>Probability and statistics</td>
</tr>
</tbody>
</table>

Note: The figure shows an example core curriculum of mathematics from the largest state by population, North Rhine-Westphalia (Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen, 2004; Ministerium für Schule, und Weiterbildung des Landes Nordrhein-Westfalen, 2007). The G8 reform in the German school system resulted in changes to the timing and requirements of specific competencies within evaluation categories. Panel (a) shows an overview of the core curriculum for mathematics as set by the state before the reform. Panel (b) shows the overview of the core curriculum for mathematics as set by the state after the reform. This figure is a translated version of figure 1.2.
Figure 1A.2: Observed and Predicted Year of School Entry

Note: The figure shows the performance of my imputation procedure in step two by comparing the predicted year of school entry with the observed year of school entry using the German Socio-Economic Panel (Goebel et al., 2019). The SIAB does not contain detailed information about individuals’ schooling background, such as the years of schooling or the state where school was attended. Therefore, I construct a treatment indicator for whether each individual was part of a G8 cohort by using data on the state of residence when the first employment spell is recorded, the type of high school leaving certificate, and the month and year of birth. The procedure follows three steps. Firstly, I define the state where an individual attended school as the state where the individual lived at the time the first employment spell is observed in the SIAB. Secondly, I assign each individual to an entry cohort. Children enter primary school in the year they turn six if their birthday is before a state-specific cut-off date. Before 1997, the cut-off was June 30th in all states. Since 1997, each state has set the cut-off date individually. Thirdly, I assign each individual to have completed high school under the G8 or G9 system based on the observed state of residency at first appearance and the imputed year of school entry.
Figure 1A.3: Test Scores and Cohort Size around the Reform

(a) Reading
RD estimate: 0.96 (2.87)

(b) Maths
RD estimate: 4.87 (11.82)

(c) Science
RD estimate: -7.44 (18.64)

(d) Cohort Size
RD estimate: 1963.42 (3186.10)

Note: The figure plots residualised standardised test scores at grade 9 from a regression of test scores on state-fixed effects weighted by 9th-grade cohort size separately for reading, maths and science. States commonly introduced the reform when pupils entered 5th grade of Gymnasium. If the timing of the introduction of the reform was correlated with a confounding factor that also correlates with one of the labour market outcomes, then this would violate the parallel trend assumption. One such factor could be students’ ability. Policymakers might wait for an academically strong cohort to implement the reform. Until 2006, all test scores are PISA-E scores for reading, maths and science. Reading test scores for 2008 and maths and science test scores for 2012 are National Assessment Study test scores. Reading test scores for 2015 and maths and science test scores for 2018 are IQB Trends test scores. Additionally, if policymakers wanted to probe the reform, they would be more likely to do so with a smaller cohort. Using administrative data on enrolled pupils on the academic track from the German federal statistical office provided by Statistisches Bundesamt (Destatis) (2023), I measure cohort size in the last year of primary school before pupils are sorted into tracks. In all regressions, the discontinuity coefficient is statistically insignificant and contextually small. Confidence intervals are confidence intervals of the mean, and the linear fit is weighted by cohort size for test score outcomes.
Figure 1A.4: Grade Progression

(a) Year-on-year Repetition Rates

-5% -2.5% 0% 2.5% 5% 7.5% 10%

Drop-out rate

5 4 3 2 1 Last

Years until high school graduation

Panel (a) presents year-on-year repetition rates over the years until high school graduation for G8 (red, triangles) and G9 (blue, circles) cohorts. The dip in drop-out rates in the third year before high school graduation is explained by the influx of students having completed the Realschule track and entering the academic track to obtain the university entrance certificate.

(b) Grade Progression under G8/G9

80 85 90 95 100

Cohort Size (Grade 7 = 100)

6 5 4 3 2 1 Last

Years until high school graduation

Panel (b) displays a grade progression index over the years until high school graduation. Cohort sizes are indexed at grade 7. My measure of grade repetition cannot distinguish between track switchers, repeaters and school drop-outs. This causes a problem if any of those three reasons for irregular progression leads to fewer graduations from the academic track.

Note: The figure shows different measures of grade progression using administrative data on enrolled pupils on the academic track from the German federal statistical office provided by Statistisches Bundesamt (Destatis) (2023). If the reform changed who attends the academic track, then the parallel trend assumption would be violated as I condition on individuals to have a recorded university entrance certificate in the data. Panel (a) presents year-on-year repetition rates over the years until high school graduation for G8 (red, triangles) and G9 (blue, circles) cohorts. The dip in drop-out rates in the third year before high school graduation is explained by the influx of students having completed the Realschule track and entering the academic track to obtain the university entrance certificate. Panel (b) displays a grade progression index over the years until high school graduation. Cohort sizes are indexed at grade 7. My measure of grade repetition cannot distinguish between track switchers, repeaters and school drop-outs. This causes a problem if any of those three reasons for irregular progression leads to fewer graduations from the academic track.
Figure 1A.5: Heterogeneity of the Effect of the G8 Reform in the Longer Run, Only University Graduates

(a) Yearly Labour Earnings

(b) Employment

(c) Experience

(d) Tenure

(e) Daily Wage

(f) Mean Wage at Firm

(g) Relative Distance to Mean Wage at Firm

Note: The figure explores the heterogeneity of the effects of the G8 reform on labour market outcomes in different years after graduating high school on a sample of only university graduates. Some school entry cohorts are only observed for a limited number of years due to the recency of the reform. I show results for estimating equation 1.8 via OLS on subsamples of school entry cohorts: cohorts 1989 to 2000 for up to six years after graduation (circles; X refers to a specification as in column 1 of table 1.2 for a subsample of university graduates); cohorts 1989 to 1999 for up to 7 years after graduation (crosses); cohorts 1989 to 1998 for up to 8 years after graduation (squares); cohorts 1989 to 1997 for up to 9 years after graduation (diamonds); cohorts 1989 to 1996 for up to 10 years after graduation (triangles). All specifications include state-fixed effects, school entry cohort-fixed effects and a treatment indicator for having graduated under the G8 system on the right-hand side. All monetary values are denoted in 2015 euro. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Figure 1A.6: Heterogeneity of the Effect of the G8 Reform in the Longer Run, Only Non-university Graduates

(a) Yearly Labour Earnings  
(b) Employment  
(c) Experience  
(d) Tenure  
(e) Daily Wage  
(f) Mean Wage at Firm  
(g) Relative Distance to Mean Wage at Firm

Note: This figure explores the heterogeneity of the effects of the G8 reform on labour market outcomes in different years after graduating high school on a sample of only non-university graduates. Some school entry cohorts are only observed for a limited number of years due to the recency of the reform. I show results for estimating equation 1.8 via OLS on subsamples of school entry cohorts: cohorts 1989 to 2000 for up to six years after graduation (circles; X refers to a specification as in column 1 of table 1.2 for a subsample of non-university graduates); cohorts 1989 to 1999 for up to 7 years after graduation (crosses); cohorts 1989 to 1998 for up to 8 years after graduation (squares); cohorts 1989 to 1997 for up to 9 years after graduation (diamonds); cohorts 1989 to 1996 for up to 10 years after graduation (triangles). All specifications include state-fixed effects, school entry cohort-fixed effects and a treatment indicator for having graduated under the G8 system on the right-hand side. All monetary values are denoted in 2015 euro. Standard errors in parentheses are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Table 1A.1: Cut-off Dates for School Entry After 1997

<table>
<thead>
<tr>
<th>Federal State (Bundesland)</th>
<th>Cut-off date</th>
<th>Year of coming into effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schleswig-Holstein</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Hamburg</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Lower Saxony</td>
<td>flexible age to start school</td>
<td>2018</td>
</tr>
<tr>
<td>Bremen</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>31.07.</td>
<td>2007</td>
</tr>
<tr>
<td>Rhine-Westphalia</td>
<td>31.08.</td>
<td>2009</td>
</tr>
<tr>
<td>Bremen</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Rhineland-Palatinate</td>
<td>31.08.</td>
<td>2008</td>
</tr>
<tr>
<td>Baden-Wurttemberg</td>
<td>31.07.</td>
<td>2005</td>
</tr>
<tr>
<td></td>
<td>31.08.</td>
<td>2006</td>
</tr>
<tr>
<td></td>
<td>30.09.</td>
<td>2007</td>
</tr>
<tr>
<td>Bavaria</td>
<td>For children turning 6 between 01.07.an 30.09., the legal guardian decides about starting school</td>
<td>2020</td>
</tr>
<tr>
<td>Saarland</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Berlin</td>
<td>31.12.</td>
<td>2005</td>
</tr>
<tr>
<td></td>
<td>30.09.</td>
<td>2017</td>
</tr>
<tr>
<td>Brandenburg</td>
<td>30.09.</td>
<td>2005</td>
</tr>
<tr>
<td>Mecklenburg-Western Pomerania</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Saxony</td>
<td>30.09.</td>
<td>2004</td>
</tr>
<tr>
<td>Saxony-Anhalt</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>Thuringia</td>
<td>31.07.</td>
<td>2003</td>
</tr>
</tbody>
</table>

Note: The table provides an overview of the changes in cut-off dates by state and year used to infer the year of school start as provided by Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland - IV C - (2022) and Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland - IV C - (2019). The SIAB does not contain detailed information about individuals' schooling background, such as the years of schooling or the state where school was attended. Therefore, I construct a treatment indicator for whether each individual was part of a G8 cohort by using data on the state of residence when the first employment spell is recorded, the type of high school leaving certificate, and the month and year of birth. The procedure follows three steps. Firstly, I define the state where an individual attended school as the state where the individual lived at the time the first employment spell is observed in the SIAB. Secondly, I assign each individual to an entry cohort. Children enter primary school in the year they turn six if their birthday is before a state-specific cut-off date. Before 1997, the cut-off was June 30th in all states. Since 1997, each state has set the cut-off date individually. Thirdly, I assign each individual to have completed high school under the G8 or G9 system based on the observed state of residency at first appearance and the imputed year of school entry.
Table 1A.2: Placebo Results - Effect of G8 Reform on Unaffected Tracks

<table>
<thead>
<tr>
<th>Event time</th>
<th>Labour Earnings (1)</th>
<th>Employed Experience (2)</th>
<th>Tenure (3)</th>
<th>College Degree (4)</th>
<th>Daily Wage (5)</th>
<th>Mean Wage at Firm (6)</th>
<th>Distance to Firm Mean Wage (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t=-5</td>
<td>-825.99</td>
<td>0.042</td>
<td>-50.83</td>
<td>-193.96</td>
<td>-0.002</td>
<td>-1.28</td>
<td>-5.73</td>
</tr>
<tr>
<td></td>
<td>(1,327.45)</td>
<td>(0.042)</td>
<td>(96.99)</td>
<td>(67.44)</td>
<td>(0.01)</td>
<td>(3.88)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>t=-4</td>
<td>-800.35</td>
<td>-0.007</td>
<td>-42.11</td>
<td>-110.27</td>
<td>-0.007</td>
<td>-1.42</td>
<td>-3.81</td>
</tr>
<tr>
<td></td>
<td>(1,040.26)</td>
<td>(0.028)</td>
<td>(40.21)</td>
<td>(60.62)</td>
<td>(0.008)</td>
<td>(2.96)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>t=-3</td>
<td>-285.69</td>
<td>0.017</td>
<td>-13.22</td>
<td>-109.34</td>
<td>-0.003</td>
<td>0.01</td>
<td>-5.72</td>
</tr>
<tr>
<td></td>
<td>(709.63)</td>
<td>(0.026)</td>
<td>(58.05)</td>
<td>(32.5)</td>
<td>(0.005)</td>
<td>(2.20)</td>
<td>(2.08)</td>
</tr>
<tr>
<td>t=-2</td>
<td>-589.91</td>
<td>-0.055</td>
<td>6.42</td>
<td>-34.37</td>
<td>0.00</td>
<td>-1.3</td>
<td>-2.97</td>
</tr>
<tr>
<td></td>
<td>(650.58)</td>
<td>(0.017)</td>
<td>(50.73)</td>
<td>(41.91)</td>
<td>(0.004)</td>
<td>(1.84)</td>
<td>(1.26)</td>
</tr>
<tr>
<td>t=0</td>
<td>-136.07</td>
<td>0.000</td>
<td>53.26</td>
<td>48.93</td>
<td>-0.002</td>
<td>-0.32</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>(1,139.77)</td>
<td>(0.017)</td>
<td>(70.22)</td>
<td>(62.25)</td>
<td>(0.003)</td>
<td>(2.99)</td>
<td>(2.12)</td>
</tr>
<tr>
<td>t=1</td>
<td>1,899.70</td>
<td>0.023</td>
<td>195.11</td>
<td>202.43</td>
<td>-0.013</td>
<td>5.01</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td>(1,950.46)</td>
<td>(0.023)</td>
<td>(146.49)</td>
<td>(115.33)</td>
<td>(0.003)</td>
<td>(5.14)</td>
<td>(6.2)</td>
</tr>
<tr>
<td>t=2</td>
<td>-1,005.79</td>
<td>-0.039</td>
<td>23.6</td>
<td>30.69</td>
<td>-0.007</td>
<td>-3.73</td>
<td>-7.94</td>
</tr>
<tr>
<td></td>
<td>(1,669.71)</td>
<td>(0.087)</td>
<td>(124.31)</td>
<td>(64.59)</td>
<td>(0.008)</td>
<td>(4.82)</td>
<td>(4.79)</td>
</tr>
<tr>
<td>t=3</td>
<td>-1,091.00</td>
<td>-0.075</td>
<td>145.97</td>
<td>78.54</td>
<td>-0.013</td>
<td>-3.53</td>
<td>-12.06</td>
</tr>
<tr>
<td></td>
<td>(1,416.82)</td>
<td>(0.039)</td>
<td>(144.11)</td>
<td>(69.78)</td>
<td>(0.01)</td>
<td>(4.02)</td>
<td>(4.59)</td>
</tr>
<tr>
<td>t=4</td>
<td>-1,690.11</td>
<td>0.045</td>
<td>75.56</td>
<td>210.19</td>
<td>0.008</td>
<td>-4.32</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>(1,391.52)</td>
<td>(0.037)</td>
<td>(147.24)</td>
<td>(110.73)</td>
<td>(0.028)</td>
<td>(4.07)</td>
<td>(7.84)</td>
</tr>
<tr>
<td>t=5</td>
<td>-1,131.06</td>
<td>0.01</td>
<td>31.37</td>
<td>-28.58</td>
<td>-0.026</td>
<td>-3.09</td>
<td>-10.49</td>
</tr>
<tr>
<td></td>
<td>(1,678.94)</td>
<td>(0.042)</td>
<td>(110.04)</td>
<td>(89.82)</td>
<td>(0.007)</td>
<td>(4.72)</td>
<td>(6.18)</td>
</tr>
</tbody>
</table>

Observations 8,552 8,552 8,552 6,092 8,552 8,552 4,015 4,008
State-Cohort Cells 143 143 143 141 143 143 134 134

Note: The table presents results from repeating the OLS estimation of equation 1.7 on a sample of individuals with no degree higher than a secondary degree (Hauptschule or no degree). After finishing high school, these graduates have a very different trajectory than individuals with a university entrance certificate. Therefore, these two groups are unlikely to be substitutes in the labour market. All treatment indicators remain unchanged. All labour market outcomes are measured six years after graduation. The dummy for the school entry cohort one year before the double graduation cohort is omitted in OLS regressions so that the estimated coefficients measure the effect of the G8 reform relative to the omitted cohort. All monetary values are denoted in 2015 euro. Each column refers to a different outcome of interest. Standard errors in parenthesis are clustered at the state level. All regressions are run on the state-school entry cohort level and are weighted by the number of observations in each state-school entry cohort cell.
Table 1A.3: Effect of the G8 Reform on Number of Graduates on the Academic Track

<table>
<thead>
<tr>
<th>Number of Graduates from Academic Track</th>
<th>Level (1)</th>
<th>Log (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of G8 Reform</td>
<td>-659.218</td>
<td>-0.074</td>
</tr>
<tr>
<td></td>
<td>(430.448)</td>
<td>(0.055)</td>
</tr>
</tbody>
</table>

State fixed effects ✓ ✓
Cohort fixed effects ✓ ✓

State-Cohort Cells 192 192

Note: The table shows results from a regression specification as in equation 1.6 with the number of graduates from the academic track as the dependent variable using administrative data on the number of pupils in each grade, on every school track by year (Statistisches Bundesamt (Destatis), 2023). My measure of grade repetition in figure 1A.4 cannot distinguish between track switchers, repeaters and, although unlikely, school drop-outs. This causes a problem if any of those three reasons for irregular progression leads to fewer graduations from the academic track. Standard errors in parenthesis are clustered at the state level. As the data on school leavers spans more years than the main sample, I include entry cohorts between 1992 and 2007 in 12 states.
Chapter 2

Income Taxes and the Timing of Marital Dissolution: Evidence from Germany

2.1 Introduction

Tax incentives influence individuals’ behaviour, especially when small changes in timing can lead to large penalties or gains. While it is well-established that couples can shift the timing of marriage (Alm and Whittington, 1997b; Fink, 2020), childbirth (Dickert-Conlin and Chandra, 1999; LaLumia et al., 2015) or death (Kopczuk and Slemrod, 2003) to qualify for monetary benefits, the timing of marital dissolution has received limited attention. I investigate whether rebates in income tax liabilities influence the timing of marital breakups.

When deciding to delay the date of marital breakup, individuals face a trade-off between the emotional costs of continuing to live together despite having decided to end the marriage and pecuniary gains. In the German income taxation system, such pecuniary gains arise by allowing married couples to file for joint taxation. Income tax liabilities under joint taxation are never higher compared to those under individual taxation for a couple with the same taxable income. However, couples are only
eligible for joint taxation if they are married and not separated for at least one day in
the calendar year. I will refer to this rebate in income tax liabilities for married couples
as the "marriage subsidy".

The German law differentiates between the date of divorce and the date of separ-
aration. Before married couples can be legally divorced, spouses must live legally
separately for at least 12 months before filing for divorce in Germany. Couples must
inform their local tax office about the separation, including the date of separation,
by filling in a designated form\(^1\) before submitting it to the tax office. Eligibility for
joint taxation is based on this document. The precise date is also recorded when
spouses start the process of a legal divorce through a solicitor. During the process of
legal divorce, at least one partner must be represented by a solicitor. Commonly, the
separation starts with the separation of “table and bed”. Accepted proofs of the date
are a spouse renting a new flat for single occupancy or registering a new residence
with the local registration office, family, friends or a new partner as witnesses in
court, and a letter of separation personally delivered and signed by the other partner.
Spouses can also ask a solicitor to draft a document to record the date of separation
used as proof in court. This option is recommended if the separation of “table and
bed” occurs in the same flat. In this case, partners must have separate economies,
not have meals together and sleep in separate beds.

Understanding the timing of separation is important to economic welfare, as de-
laying the date of separation might increase the risk of domestic violence between
partners. Additionally, prolonging the time until separation and divorce could reduce
courtship time before the next marriage, leading to potentially less stable subsequent
marriages (Roeder and Ullmann, 2019), which in turn might negatively affect fertility
(Alesina and Giuliano, 2006). This tax rule is also a sizeable expense in the yearly
budget. Over the last 30 years, about 150,000 divorces were recorded in Germany.
A back-of-the-envelope calculation suggests that the German government spends

\(^1\)Figure 2B.1 in the appendix shows the form which is available on https://www.elster.de/eportal/
formulare-leistungen/alleformulare/elevelegetrenntlebend
nearly EUR 20 million yearly on this tax rule. As this policy has been in place for many decades, total spending on this tax rule amounts to well above EUR 500 million for the last 30 years.

Using rich survey data from the German Socio-Economic Panel, I establish that the average yearly marriage subsidy is about 3% of pre-tax earnings of the average household in my sample, which is equivalent to about EUR 1,500 measured in 2014 prices. Between 1984 and 2017, I observe a bunching of separations for married couples in the first quarter, especially in January, and a missing mass of separations in the last quarter, especially in December. Over the same period, I do not find any bunching for unmarried couples. This bunching is consistent with spouses postponing their separation date in response to the discrete granting of the marriage subsidy. Although not all separations occurring in the first quarter are tax-motivated, I consider separations occurring in the first quarter as potentially delayed separations in response to the marriage subsidy. I find that an increase in the marriage subsidy by EUR 1,000 is associated with a 2.9 percentage point higher probability of separating in the first quarter of the following calendar year. In my preferred specification, identifying variation of this coefficient mostly stems from non-linearities between the income levels of both spouses and the marriage subsidy. Surprisingly, my estimate is almost identical to the incentive to shift the timing of marriage forward, which is about 3.1 percentage points (Fink, 2020). The results are robust to a series of placebo exercises and alternative specifications. My estimates suggest that each additional child is associated with an increase in the probability of separation in the first quarter by about 10 percentage points. Thus, the presence of children and fertility considerations likely play an important role in determining the timing of separation besides the marriage subsidy. These findings contribute to the literature by providing strong evidence that not only the decision to separate is subject to tax incentives but also the timing.

I describe the distribution of characteristics for couples delaying their separation
timing, employing techniques similar to identifying compliers in instrumental variable settings (Angrist and Pischke, 2009). Over 17% of all separations reported in the first quarter are delayed in response to the tax rule. I show that delaying couples are more likely to be over 40 years old and are slightly more likely to be married for more than five years, suggesting that delaying couples have completed family planning. Women in delaying couples are less likely to contribute less than half to households’ gross income.

Finally, I add to the literature on the relationship between tax incentives and the probability of divorce. As delaying the date of separation requires spouses to cooperate, some couples might be closer to reconciling their marriage. If a married couple wants to divorce in Germany, spouses must have lived separately for at least 12 months before the marriage can be legally divorced. Thus, shifting the date of separation to a later date results in spouses extending this waiting period, which might help rectify their relationship. Estimating a Cox proportional hazards model (Cox, 1972) of the probability of divorce, I find that couples who separate in the first quarter of a year are equally likely to divorce at various time horizons as couples separating in other months. While timing decisions neither delay divorce nor affect its probability, higher earnings of wives are associated with swifter divorces after separation.

The rest of the paper is structured as follows. Section 2 summarises key findings of previous literature on income tax incentives and marital decisions. Section 3 describes the German income tax system and its incentive for marital timing decisions. Section 4 describes the sample selection and how the marriage subsidy is calculated alongside some descriptive statistics. Section 5 presents the empirical strategy. Section 6 presents the main results, including different robustness checks, and describes delayer characteristics. Section 7 briefly describes the German divorce law and compares the probability of divorce for different time horizons, and section 8 concludes.
2.2 Income Tax Incentives and Marital Decisions

This paper contributes to the literature on estimating the incentive of taxation on decision-making and shifting the timing of these decisions. My study is closely related to Fink (2020) investigating whether couples shift the timing of marriage in response to income tax laws. Using the same German survey data as I do in this paper, the author establishes that couples marry more often in December in Germany than in other countries. Using a logit model, he estimates whether couples with higher tax gains through marriage are more likely to marry in December. The author’s preferred specification suggests that an increase in the marriage subsidy by EUR 1,000 is associated with a 3.1 percentage point higher probability of marrying in the last quarter of a year. Roeder and Ullmann (2019) show that couples who marry in December experience, on average, shorter-lasting marriages. They argue that these couples are more likely to be mismatched in long-term relationships due to foregone courtship time. For the US, Alm and Whittington (1997b) estimate that couples can also delay the timing of marriage to avoid a marriage penalty, though the estimated effect is small. A more recent paper by Frazier and McKeehan (2018) reconfirms these findings using US data from the mid-1980s to 2011. They estimate that a one percent increase in the size of the marriage penalty relative to a couple’s income raises the probability of later marriages by 1.2 percentage points. I add to the literature by showing that the marriage subsidy affects the timing of separation. I also describe the characteristics of these delayers, showing that these couples are more likely to be over the age of 40 and married for more than five years, which underlines the importance of family planning considerations in marital timing decisions.

My results relate to work on tax incentives and the probability of separation and divorce. Alm and Whittington (1997a) investigate how tax incentives affect the probability of divorce in the US. Using a discrete-time hazard model, they show that married couples react to tax incentives when deciding whether to divorce. Their
results suggest that responses are typically small, though women react more strongly to tax incentives than men. They find that a reduction in the marriage penalty by 50% is associated with a 0.27 percent higher aggregated divorce rate, while they find considerable heterogeneity among different income groups. Taking a wider angle, Dickert-Conlin (1999) investigates the effect of the US tax and transfer systems on the decision to separate. The author’s results are twofold. On the one hand, the effect of transfers on the probability of separation when controlling for the marriage tax alone is not statistically significant. On the other hand, conditional on marriage transfer penalties, the estimates suggest that taxes weakly affect the separation decision. My study provides evidence that, conditional on having decided to separate, the probability of divorce is unaffected by whether a couple may have delayed the date of separation.

2.3 The German Income Tax System and Incentives on Marital Timing Decisions

The German income taxation system allows married couples to benefit from joint taxation through income splitting. Income tax liabilities are calculated by applying the tax code to half the sum of both partners’ taxable incomes and multiplying the resulting amount by two. As the German tax code is progressive, joint taxation leads to liabilities that are never higher than under individual taxation for the same level of income. While the overall tax burden under joint taxation does not depend on the income distribution within a household, the advantage of joint compared to individual taxation increases the more unequal household income is distributed among the spouses. I refer to the difference in income tax liabilities between joint and individual taxation as the marriage subsidy.

Figure 2.1 shows marriage subsidies for exemplary realisations of within-marriage income distributions for the tax year 2010. Although the German tax code has been
reformed several times, the general structure of the income tax code has not changed over the past decades. For a given level of household income, spouses with more unequal earnings have larger marriage subsidies than spouses with more similar incomes. The marriage subsidy is highest if one partner has no earnings while the other partner earns all the taxable household income. The marriage subsidy can increase or decrease with higher taxable household income depending on the within-marriage distribution of taxable income. If a household’s taxable income increases and the within-marriage income distribution is fixed, the marriage subsidy decreases as the marginal tax rates of spouses’ individual incomes become more similar. The splitting advantage disappears when individual incomes are subject to the same marginal tax rate.

Couples are eligible for joint taxation if they are married for at least one day in a calendar year. Recent research has shown evidence consistent with the idea that the timing of marriage responds to the German marriage subsidy (Roeder and Ullmann, 2019; Fink, 2020). Similarly, if a couple decides to end their marriage, the marriage subsidy provides an incentive to postpone the timing of separation into the following calendar year. Importantly, couples have an incentive to postpone the timing of separation and not the timing of divorce. According to German law, the tax-relevant event is the separation and not the date of the legal divorce.

Figure 2.2 presents histograms of the months of separation for married couples in panel (a) and unmarried couples in panel (b). The share of separations occurring in January among married couples is approximately 15%. This proportion is considerably higher than the share of separations in any other month. In contrast, the share of separations in December is considerably smaller at 5.5%. Unmarried couples are not subject to any tax-related incentives regarding their timing decision. Panel (b) suggests that separation timing follows a seasonal pattern as the share of separations occurring in January is also moderately larger compared to other months. However, the contrast is far less prominent than for married couples. The share of
separations in December is close to the average.

The observed bunching of separations in the first quarter and the missing mass of separations in the last quarter for married, but not for unmarried couples, is consistent with spouses postponing their separation date in response to the discrete granting of the marriage subsidy. Although not all separations occurring in the first quarter are tax-motivated, I regard these separations as potentially delayed separations in response to the marriage subsidy.

2.4 Data and Descriptive Statistics

I use income and couple history data drawn from the 34th version of the German Socio-Economic Panel (Goebel et al., 2019). The German Socio-Economic Panel (hereafter: SOEP) is a survey that includes almost 15,000 households and about 30,000 persons.

2.4.1 Sample Selection

In the survey, respondents are asked to provide information on their marital biography (Goebel, 2012). I select a sample of individuals who reported having separated and gave the exact month and year of separation. The German questionnaire uses the ambiguous word “Coupled” (Partnerschaft), which can also include married couples. Therefore, I define a married individual as having separated if they changed their reported marital status from “Married” or “Coupled” to “Married, separated”. I identify a separation of an unmarried individual as the change in marital status from “Coupled” to “Single”. Counting married and unmarried couples, I identify 5,819 individuals who have separated. I drop 24 observations for whom the marital status changed with the partner's death. The SOEP provides an indicator variable for whether a separated, married couple eventually filed for divorce. The variable takes on the value one if the last separation spell of the last marriage ended with a divorce. If the
last separation spell did not end in a divorce, then the marriage is considered and coded as continuing. For these spells, the latest year recorded is the year in which the most recent interview was conducted (Goebel, 2012). Therefore, the indicator is right-censored. 148 married individuals who did not respond to this question are dropped from the sample. Following the relevant literature, I only keep the first separation recorded for each individual in the sample, which drops another 753 observations. Avoiding double counting, I define the unit of observation on the couple level. Therefore, I match each individual with their partner, yielding 2,597 separating couples, of which 1,452 are married and 1,145 are unmarried.

When merging relevant covariates, some couples are dropped due to missing responses. 1,255 individuals have missing data due to non-responses to the questionnaire for income data and the number of children for the relevant year of interest. I drop observations for which at least one individual has missing income data as I am unable to calculate the marriage subsidy for these couples. Of all remaining couples in the sample, an additional 122 individuals have missing data on surveyed education measures and employment status. In my main results, I also drop couples for which at least one individual has missing education measures or employment status. Additionally, I drop 13 same-sex couples as the small sample size does not allow for any heterogeneity analysis of this group. After merging relevant covariates, the final sample counts 1,334 couples. As I can still calculate the marriage subsidy for 61 couples with non-missing income data but missing education measures and employment status, I present results without covariates in column 5 of table 2B.2 in the appendix. The results are very similar to my corresponding baseline results in table 2.2.

### 2.4.2 Calculating the Marriage Subsidy

While the survey includes questions about individuals’ income and marital status, it does not include information on tax liabilities. Therefore, I use a three-step approach to approximate the marriage subsidy of a couple using self-reported income data. I
calculate the marriage subsidy as the difference between the sum of individual tax liabilities and the tax liability under joint taxation.

First, I construct each individual’s taxable income by summing up earnings from primary and secondary employment and income from self-employment. From this pre-tax income, I subtract social security contributions, including mandatory health, unemployment, pension, and care insurance payments, assuming all contributions are entirely deductible. I attribute half of the child allowances to both women and men. The resulting taxable income abstracts from tax rules that only apply to self-employed individuals and income-related exemptions. I also abstract from the solidarity surcharge, which was introduced in 1991, and church taxes, as the data does not include regular surveys about religious denominations. The solidarity surcharge and church taxes are both calculated as a percentage of income tax liabilities. Therefore, I likely underestimate the overall tax liabilities under both individual and joint taxation and, thus, the resulting marriage subsidies.

Second, I apply the tax code to the resulting taxable income for each individual separately. This yields the total tax liability of the household under individual taxation. To calculate the total tax liability of a married couple under joint taxation, I average the taxable incomes of both spouses, apply the tax code to this average income, and multiply the resulting amount by two. Lastly, I compute the marriage subsidy as the difference between the total tax liability of the household under individual taxation and the total tax liability under joint taxation.

---

2 Employee contributions are capped to a varying contribution ceiling, depending on the year and the insurance type. I take these into account when deducting social security contributions from pre-tax earnings.

3 Lusardi and Mitchell (2011) and Bucher-Koenen and Lusardi (2011) show that self-employed individuals are more financially literate and presumably more tax literate as well. Abstracting from tax rules that only apply to self-employed individuals could bias my estimates if the actual marriage subsidies for self-employed were systematically different and self-employed individuals systematically differ in delaying the separation date.

4 In 1991, the solidarity surcharge was introduced to finance the rise in government spending caused by developments in the Middle East, in Southeast and East Europe at the time and the German reunification, according to the bill. The surcharge was first introduced for 12 months only and was set to 7.5% of the income tax liability. In 1993 and 1994, no solidarity surcharge was imposed. Between 1995 and 1997, the surcharge was set to 7.5% before being reduced to 5.5% in 1998. Since 1998, the solidarity surcharge has been stable at 5.5% for the remainder of the years in the sample.

5 Depending on the religion and the state of residency, church taxes amount to 8% to 9% of income tax liabilities.
2.4.3 Descriptive Statistics

Table 2.1 shows descriptive statistics for my sample of 1,334 separating couples. All monetary values are denoted in 2014 euro. About 61% of the couples in my data are married. These couples have an average marriage subsidy of EUR 1,506. A third of all couples in my data separate in the first quarter. Nearly two-thirds of marriages in the sample end in a divorce. For 25% of couples, at least one partner has obtained a college degree. At least one partner is not employed in over a third of couples. Men are more often employed and work close to 45 hours a week on average, while women are less frequently employed and work about ten hours less per week. Households’ average yearly after-tax earnings amount to slightly above EUR 43,000. About one-fifth of couples have children, with the average parenting couple having between one and two children. 21% of the couples in the sample are located in the area of former East Germany.

Figure 2.3 presents the average marriage subsidy by quarter of separation for married couples and the hypothetical marriage subsidy for unmarried couples. The marriage subsidy is higher for married couples who separate in the first quarter than for couples separating in all other quarters of a year. In contrast, there is no difference in the hypothetical marriage subsidies for unmarried couples separating in the first quarter compared to unmarried couples separating in other quarters. Therefore, the figure provides another piece of evidence that some separations occurring in the first quarter of a year are being shifted in response to tax incentives.

2.5 Empirical Strategy

Conditional on the decision to separate, couples can choose to separate immediately or postpone the separation into the following calendar year, thereby benefiting from tax advantages through joint taxation for one more year. I can estimate whether the probability of separating in the first quarter of a year correlates with the marriage
subsidy by estimating the following linear probability model via OLS:

\[ Q_{\text{1Separation}_i} = \alpha + \beta \text{MarriageSubsidy}_i + X_i \gamma + \epsilon_i. \]  

(2.1)

The outcome variable is \( Q_{\text{1Separation}_i} \), which takes on the value of one if the separation of couple \( i \) occurred in the first quarter of a year and zero otherwise. \( \text{MarriageSubsidy}_i \) represents the marriage subsidy defined as the reduction in total tax liability by filing as a married couple compared to the combined tax liability if both spouses filed individually. The parameter of interest is \( \beta \). \( X_i \) is a vector of controls, and \( \epsilon_i \) is an idiosyncratic error term. A similar approach is adopted in Alm and Whittington (1997b), LaLumia et al. (2015) and Fink (2020). For all separations occurring in the first quarter, I use the marriage subsidy for the year before the separation was reported to reflect accurately the expected monetary gain of a delayed separation. All other variables are measured at the time of separation. The results are qualitatively and quantitatively similar when equation 2.1 is estimated as a logit model via maximum likelihood.

After estimating equation 2.1 without any covariates to calculate the unconditional correlation between the timing of separation and the marriage subsidy, I sequentially add additional explanatory variables. I account for time trends and life cycle effects by including a set of separation year dummies and a full set of age dummies for both spouses, respectively. The sample includes separations from 1985 to 2017. Over that time, the tax code has changed in 1981, 1986, 1988, 1990, 1996, 1998, 1999, 2000, 2001, 2002, 2004, 2005, 2007, 2009, 2010, 2013, 2014, 2015, 2016 and 2017. While some adjustments only matched inflation by changing the cut-offs for progressivity brackets, including the personal allowance, many represent a substantial change in marginal tax rates. An example of this is the introduction of an additional tax bracket in 2007. An implicit assumption is that neither mean reversion in income nor changes in inequality are correlated with year-to-year changes in the tax code (Gruber and Saez, 2002). Suppose a couple receives a positive income shock the year before separation. If changes to the tax code were in response to income shocks, then the
marriage subsidy gained through postponing might differ from the expected monetary gain which could lead to measurement error in the marriage subsidy variable and potentially bias in my estimates. However, as income shares of the top 10% and top 1% have remained stable over the sample period, reforms leading to substantial changes in marginal tax rates are likely uncorrelated with such changes in income inequality (Frieden et al., 2023). Age effects likely influence the decision to delay a separation if spouses have not completed family planning but are nearing the limits of fertility (Alm and Whittington, 1997b). Spouses who decide to separate at the end of a year might delay their separation date to after the festive season if they have children. Therefore, I add the number of children as a control. Additionally, I include both spouses’ labour incomes. Figure 2.1 illustrates that the level of taxable household income and the distribution of income within the household are correlated with the marriage subsidy. At the same time, the relationship between household income and the decision to delay might be ambiguous. The level of spouses’ education may also influence the timing decision, which is captured by a control for whether at least one of the spouses has attended higher education. As the marriage subsidy is largest when one partner is not employed, I include a dummy taking on the value one if at least one partner is not employed and zero otherwise. Finally, I add a regional indicator to capture the differences between the East and West of Germany. As my sample spans over three decades, cultural differences between individuals living in either region of Germany might affect the decision to delay separation.

The model does not include any control for financial literacy, which is a concern if financial literacy is correlated with a separation in the first quarter and the marriage subsidy. Although financially literate couples are more likely to know about the advantages of delaying the date of separation to remain eligible for joint taxation for another year, better financial literacy is unlikely to systematically affect the marriage subsidy. The marriage subsidy for a given year is a function of total taxable household earnings and the within-marriage distribution of taxable income. One concern could be that financially literate spouses strategically allocate deductions to their pre-tax
incomes. The German tax code lists more than 530 ways of deducting expenses from pre-tax income (Kirchhof, 2011). Many of these possible deductions are related to education or retirement savings, which are non-transferable between spouses (Doerrenberg et al., 2017). Additionally, most income-related deductions are third-party reported and deducted from a person’s income automatically, leaving hardly any leeway to allocate deductions between spouses, even for financially literate couples.

When estimating regression model 2.1 with all controls included, the parameter of interest, $\beta$, is mostly identified by non-linearities between the marriage subsidy and the income level of both spouses.

2.6 Results

2.6.1 Baseline Results

Table 2.2 presents estimation results of the correlation between the marriage subsidy, a reduction in income tax liabilities through joint taxation, and the probability of separating in the first quarter of a year. Column 3 of table 2B.2 in the appendix shows results from a logit model yielding quantitatively similar results. Couples with higher marriage subsidies are more likely to separate in the first quarter of the following calendar year. Column 1 represents the most parsimonious model, in which the indicator of a first-quarter separation is regressed only on the marriage subsidy. The estimated coefficient is 0.11, suggesting that an increase in the marriage subsidy by EUR 1,000 is associated with a higher probability of a first-quarter separation by 11 percentage points. The effect is statistically significant on all conventional levels. This simple specification computes a correlation where the coefficient is identified by variation in the marriage subsidy arising from the distribution of income between spouses and the household income level, changes in the tax code over time and the number of children.

After adding separation year-fixed effects and age-fixed effects for both spouses in column 2, the estimated coefficient on the marriage subsidy drops to just over 4
percentage points. Age and income likely correlate in a non-linear way. Therefore, the drop in the coefficient might be due to age dummies absorbing contamination from the relationships between age, the marriage subsidy, and considerations of family planning after a separation. In columns 3 to 6, I sequentially add controls for the number of children, both spouses’ income, whether at least one partner has a college degree, whether at least one partner is not employed and regional differences. After including a control for the number of children in the household, the coefficient decreases to about 3 percentage points. For each child, the probability of a separation in the first quarter increases by about ten percentage points. This underlines that children play a key role in the timing decision. A likely explanation is that parents try to postpone the separation after the festive season at the end of the year. The marriage subsidy is correlated with having children through child allowances. Child allowances are deducted from households’ taxable incomes. Large changes in the marriage subsidy are driven by the within-household income distribution rather than the household income level for a given distribution between spouses. Therefore, the change in the coefficient after adding the number of children in the household as a control is limited. In column 4, the estimated effect of the marriage subsidy changes only slightly after including controls for both spouses’ income. The coefficients on either spouse’s income are small and statistically insignificant, as most of the variation in income is likely already captured by the full set of age dummies. Adding controls for whether any partner has obtained a university degree, any partner is not employed, or in which region the couple lives are statistically insignificant and do not change the results. For the richest specification in column 6, an increase in the marriage subsidy by EUR 1,000 is correlated with a 2.9 percentage point higher probability of a first-quarter separation, which is statistically significant on a 5% level. The results are consistent with the hypothesis that couples shift the timing of marital decisions related to separation in response to monetary gains arising from the marriage subsidy.

I find quantitatively and qualitatively similar results when controlling for the number of children either non-parametrically or by a parenting dummy, taking on the value one if the couple has at least one child and zero otherwise. In the former case, the estimated coefficient on the marriage subsidy is 0.0248 (SE: 0.0133), and 0.0250 (SE: 0.0132) in the latter case.
2.6.2 Robustness Checks

Placebo Exercise

The first robustness check is a placebo test estimating the incentive on the timing decision of unmarried couples, who should not be affected by the hypothetical marriage subsidy. Table 2B.1 in the appendix presents results of estimating a regression specification as in equation 2.1 on a sample of unmarried couples. Estimates from the most parsimonious model are shown in column 1. The coefficient is about 0.145 and statistically significant. However, the coefficient decreases to 0.019 and is not statistically significant after including dummies for the year of separation and the age of both spouses. Thus, most of the variation in the hypothetical marriage subsidy is likely driven by the correlation of age with earnings. The estimated coefficient on the hypothetical marriage subsidy remains statistically insignificant on all conventional levels after sequentially including the same set of controls as before. The results from this placebo test support the hypothesis that the marriage subsidy is correlated with shifting the timing of marital decisions related to separation.

Alternative Specifications

Table 2B.2 shows the results of different robustness checks related to non-linear relationships between income and the marriage subsidy, shortcomings of the linear probability model, and time until the turn of the year conditional on couples having made the decision to separate.

Columns 1 and 2 include a 5-piece spline of both partners’ pre-tax labour earnings and a 5-piece spline of total household net income, respectively. The inclusion of flexible controls for income is motivated by a specification in LaLumia et al. (2015) investigating the incentive of tax savings on the timing decision of childbirth. LaLumia et al. (2015) argue that if a tax code produces strong non-linearities in income and tax gains, controlling only linearly for income might result in biased estimates. Suppose there are additional factors unrelated to the marriage subsidy that impact the timing
decision of separation. If these factors are also non-linearly correlated with income, then my estimates in table 2.2 might be biased. Estimates in columns 1 and 2 of table 2B.2 alleviate these concerns as the coefficients are similar to my main results. When splines are included for both spouses’ individual incomes, the estimate is lower than my baseline results and also less precisely estimated. The point estimate for the marriage subsidy is larger and statistically significant on all conventional levels if a spline for households’ net income is included. These robustness checks ease concerns related to non-linear relationships between income and the marriage subsidy, and other income-related factors for separating in the first quarter uncorrelated to the marriage subsidy.

Column 3 of table 2B.2 presents results from estimating the correlation between the marriage subsidy and marital timing decisions related to separation using a logit model estimated via maximum likelihood. Coefficients represent average marginal effects. The point estimate suggests that a EUR 1,000 increase in the marriage subsidy is associated with a 3.2 percentage point higher probability of a first-quarter separation. The effect is statistically significant on all conventional levels and larger compared to the coefficient for the richest specification in the baseline results.

Another concern is that couples who decide to separate early in a year might find it more difficult to postpone the timing of separation conditional on the marriage subsidy. Column 4 probes the robustness of my results by including a linear control for the number of months until January of the following calendar year for all separations occurring outside the first quarter. I also include an interaction term between the linear control and the marriage subsidy. The coefficient on the linear control for the number of months until January of the following year has the expected negative sign and is statistically significant. The further away the turn of the year is, the less likely a couple is to separate in the first quarter. The estimate on the interaction term is precisely estimated but small. The coefficient on the marriage subsidy suggests that
an increase in the marriage subsidy by EUR 1,000 is associated with an increase in the probability of separation in the first quarter by 2.5 percentage points, which is similar to my baseline estimates in table 2.2.

**Bandwidth Checks**

I check the sensitivity of my results by changing the window around the turn of the year. I start by estimating a regression model as equation 2.1 using a subsample of couples separating in December and January only, before gradually including more months around the turn of the year in the sample.

Table 2B.3 shows results from this robustness exercise. If the sample contains only separations occurring in January and December, then the estimated coefficient on the marriage subsidy is about 0.043. The coefficient is larger than the baseline results but imprecisely estimated. One reason for the imprecision might be the smaller sample of 175 couples combined with a full set of spouses’ age and year dummies. In contrast to the baseline results, the estimated effect of having children is statistically insignificant for this sample. The estimated coefficient on the marriage subsidy is close to zero when expanding the sample gradually to include separations between November and February. This particularity relates to previous findings on the timing of marriage in the German context, as most days of mourning fall into November. For example, Fink (2020) shows that the number of marriages occurring in November is lower than in any other month. Similar to my results, the author finds no correlation between the marriage subsidy and marrying in the advantageous quarter. Figure 2.2 illustrates that fewer separations occur in November, suggesting that couples likely avoid November as the month of separation, even if they are not willing to shift the date of separation into the new year. One explanation consistent with a zero-correlation result for this subsample is that couples who report a separation in November are, on average, less able to shift the date of separation, regardless of any outside factors. When including November separations in the sample, the measured correlation decreases as these couples do not change the timing of their
separation in response to the marriage subsidy or holidays of mourning. Therefore, the timing of separation for these couples might be more likely to be driven by non-tax reasons when comparing November separations with separations occurring in the first quarter. When the months of October and March are included in the sample, the coefficient on the marriage subsidy becomes statistically significant at the 10% level. The coefficient is 0.039, which is larger compared to the baseline results, suggesting that the effect is stronger for separations occurring close to the turn of the year. The estimated correlation between the marriage subsidy and a first-quarter separation is the largest for a sample including separations between September and April. The estimate decreases after adding separations reported for August and May.

Overall, this bandwidth analysis indicates that the correlation between the marriage subsidy and separation in the first quarter is stable for different windows around the turn of the year. Compared to the baseline results, the estimated coefficients are larger for separations occurring closer to the turn of the year.

2.6.3 Characteristics of Delayers

My estimation results provide evidence that is consistent with shifting behaviour in the timing of marital decisions related to separation in responses to tax gains from joint taxation. However, not all separations occurring in January are tax-motivated and about 6% of separations of married couples separate in December. Although I am unable to individually identify which couples decided to delay their separation until the following first quarter in response to the marriage subsidy, I can describe the distribution of characteristics for these "delayers". I follow Angrist and Pischke (2009) to calculate delayers’ characteristics ratios for a set of binary characteristics. The ratio is calculated as:

---

7I use the term "delayers" in a similar way as the term compliers in instrumental variables settings. In my paper, delayers are defined as couples who separate in the first quarter of a year to receive the marriage subsidy but would have separated earlier without the marriage subsidy.
\[
\frac{E[Q_1\text{Separation}_i|MS_{i}^{\text{high}} = 1, x_i = 1] - E[Q_1\text{Separation}_i|MS_{i}^{\text{high}} = 0, x_i = 1]}{E[Q_1\text{Separation}_i|MS_{i}^{\text{high}} = 1] - E[Q_1\text{Separation}_i|MS_{i}^{\text{high}} = 0]}.
\] (2.2)

\(x_i\) denotes a characteristic of couple \(i\). \(Q_1\text{Separation}_i\) is a dummy variable referring to couple \(i\)’s actual timing decision, taking on the value one if couple \(i\) separates in the first quarter and zero otherwise. \(MS_{i}^{\text{high}}\) is a dummy variable taking on the value one if couple \(i\)’s marriage subsidy is higher than the median marriage subsidy and zero otherwise. The numerator equals a regression of \(Q_1\text{Separation}_i\) on the dummy variable \(MS_{i}^{\text{high}}\) conditional on the characteristic of interest for couple \(i\) being equal to one, describing the difference in the probability of a first-quarter separation between high and low marriage subsidy couples conditional on couples being of a certain characteristic. The denominator equals an overall regression of \(Q_1\text{Separation}_i\) on the dummy variable \(MS_{i}^{\text{high}}\), representing the difference in the probability of a first-quarter separation between high and low marriage subsidy couples, overall couples. Angrist and Pischke (2009) interpret this ratio as the relative probability that a complying couple is of a certain characteristic.

Table 2.3 presents delayers’ characteristics ratios. Column 1 shows the unconditional mean of each characteristic in the sample of married couples. Column 2 shows the calculated mean of each characteristic conditional on being a delimiter. Column 3 shows the ratio between column 1 and column 2 as calculated in equation 2.2. Overall, nearly 17% of all couples who separate in the first quarter are delayers. Delaying couples are less likely to be parenting, although the ratio is not statistically significant. Together with my baseline estimates, this result implies that although children provide a strong motive for a delayed separation, high marriage subsidy couples who delay separation are not necessarily more likely to have children. This observation might explain the limited effect of including the number of children as a control variable in the baseline results as shown in table 2.2. Delayers are more likely to live in East Germany but are similar to other couples separating in the first quarter regarding higher education and unemployment. While figure 2.1 shows that
the marriage subsidy is larger for households with unequal within-marriage income distributions, women are less likely to contribute less than half to households’ gross income in delaying couples. This result suggests potential bargaining between spouses about the timing decision. Delaying couples are slightly more likely to be married for more than five years. Lastly, spouses who delay their date of separation are more likely to be over the age of 40, which indicates that these couples have likely finished family planning at the time of separation.

2.7 Are Delayers Less Likely to Divorce?

Conditional on having decided to separate, married couples can delay their separation into the first quarter of the following calendar year to receive the marriage subsidy for one more year. Delaying the timing of separation requires a couple to coordinate, which could be a first step towards spouses reconciling. In this section, I investigate whether couples separating in the first quarter of a year are less likely to divorce. My findings suggest that the timing of separation does not impact the probability of divorce, whereas higher labour earnings of women are associated with divorces occurring sooner after separation.

2.7.1 Divorce Laws in Germany

Married couples must live separately for at least 12 months before filing for divorce in Germany. This grace period serves at least two purposes. Firstly, living separately for an extended period acts as an irrefutable signal that the marriage is disrupted and requires to be dissolved. Secondly, lawmakers try to ensure that marriages are not prematurely dissolved by giving plenty of time for reconciliation. Marriages are divorced based on a disruption principle, meaning marriage can fail without one spouse being responsible.\(^8\) My estimation results support the hypothesis that married

\(^8\)In the 1970s, German divorce laws changed substantially. Before, spouses used to be divorced based on a fault principle (Schuldprinzip), meaning that a marriage was regarded as failed if one spouse disrupted the marriage. The spouse blamed for the disruption would usually face legal consequences and obligations. The legislation introduced during the 1970s invokes a disruption
couples delay the timing of separation in response to the marriage subsidy. Thus, spouses seem willing to extend the 12-month waiting time before legal divorce by deferring the timing of separation, conditional on having decided to separate.

Before showing estimation results, I establish some facts about the distribution of month of divorce in figure 2.4 for Germany. I also show the distribution for Austria, Denmark and The Netherlands as these countries are geographically close with a culture similar to Germany for comparison. None of these countries has a separation period or marriage subsidy comparable to the German model. Divorces are nearly uniformly distributed throughout the year in Austria, Denmark and The Netherlands, whereas more than 25% of divorces in Germany occur in February and March. The share of divorces in January is about 7%. The reason why the share of divorces in January is relatively low, although figure 2.2 shows a large share of separations occurring in January, lies in the timing of events between separation and divorce. Spouses must have lived separately for at least 12 months before the day of the final court hearing. Lawyers can apply to start the legal divorce process up to eight weeks before these 12 months have passed. The process usually lasts four to six months or, depending on the complexity of the divorce case, up to 12 months. Therefore, a couple separating in January can apply to start the divorce process in November of the same year. The marriage will then be divorced between February and April, leading to a large share of divorces occurring in February and March.

2.7.2 Probability of Divorce for Delayers and Non-Delayers

I estimate a Cox proportional hazards model (Cox, 1972) to investigate whether couples separating in the first quarter of a year are less likely to divorce than couples separating in other months. The Cox proportional hazard model of divorce after separation is given by principle (Zerrütungsprinzip). The disruption principle is based on the idea that marriage can fail without one spouse being responsible, as described in §1565 Abs.1 S1. BGB in the German law text. This legislation is still in place today. The shift in divorce law represented the idea that neither courts nor the spouses themselves were able to identify the causes for the marriage disruption or to which extent each spouse contributed to them (Schlüter, 2009).
where $x$ is a set of controls measured in the year of separation analogous to those included in the baseline results in table 2.2. $h_0(t)$ denotes the baseline hazard and is a function of time in months $t$ between the reported start and end of the separation. I define failure as a divorce after separating. The controls in the model affect the relative risk of divorce but not the shape of the overall hazard function. A change in a regressor leads to a change in the hazard rate by

$$\frac{\partial h(t|x, \beta)}{\partial x_j} = \beta_j h(t|x, \beta),$$

which does not require knowledge of $h_0(t)$. I use the Breslow method for handling tied failures. I assume that the hazard ratio is constant over time such that at any point along the time dimension, the hazard of divorce is constant between couples that separate in the first quarter to those separating in other months. Transition times are grouped where the hazard within the month is constant. The separation data is right-censored as some couples were still married for the last recorded year. 401 couples have divorced and are non-censored, while 282 are still married and are right-censored. 136 couples are known to divorce, but the exact month is not recorded. Excluding these 136 couples from the analysis yields quantitatively similar results.

Figure 2B.2 shows descriptive evidence for whether couples separating in the first quarter are less likely to divorce. Panel (a) in figure 2B.2 shows the Kaplan-Meier estimate of the survival function, and panel (b) shows the Nelson-Aalen estimate of the cumulative hazard function. Both figures suggest that couples separating in the first quarter are slightly more likely to divorce between 22 and 40 months after separating. Overall, the probability of divorce between couples separating in the first quarter and those separating in different months does not appear to be systematically different.

Table 2.4 presents coefficients of the Cox proportional hazard model and post-estimation diagnostics. The coefficient on whether a couple separated in the first
quarter of a year is small and statistically insignificant. Only the coefficient on wives’ labour earnings is statistically significant on all conventional levels. The results suggest that the timing of separation is uncorrelated with the timing and the probability of divorce. Only higher earnings of women are correlated with swifter divorces after separation. Table 2.4 also provides post-estimation checks of zero slope to test whether the log hazard-ratio function is constant over time. The results indicate no violation of the proportional-hazards assumption.

2.8 Discussion and Conclusion

I investigate whether income tax has an incentive on the timing of separation in the context of the German income tax system. Pecuniary gains arise by allowing married couples to file for joint taxation. This tax benefit, the "marriage subsidy", is granted if a couple is married for at least one day in a calendar year.

Using the German Socio-Economic Panel, a high-quality survey containing detailed information on individual characteristics and relationship histories, I estimate that an increase in the marriage subsidy by EUR 1,000 is associated with a 2.9 percentage point higher probability of separating in the first quarter of the following calendar year. The results are robust to a series of robustness checks. The estimate is similar to the correlation measured between the marriage subsidy and the timing of marriage, which is about 3.1 percentage points (Fink, 2020). I show that delaying couples are more likely to be over 40 years old and are slightly more likely to be married for more than five years, suggesting that delaying couples have completed family planning. Women are less likely to contribute less than half of the households’ gross income in delaying couples. As coordinating on the timing of separation is a potential first step in reconciling a relationship, I estimate a Cox proportional hazards model of the probability of divorce, finding that couples who separate in the first quarter of a year are equally likely to divorce at various time horizons as couples separating in other months. This is a surprising finding as agreeing on an advantageous timing of
separation requires some cooperative behaviour, which could be seen as a first step in reconciling the marriage.

Based on my results, one avenue to explore further is to compare spouses’ expected utility change after separation with realised changes, as suggested by Friedberg and Stern (2014). Asking an individual about their presumed level of happiness after separation and their guess about the partner’s suspected level of happiness after separation, linked with observed outcomes after separation, could give rise to bargaining within the marriage about the timing of separation. Indicative of these ideas are my additional results in appendix 2.8 and figure 2B.3. Tracing out the trajectory of overall life satisfaction after separation suggests that men fully recover after three years, regardless of the timing of separation. Interestingly, women who separate in the first quarter need two more years to fully recover from separation compared to women who separate in other months. This could suggest bargaining within marriages about the timing of separation.

My results provide convincing support for the hypothesis that tax incentives influence the timing of separation. As such, my results are relevant from a policy perspective. Firstly, this tax rule is a sizeable yearly expense. About 150,000 divorces are recorded in Germany every year. Taking my descriptive statistics at face value implies that roughly \( \frac{150,000}{0.656} \approx 229.000 \) couples separate each year. One-third, or over 75,000, of these couples separate in the first quarter. Based on my calculations following Angrist and Pischke (2009), nearly 17% of couples who separate in the first quarter are delayers. Thus, there are approximately 12,750 couples who delay their separation in response to the tax rule. These couples delay their separation to the first quarter of the following calendar year in response to the tax rule but would have separated in the previous calendar year otherwise. As the average marriage subsidy is EUR 1,506, a back-of-the-envelope calculation would suggest that the German government spends EUR \( 1,506 \times 12,750 \approx EUR \)
19,201,500 yearly in 2014 prices on this tax rule as foregone income tax revenue. As this policy has been in place for over thirty years, total spending amounts to well above half a billion euro. Secondly, future fertility considerations and the marriage subsidy steer the timing decision of separation in opposite directions. Therefore, the tax code might have unexpected regressive effects if younger couples with lower incomes are less likely to receive the monetary benefit due to future plans about fertility.

Discretely granting the marriage subsidy does not encourage the continuation of marriage as the probability of divorce remains unaffected by the timing decision of separation. If the goal of policymakers is to encourage the reconciliation of marriages, then they would need to search for different incentives. The tax code, in its current format, might have net negative welfare implications related to delayed separation, such as an increased risk of domestic violence or opportunity costs of foregone income tax revenue. My results do not allow for any conclusion on whether the government should provide a marital tax break altogether. However, one recommendation for policymakers derived from my results could be to grant the marriage subsidy on a pro-rata basis. For example, if a couple separates in January, then 1/12 of the expected marriage subsidy will be granted. Although this rule still bears an incentive to delay separation until the next month for any month, the associated negative welfare implications are much smaller while retaining the current status quo of allowing for a tax break for married and not separated couples.
References - Chapter 2


Figure 2.1: Marriage Subsidy for Exemplary Within-Marriage Distributions of Taxable Household Income

Note: The figure shows marriage subsidies for exemplary realisations of within-marriage income distributions for the tax year 2010. The horizontal axis represents the total taxable income of a household measured in EUR 1,000. Given the same household income level, couples with one partner earning all income (red, solid) have the largest gain from joint taxation. Comparing couples where one partner earns 80 percent, while the other earns 20 percent of taxable household income (blue, dotted), the marriage subsidy is higher than for a couple with a 60 percent, 40 percent split (green, dash-dot). Depending on the within-marriage distribution of taxable income, the marriage subsidy can increase and decrease with higher levels of taxable household income. Humps occur when spouses’ incomes are taxed differently at the margin. As total taxable income increases, the difference in marginal tax rates becomes smaller. Eventually, both incomes will be taxed subject to the same marginal tax rate, so the tax advantage disappears. As total taxable income increases further, one income enters the next higher tax bracket so that the marriage subsidy increases again.
Figure 2.2: Distribution of Month of Separation by Marital Status

(a) Married Couples

(b) Unmarried Couples

Note: The figure shows histograms of the month of separation for married couples in panel (a) and unmarried couples in panel (b) based on all 2,597 couples that I identify in the SOEP as having separated. The bars are scaled such that the height of each bar can be interpreted as a percentage. The unit of observation is on the couple level. Of all 2,597 identified couples that report a separation, 1,452 are married, and 1,145 are unmarried. The reported month and year of separation are taken from the questionnaire of the male partner in the relationship. The results do not change when the reported dates are taken from the female partner's questionnaire.
Figure 2.3: Marriage Subsidy by Quarter of Separation

Note: The figure shows the average marriage subsidy by quarter of separation for married couples (circles, red) and the average hypothetical marriage subsidy for unmarried couples (triangles, blue) using the sample of 819 married couples, and 515 unmarried couples based on the SOEP. The statistics at Q2-Q4 Pooled represent the average marriage subsidy and hypothetical marriage subsidy across separations occurring in the second, third and fourth quarter of a year for married and unmarried couples, respectively. Vertical lines represent 95% confidence intervals of the mean. All monetary values are denoted in 2014 euro.
Figure 2.4: Distribution of Month of Divorce by Country

(a) Germany

(b) Austria

(c) Denmark

(d) The Netherlands

Note: The figure shows the distribution of month of divorce for Germany, Austria, Denmark and The Netherlands. While Germany is geographically close to these countries, none has a separation period or a marriage subsidy similar to the German model. The data for each country includes years between 2005 and 2018. The data for Germany is based on the SOEP. Data for Austria, Denmark and The Netherlands are provided by Statistis Austria, Statistics Denmark and Statistics Netherlands, respectively.
Table 2.1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Separation in First Quarter</td>
<td>1,334</td>
<td>0.347</td>
<td>0.476</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Explanatory Variable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marriage Subsidy</td>
<td>819</td>
<td>1,506.52</td>
<td>1,815.27</td>
<td>0</td>
<td>13,611</td>
</tr>
<tr>
<td><strong>Couple Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1,334</td>
<td>0.614</td>
<td>0.487</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Marriage Ends in Divorce</td>
<td>819</td>
<td>0.656</td>
<td>0.475</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Any Partner has College Degree</td>
<td>1,334</td>
<td>0.244</td>
<td>0.4299</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Any Partner Not Employed</td>
<td>1,334</td>
<td>0.362</td>
<td>0.481</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HH Pre-Tax Labour Income</td>
<td>1,334</td>
<td>49.79</td>
<td>38.08</td>
<td>0</td>
<td>530.51</td>
</tr>
<tr>
<td>HH Net Labour Income</td>
<td>1,334</td>
<td>43.08</td>
<td>28.04</td>
<td>0</td>
<td>334.28</td>
</tr>
<tr>
<td>Reported Working Hours Men</td>
<td>1,072</td>
<td>44.51</td>
<td>10.20</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>Reported Working Hours Women</td>
<td>932</td>
<td>34.75</td>
<td>12.64</td>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>Age Women</td>
<td>1,334</td>
<td>35.74</td>
<td>10.551</td>
<td>18</td>
<td>89</td>
</tr>
<tr>
<td>Age Men</td>
<td>1,334</td>
<td>38.72</td>
<td>10.902</td>
<td>18</td>
<td>93</td>
</tr>
<tr>
<td>Parenting</td>
<td>1,334</td>
<td>0.218</td>
<td>0.413</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of Children</td>
<td>291</td>
<td>1.68</td>
<td>0.923</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>East</td>
<td>1,334</td>
<td>0.216</td>
<td>0.412</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note:* The table presents descriptive statistics for my sample based on the SOEP. The data contains only female-male couples, as the sample size is too small to allow for any inference about other couple formations. *Obs* refers to the number of couples in the sample. *Separation in First Quarter* is a dummy variable that takes on the value one if the couple separated in the first quarter of a year, and zero otherwise. *Marriage Subsidy* and *Marriage Ends in Divorce* are only recorded for married couples. *Any Partner has College Degree* and *Any Partner Unemployment* are dummy variables that take on the value one if at least one partner has a college degree or is not employed, respectively. *Reported Working Hours Men* and *Reported Working Hours Women* are based on reported working hours. *Number of Children* is only recorded for couples that have children. All monetary values are measured in 2014 prices.
<table>
<thead>
<tr>
<th>Dependent Variable: ( Q_1 \text{Separation}_i ), if couple ( i )'s separation occurred in Q1 and 0, otherwise.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marriage Subsidy</td>
<td>0.1107</td>
<td>0.0406</td>
<td>0.0303</td>
<td>0.0272</td>
<td>0.0279</td>
<td>0.0292</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.0106)</td>
<td>(0.0107)</td>
<td>(0.0126)</td>
<td>(0.0128)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Number of Children</td>
<td>0.1043</td>
<td>0.1033</td>
<td>0.1027</td>
<td>0.1012</td>
<td>0.1012</td>
<td>0.1012</td>
</tr>
<tr>
<td></td>
<td>(0.0231)</td>
<td>(0.0232)</td>
<td>(0.0233)</td>
<td>(0.0233)</td>
<td>(0.0233)</td>
<td></td>
</tr>
<tr>
<td>Husband's Labour</td>
<td>0.0011</td>
<td>-0.0008</td>
<td>-0.0004</td>
<td>0.0011</td>
<td>0.0008</td>
<td>0.0004</td>
</tr>
<tr>
<td>Earnings</td>
<td>(0.0065)</td>
<td>(0.0066)</td>
<td>(0.0066)</td>
<td>(0.0065)</td>
<td>(0.0066)</td>
<td>(0.0066)</td>
</tr>
<tr>
<td>Wife's Labour</td>
<td>-0.0094</td>
<td>-0.0139</td>
<td>-0.0142</td>
<td>-0.0094</td>
<td>-0.0139</td>
<td>-0.0142</td>
</tr>
<tr>
<td>Earnings</td>
<td>(0.0098)</td>
<td>(0.0107)</td>
<td>(0.0107)</td>
<td>(0.0098)</td>
<td>(0.0107)</td>
<td>(0.0107)</td>
</tr>
<tr>
<td>Any Partner has</td>
<td>0.0318</td>
<td>0.0301</td>
<td>0.0318</td>
<td>0.0318</td>
<td>0.0301</td>
<td>0.0318</td>
</tr>
<tr>
<td>College Degree</td>
<td>0.0441</td>
<td>0.0441</td>
<td>0.0441</td>
<td>0.0441</td>
<td>0.0441</td>
<td>0.0441</td>
</tr>
<tr>
<td>Any Partner</td>
<td>-0.0325</td>
<td>-0.0353</td>
<td>-0.0325</td>
<td>-0.0325</td>
<td>-0.0353</td>
<td>-0.0325</td>
</tr>
<tr>
<td>Not Employed</td>
<td>(0.0415)</td>
<td>(0.0415)</td>
<td>(0.0415)</td>
<td>(0.0415)</td>
<td>(0.0415)</td>
<td>(0.0415)</td>
</tr>
<tr>
<td>East</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0462)</td>
<td></td>
</tr>
<tr>
<td>Year Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Age Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.2</td>
<td>0.472</td>
<td>0.493</td>
<td>0.493</td>
<td>0.494</td>
<td>0.494</td>
</tr>
<tr>
<td>Observations</td>
<td>819</td>
<td>819</td>
<td>819</td>
<td>819</td>
<td>819</td>
<td>819</td>
</tr>
</tbody>
</table>

**Note:** The table shows the estimated correlation between separation in the first quarter of the next calendar year and the marriage subsidy. Coefficients are estimated via OLS using my sample of married couples based on the SOEP. The marriage subsidy is measured in EUR 1,000 for the year before the separation was reported for married couples separating in the first quarter of a year to reflect accurately the expected monetary gain of a delayed separation. All other variables are measured at the time of separation. Column 1 shows results from a specification regressing the indicator variable for first-quarter separations only on the marriage subsidy. Year-fixed effects and age-fixed effects of both spouses are added in column 2. In column 3, I add the number of children living in the household. Column 4 also includes controls for spouses’ labour incomes measured in EUR 10,000. In column 5, I add controls for whether at least one partner has a college degree and if at least one partner is not employed. Column 6 controls for regional differences. All monetary values are denoted in 2014 euro. Standard errors in parentheses are robust to heteroskedasticity.
Table 2.3: Complier-Characteristics Ratios for High and Low Marriage Subsidies

| Characteristic                              | \( E(x) \) (1) | \( E(x|D_1 > D_0) \) (2) | \( \frac{E(x|D_1 > D_0)}{E(x)} \) (3) |
|--------------------------------------------|----------------|--------------------------|----------------------------------|
| Parenting                                  | 0.271          | 0.193                    | 0.711 (0.499)                    |
| East                                       | 0.188          | 0.239                    | 1.27 (0.622)                     |
| Any Partner has College Degree             | 0.259          | 0.184                    | 0.709 (0.473)                    |
| Any Partner Not Employed                   | 0.364          | 0.158                    | 0.433 (0.391)                    |
| Wife's Earnings Contribute Less Than 50% to HH Gross Income | 0.742          | 0.615                    | 0.828 (0.171)                    |
| Marriage Lasted Longer Than 5 Years         | 0.755          | 0.763                    | 1.011 (0.159)                    |
| Husband Older Than 40                      | 0.529          | 1.001                    | 1.893 (0.371)                    |
| Wife Older Than 40                         | 0.394          | 0.714                    | 1.811 (0.41)                     |

Note: The table shows delayers’ characteristics ratios following Angrist and Pischke (2009) using my sample of married couples based on the SOEP. All characteristics are binary. I use the term "delayers" analogously to the term "compliers" in instrumental variables settings. In my paper, delayers are defined as couples that separate in the first quarter of a year to receive the marriage subsidy but would have separated earlier without the marriage subsidy. Column 1 shows the unconditional mean of each characteristic in the sample of married couples. Column 2 shows the calculated mean of each characteristic conditional on being a delayer. Column 3 shows the ratio between column 1 and column 2. The ratio in column 3 is calculated as:

\[
\frac{P[x_i = 1|Q_{1\text{Separation}} > Q_{1\text{Separation}}_0]}{P[x_i = 1]} = \frac{E[Q_{1\text{Separation}}|MS_{i}^{\text{high}} = 1, x_i = 1] - E[Q_{1\text{Separation}}|MS_{i}^{\text{high}} = 0, x_i = 1]}{E[Q_{1\text{Separation}}|MS_{i}^{\text{high}} = 1] - E[Q_{1\text{Separation}}|MS_{i}^{\text{high}} = 0]},
\]

where \( x_i \) denotes a characteristic of couple \( i \). \( Q_{1\text{Separation}} \) and \( Q_{1\text{Separation}}_0 \) refer to couple \( i \)’s timing decision under a higher-than-median marriage subsidy and a lower-than-median marriage subsidy, respectively. In the table, I denote these by \( D_1 \) and \( D_0 \), respectively. \( MS_{i}^{\text{high}} \) is a dummy variable referring to couple \( i \)’s actual timing decision, taking on the value one if couple \( i \) separates in the first quarter and zero otherwise. Standard errors are in parenthesis. The sample size is 819 for all characteristics.
Table 2.4: Probability of Divorce - Cox Proportional Hazard Model

<table>
<thead>
<tr>
<th></th>
<th>Coefficient (1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separation in First Quarter</td>
<td>0.02 (0.117)</td>
<td>-0.048</td>
<td>2.45</td>
<td>0.117</td>
</tr>
<tr>
<td>Marriage Subsidy</td>
<td>0.009 (0.039)</td>
<td>-0.001</td>
<td>0.00</td>
<td>0.982</td>
</tr>
<tr>
<td>Husband's Labour Earnings</td>
<td>-0.023 (0.016)</td>
<td>-0.018</td>
<td>0.25</td>
<td>0.619</td>
</tr>
<tr>
<td>Wife's Labour Earnings</td>
<td>0.126 (0.036)</td>
<td>-0.014</td>
<td>0.21</td>
<td>0.649</td>
</tr>
<tr>
<td>Number of Children</td>
<td>-0.054 (0.071)</td>
<td>0.028</td>
<td>1.03</td>
<td>0.309</td>
</tr>
<tr>
<td>East</td>
<td>-0.11 (0.128)</td>
<td>0.013</td>
<td>0.13</td>
<td>0.718</td>
</tr>
<tr>
<td>Any Partner has College Degree</td>
<td>-0.198 (0.123)</td>
<td>0.052</td>
<td>2.36</td>
<td>0.124</td>
</tr>
<tr>
<td>Any Partner Not Employed</td>
<td>0.061 (0.117)</td>
<td>-0.034</td>
<td>1.13</td>
<td>0.287</td>
</tr>
<tr>
<td>Year Dummies</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Dummies</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Test</td>
<td>116.49</td>
<td>126</td>
<td>0.717</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>819</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Failures/Divorces</td>
<td>537</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time at Risk</td>
<td>27,190</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Pseudo-Likelihood</td>
<td>-2988.78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table shows the results of a Cox proportional hazard model in column 1 and post-estimation diagnostics in columns 2-4. Column 1 shows regression coefficients. The proportional-hazards assumption implies that the estimated coefficient from the Cox model is the same for all failure time $t_j$. Under this assumption, the sum of the scaled Schoenfeld residual at failure time $t_j$ and the estimated coefficient from the Cox model do not change over $t_j$. Column 2 presents the slope from a linear fit line of the scaled Schoenfeld residual sum plus the Cox model's estimated coefficient over time. Column 3 shows the corresponding $\chi^2$ test statistic. Column 4 presents the p-value. The marriage subsidy is measured in EUR 1,000 for the year before the separation was reported for married couples separating in the first quarter to reflect accurately the expected monetary gain of a delayed separation. Spouses' labour incomes are measured in EUR 10,000. All monetary values are denoted in 2014 euro. Standard errors in parentheses are robust to heteroskedasticity.
Appendix 2A: Additional Results: The Effect of Separation on Overall Life Satisfaction

My results suggest that the timing of separation is subject to outside factors, such as incentives from the income tax system. Couples face a trade-off between unhappiness from continuing to live together and receiving the marriage subsidy. This unhappiness is likely heterogeneous across individuals and potentially correlated with the impact of separation on overall life satisfaction. The marginal couple who shifts their separation date is the one that is indifferent between the potential gain from the marriage subsidy and the incurred unhappiness due to postponing the separation date. Conditional on the marriage subsidy, the timing decision becomes a function of the impact of separation on overall life satisfaction before and after the separation occurs. My results of exploring the trajectories of overall life satisfaction around the time of separation show that men fully recover three years after separation, regardless of the timing of separation. Women who separate in the first quarter need two more years to fully recover from separation.

I estimate the trajectories of overall life satisfaction around the time of separation through a series of event-studies of the form:

\[ y_{i,t} = \sum_{e=-5}^{5} \beta_e 1[e = t - s] + \sum_{e=-5}^{5} \delta_e 1[e = t - s] \cdot Q1\text{Separation}_i + \beta_{\text{Marriage Subsidy}} y_{i,t} + \alpha_{age} + \tau_t + \eta_{i,t} \]  

(2A.1)

In this event-study model, the parameters of interest are the coefficients \( \beta_{-5} \) to \( \beta_5 \) and \( \delta_{-5} \) to \( \delta_5 \). The former coefficients measure the effect of separation on overall life satisfaction \( y_{i,t} \) relative to the year of separation \( s \). The latter set of coefficients allows for the effect to be different between individuals separating in the first quarter (\( Q1\text{Separation}_i = 1 \)) and those separating in other months (\( Q1\text{Separation}_i = 0 \)).

\(^9\)This idea is similar to concepts in the "Bunching" literature (Kleven, 2016), where a couple wants to maximise after-tax income or consumption but incurs costs or disutility from exerting effort. This disutility usually depends on an ability parameter leading to heterogeneity in disutility across couples.
I exclude the event-time dummy for the year before the separation such that the event-time coefficients measure the effect on overall life satisfaction relative to the year before the separation. Overall life satisfaction is measured on a zero to ten scale, where ten is the highest. I include a linear control for the marriage subsidy, a full set of age dummies to control for life cycle effects and a full set of year dummies to control for time trends common to all individuals. $\eta_{i,t}$ is an idiosyncratic error term. I estimate equation 2A.1 pooled and separately for husbands and wives. Similar to the regression models in Leopold and Kalmijn (2016), equation 2A.1 can track short-term and long-term changes in overall life satisfaction along the separation process.

Figure 2B.3 presents event-studies for the effect of separation on overall life satisfaction. All results should be interpreted as correlations between separation and overall life satisfaction, as the timing of separation is not random. Panel (a) shows results from a pooled sample of men and women. The event-time dummies for pre-separation years are statistically insignificant for individuals separating in the first quarter of the year. This suggests that their level of overall life satisfaction leading up to separation is unaffected. The largest estimated effect occurs one year after separation with a coefficient of -0.73, which is precisely estimated. The negative impact of separation fades out after four years. However, the event-study plot for individuals separating in other months of the year differs in pre-trends and post-event trajectories. The estimated pre-event coefficients are positive and statistically significant, suggesting that overall life satisfaction levels decrease before the separation happens. The largest estimated effect also occurs one year after the event, with a point estimate of -0.5. Individuals separating in other months of the year appear to recover more quickly from separation, as the estimated effect is statistically insignificant after two years. The confidence intervals overlap for all estimated coefficients so that the differences between timing groups are not statistically significant.

Panel (b) and (c) of figure 2B.3 show event-study plots for a sample of only respondents in the SOEP being asked: “In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered?”.

---

10Data for this variable is derived from respondents in the SOEP being asked: “In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered?”.
husbands and wives, respectively. Overall life satisfaction for men separating in other months of the year exhibits a statistically significant pre-trend three or more years before separation. The coefficient for two years before separation is statistically insignificant. These pre-trends indicate bargaining occurring before separation. There is no such pre-trend for men separating in the first quarter. Post-separation trajectories are similar across timing groups. As before, confidence intervals overlap for all estimated coefficients. Overall life satisfaction of women separating in other months also exhibits a statistically significant pre-trend for years before separation, in contrast to women separating in the first quarter of a year. While women who separate in the first quarter fully recover after four years, women separating in other months recover two years earlier. The difference in the effect of separation between the timing groups is statistically significant for the third year after separation.

Overall, the results suggest that husbands do not react differently to a separation regardless of the timing, and the effects are short-lived. Wives who separate in other quarters of the year recover more swiftly compared to wives who separate in the first quarter. Nevertheless, the differences between the timing groups are mostly statistically insignificant. If couples separate in other quarters of a year, both spouses appear to be on a negative trajectory before the separation happens.

The transitory nature of the effects corresponds to predictions derived from the crisis model (Johnson and Wu, 2002). The crisis model builds on the assumption that separation and divorce are emotionally troubling. Both events cause stress due to adjusting to a new housing situation, living under different financial constraints, dividing consumption goods, and reorganising social interactions. As the stress caused by these adjustments is temporary, the negative effect of separation and divorce on overall life satisfaction is short-lived (Stroebe et al., 2007).

Figure 2B.4 presents additional event-study plots for labour earnings. The pre-event coefficients for the effect on husbands’ labour earnings are nearly all statistically...
insignificant and close to zero for both timing groups. However, earnings of husbands separating in the first quarter drop in the second year after the event but recover after three years. For husbands separating in other months, the estimated effect on earnings is persistent, suggesting that earnings are nearly EUR 4,500 lower five years after the event. Women’s earnings increase immediately after separation. After that, the effect is imprecisely measured. Figure 2B.5 shows additional event-study plots for reported hours worked. Men do not appear to adjust their labour supply regardless of the timing of separation. Mirroring the results in labour earnings, women increase their weekly hours worked by about three hours in both timing groups.
**Figure 2B.1: Declaration of Permanently Living Separately**

### Erklärung zum dauernden Getrenntleben

<table>
<thead>
<tr>
<th>Eingangsstempel</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Steuernummer</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Steuernummer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifikationsnummer</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Identifikationsnummer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>An das Finanzamt</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] An das Finanzamt</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wohntwese: bisheriges Finanzamt</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Wohntwese: bisheriges Finanzamt</td>
</tr>
</tbody>
</table>

### Angaben zur Person

<table>
<thead>
<tr>
<th>Antragstellende Person / Name</th>
<th>Eingangsstempel</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Adresse</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Adresse</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wohnort</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Wohnort</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rufnummer</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Rufnummer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Steuerleitstelle</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Steuerleitstelle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tag der Trennung</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Tag der Trennung</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Name</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ehegatte/Lebenspartner</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Ehegatte/Lebenspartner</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geburtsdatum</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] Geburtsdatum</td>
</tr>
</tbody>
</table>

### Ein dauerndes Getrenntleben ist anzunehmen, wenn die zum Wesen der Ehe/Lebenspartnerschaft gehörige Lebens- und Wirtschaftsgemeinschaft nach dem Gesamtbild der Verhältnisse auf die Dauer nicht mehr besteht. Dabei ist unter Lebensgemeinschaft die räumliche, persönliche und geistige Gemeinschaft der Ehegatten/Lebenspartner, unter Wirtschaftsgemeinschaft die gemeinsame Erledigung der die Ehegatten/Lebenspartner gemeinsam berührenden wirtschaftlichen Fragen ihres Zusammenlebens zu verstehen.

Ich bestätige, dass die vorstehenden Voraussetzungen des dauernd Getrenntlebens in meiner Ehe/Lebenspartnerschaft vorliegen. Ein Getrenntleben aus anderen Gründen (z.B. räumliche Trennung wegen verschiedener Arbeitsorte oder wegen Fehlens einer gemeinsamen Wohnung) liegt nicht vor.

### Datenschutzhinweis:

Informationen über die Verarbeitung personenbezogener Daten in der Steuerverwaltung und über Ihre Rechte in der Datenschutz-Grundverordnung sowie über Ihre Ansprechpartner in Datenschutzfragen finden Sie unter www.finanzamt.de (unter der Rubrik „Datenschutz“) oder erhalten Sie bei Ihrem Finanzamt.

### Note:

The figure shows the declaration of permanently living separately (Erklärung zum dauernden Getrenntleben). The form is filled out and submitted to the tax office through the official digital tax administration system (ELSTER) via the website https://www.elster.de/eportal/formulare-leistungen/alleformulare/elevgetrenntlebend. Eligibility for joint taxation is based on the date declared on this document.
Figure 2B.2: Survivor and Cumulative Hazard Function Estimates by Quarter of Separation

(a) Kaplan-Meier Estimate of Survival Function

(b) Nelson-Aalen Estimate of Cumulative Hazard Function

Note: The figure shows the Kaplan-Meier estimate of the survival function in panel (a) and the Nelson-Aalen estimate of the cumulative hazard function in panel (b) for my sample of 819 married couples based on the SOEP. The red, solid line represents married couples separating in the first quarter. The blue, dotted line represents couples separating in other months.
Figure 2B.3: Event-Study Results of Overall Life Satisfaction

(a) Pooled

(b) Husband

(c) Wife

Note: The figure presents event-study plots with estimates from equation 2A.1 with overall life satisfaction as the outcome. Overall life satisfaction is recorded on a 0-10 scale, where 10 is the highest level of satisfaction and 0 is the lowest. Respondents in the SOEP are asked: “In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered?”. The marriage subsidy is measured in EUR 1,000 for the year before the separation was reported for married couples separating in the first quarter of a year to reflect accurately the expected monetary gain of a delayed separation. The longitudinal sample is based on the sample described in table 2.1 and consists of 775 husbands and 794 wives. Each individual is observed at least once before and once after separation and at least 5 times in the event window. Vertical lines represent 95% confidence intervals. All monetary values are denoted in 2014 euro. Standard errors are clustered at the individual level.
Note: The figure presents event-study plots with estimates from equation 2A.1 with individual labour earnings as the outcome. The marriage subsidy is measured in EUR 1,000 for the year before the separation was reported for married couples separating in the first quarter of a year to reflect accurately the expected monetary gain of a delayed separation. Labour incomes are measured in EUR 10,000. The longitudinal sample is based on the sample described in table 2.1 and consists of 775 husbands and 794 wives. Each individual is observed at least once before and once after separation and at least 5 times in the event window. Vertical lines represent 95% confidence intervals. All monetary values are denoted in 2014 euro. Standard errors are clustered at the individual level.
Note: The figure presents event-study plots with estimates from equation 2A.1 with reported weekly hours worked as the outcome. The marriage subsidy is measured in EUR 1,000 for the year before the separation was reported for married couples separating in the first quarter of a year to reflect accurately the expected monetary gain of a delayed separation. The longitudinal sample is based on the sample described in table 2.1 and consists of 775 husbands and 794 wives. Each individual is observed at least once before and once after separation and at least 5 times in the event window. Vertical lines represent 95% confidence intervals. All monetary values are denoted in 2014 euro. Standard errors are clustered at the individual level.
Table 2B.1: Placebo Results - The Incentive of the Hypothetical Marriage Subsidy of Unmarried Couples on Postponing Separation Until the First Quarter of a Year

<table>
<thead>
<tr>
<th>Dependent Variable: $Q_1Separation_i = 1$, if couple $i$’s separation occurred in Q1 and 0, otherwise.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical Marriage Subsidy</td>
<td>0.145</td>
<td>0.0195</td>
<td>0.0133</td>
<td>0.02</td>
<td>0.0257</td>
<td>0.0212</td>
</tr>
<tr>
<td></td>
<td>(0.0219)</td>
<td>(0.0245)</td>
<td>(0.0248)</td>
<td>(0.0291)</td>
<td>(0.0298)</td>
<td>(0.0298)</td>
</tr>
<tr>
<td>Number of Children</td>
<td>0.0791</td>
<td>0.0728</td>
<td>0.0796</td>
<td>0.0827</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0398)</td>
<td>(0.0407)</td>
<td>(0.0402)</td>
<td>(0.0405)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Husband’s Labour Earnings</td>
<td>-0.01</td>
<td>-0.0205</td>
<td>-0.0245</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0138)</td>
<td>(0.0152)</td>
<td>(0.0158)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wife’s Labour Earnings</td>
<td>-0.0008</td>
<td>-0.0173</td>
<td>-0.0235</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0185)</td>
<td>(0.0207)</td>
<td>(0.0215)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any Partner has College Degree</td>
<td>0.1338</td>
<td>0.1415</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0689)</td>
<td>(0.0691)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any Partner Not Employed</td>
<td>-0.0796</td>
<td>-0.0799</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0641)</td>
<td>(0.0641)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>-0.092</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0623)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Age Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.108</td>
<td>0.536</td>
<td>0.541</td>
<td>0.542</td>
<td>0.548</td>
<td>0.551</td>
</tr>
<tr>
<td>Observations</td>
<td>515</td>
<td>515</td>
<td>515</td>
<td>515</td>
<td>515</td>
<td>515</td>
</tr>
</tbody>
</table>

Note: The table shows placebo results of the estimated correlation between separation in the first quarter of the next calendar year and the hypothetical marriage subsidy. Coefficients are estimated via OLS using my sample of unmarried couples based on the SOEP. The hypothetical marriage subsidy of unmarried couples is measured in EUR 1,000 for the year before the separation was reported for separations occurring in the first quarter to reflect accurately the expected hypothetical monetary gain of a delayed separation. All other variables are measured at the time of separation. Column 1 shows results from a specification regressing the indicator variable for first-quarter separations only on the marriage subsidy. Year-fixed effects and age-fixed effects of both spouses are added in column 2. In column 3, I add the number of children living in the household. Column 4 also includes controls for spouses’ labour incomes measured in EUR 10,000. In column 5, I add controls for whether at least one partner has a college degree and if at least one partner is not employed. Column 6 controls for regional differences. All monetary values are denoted in 2014 euro. Standard errors in parentheses are robust to heteroskedasticity.
Table 2B.2: Robustness Check - Alternative Specifications

<table>
<thead>
<tr>
<th></th>
<th>Income Spline (1)</th>
<th>HH Net Inc Spline (2)</th>
<th>Logit (3)</th>
<th>Linear Distance (4)</th>
<th>Missing Educ + Empl (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marriage Subsidy</td>
<td>0.0271 (0.0142)</td>
<td>0.0334 (0.0127)</td>
<td>0.0324 (0.0122)</td>
<td>0.0251 (0.0104)</td>
<td>0.0386 (0.0100)</td>
</tr>
<tr>
<td>Number of Children</td>
<td>0.103 (0.0234)</td>
<td>0.1051 (0.0236)</td>
<td>0.0888 (0.0176)</td>
<td>0.0357 (0.0146)</td>
<td></td>
</tr>
<tr>
<td>Husband’s Labour Earnings</td>
<td></td>
<td>-0.0004 (0.0063)</td>
<td>0.001 (0.0049)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wife’s Labour Earnings</td>
<td></td>
<td>-0.0130 (0.0112)</td>
<td>-0.0023 (0.0073)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any Partner has College Degree</td>
<td>0.0141 (0.0455)</td>
<td>0.0186 (0.0446)</td>
<td>0.0239 (0.0408)</td>
<td>0.024 (0.0287)</td>
<td></td>
</tr>
<tr>
<td>Any Partner Not Employed</td>
<td>-0.012 (0.0529)</td>
<td>-0.0099 (0.0481)</td>
<td>-0.0459 (0.0393)</td>
<td>-0.0068 (0.0278)</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>0.0442 (0.0481)</td>
<td>0.0447 (0.0465)</td>
<td>0.0319 (0.0438)</td>
<td>0.0399 (0.0286)</td>
<td></td>
</tr>
<tr>
<td>Linear Distance to December</td>
<td></td>
<td></td>
<td></td>
<td>-0.1054 (0.0044)</td>
<td></td>
</tr>
<tr>
<td>Marriage Subsidy × Linear</td>
<td></td>
<td></td>
<td></td>
<td>-0.0039 (0.0017)</td>
<td></td>
</tr>
<tr>
<td>Distance to December</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Income Splines</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Income Spline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Age Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R²</td>
<td>0.501</td>
<td>0.496</td>
<td>0.787</td>
<td>0.472</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>819</td>
<td>819</td>
<td>761</td>
<td>819</td>
<td>880</td>
</tr>
</tbody>
</table>

Note: The table shows results from alternative specifications of estimating equation 2.1. Column 1 includes 5-piece splines of both spouses’ labour incomes. Column 2 includes a 5-piece spline of households’ net income. Column 3 presents results from estimating the effect of the marriage subsidy on marital timing decisions related to separation using a logit model estimated via maximum likelihood. Coefficients are shown as average marginal effects. Column 4 includes a linear control for the number of months until December for all separations outside the first quarter. I also include an interaction term between the linear control and the marriage subsidy. Column 5 estimates equation 2.1 on a sample of all couples with non-missing income data but with missing educational measures and employment status. Labour incomes and household net income are measured in EUR 10,000. All monetary values are denoted in 2014 euro. Standard errors in parentheses are robust to heteroskedasticity.
Table 2B.3: Robustness Check - Bandwidth Analysis

<table>
<thead>
<tr>
<th>Months included in the Sample</th>
<th>Dec - Jan (1)</th>
<th>Nov - Feb (2)</th>
<th>Oct - Mar (3)</th>
<th>Sep - Apr (4)</th>
<th>Aug - May (5)</th>
<th>Baseline (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marriage Subsidy</td>
<td>0.0429</td>
<td>0.0023</td>
<td>0.0392</td>
<td>0.0432</td>
<td>0.0304</td>
<td>0.0292</td>
</tr>
<tr>
<td></td>
<td>(0.0341)</td>
<td>(0.0257)</td>
<td>(0.0202)</td>
<td>(0.0165)</td>
<td>(0.0148)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Number of Children</td>
<td>0.0131</td>
<td>0.037</td>
<td>0.085</td>
<td>0.0935</td>
<td>0.1062</td>
<td>0.1012</td>
</tr>
<tr>
<td></td>
<td>(0.0431)</td>
<td>(0.0349)</td>
<td>(0.0309)</td>
<td>(0.0282)</td>
<td>(0.0262)</td>
<td>(0.0233)</td>
</tr>
<tr>
<td>Husband’s Labour Earnings</td>
<td>0.0343</td>
<td>0.0237</td>
<td>-0.0022</td>
<td>-0.0077</td>
<td>-0.0022</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.0207)</td>
<td>(0.0151)</td>
<td>(0.0104)</td>
<td>(0.008)</td>
<td>(0.0069)</td>
<td>(0.0066)</td>
</tr>
<tr>
<td>Wife’s Labour Earnings</td>
<td>0.0198</td>
<td>-0.0135</td>
<td>0.0096</td>
<td>-0.0025</td>
<td>-0.0206</td>
<td>-0.0142</td>
</tr>
<tr>
<td></td>
<td>(0.0261)</td>
<td>(0.0263)</td>
<td>(0.021)</td>
<td>(0.0164)</td>
<td>(0.0127)</td>
<td>(0.0107)</td>
</tr>
<tr>
<td>Any Partner has College Degree</td>
<td>-0.0754</td>
<td>0.0065</td>
<td>0.0321</td>
<td>0.0757</td>
<td>0.063</td>
<td>0.0301</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.084)</td>
<td>(0.0696)</td>
<td>(0.0595)</td>
<td>(0.0508)</td>
<td>(0.0441)</td>
</tr>
<tr>
<td>Any Partner Not Employed</td>
<td>-0.0909</td>
<td>-0.0014</td>
<td>0.0116</td>
<td>-0.0153</td>
<td>-0.041</td>
<td>-0.0353</td>
</tr>
<tr>
<td></td>
<td>(0.1432)</td>
<td>(0.0864)</td>
<td>(0.0837)</td>
<td>(0.0549)</td>
<td>(0.0466)</td>
<td>(0.0415)</td>
</tr>
<tr>
<td>East</td>
<td>0.0657</td>
<td>0.1297</td>
<td>0.0529</td>
<td>0.0252</td>
<td>0.0588</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>(0.1074)</td>
<td>(0.0887)</td>
<td>(0.0681)</td>
<td>(0.0602)</td>
<td>(0.0528)</td>
<td>(0.0462)</td>
</tr>
<tr>
<td>Year Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Age Dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R²</td>
<td>0.933</td>
<td>0.821</td>
<td>0.767</td>
<td>0.659</td>
<td>0.561</td>
<td>0.494</td>
</tr>
<tr>
<td>Observations</td>
<td>175</td>
<td>299</td>
<td>437</td>
<td>557</td>
<td>694</td>
<td>819</td>
</tr>
</tbody>
</table>

Note: The table shows results from estimating the regression model in equation 2.1 on subsamples of different windows around the turn of the year. The dependent variable is whether a couple separated in the first quarter of a year. Column 1 presents results based on a subsample of married couples separating in December and January. Columns 2 to 5 gradually increase the window, including more months before and after the turn of the year. The sample size increases the more months are included in the sample. Year-fixed effects and age-fixed effects of both spouses are included in each column. Labour incomes and household net income are measured in EUR 10,000. All monetary values are denoted in 2014 euro. Standard errors in parentheses are robust to heteroskedasticity.
Chapter 3

Who cares? The Role of Childcare in Reducing Labour Market Inequalities

3.1 Introduction

Having children is one of the most costly events for a family when measured in monetary terms of foregone employment and earnings, in particular for mothers. As the employment of men is virtually unaffected by the arrival of a child, most remaining inequality in the labour market is driven by unequal effects of having children between men and women. Losses in labour market outcomes for women exist across countries with varying intensity, though the long-term impact of childbirth in Germany is among the largest (Kleven et al., 2019).¹

Employment rates of German mothers with young children were traditionally low. In the early 2000s, two out of three women with a child between the age of one and three was not employed. Low employment rates imply high costs of having children for mothers. Kuziemko et al. (2020) show that women underestimate these costs when making investment decisions to human capital, resulting in a potentially inefficiently high level of human capital investment. German policymakers have identified the issue of an ageing population and have implemented reforms in education to

¹Landais et al. (2020) show that losses in earnings are identical between biological and adoptive mothers. In this paper, I investigate the effect for biological mothers only.
increase the pool of skilled workers (Huebener and Marcus, 2015), maternal leave to increase fertility (Raute, 2019) and childcare policies to increase participation in the labour market. How effective are these expansions in childcare provision in increasing mothers’ activity in the labour market?

Using administrative data linked with childcare data, I leverage variation in childcare coverage on the local level to estimate the short and long-run effects of childcare expansions on mothers’ labour market outcomes. I contribute to the literature by studying not only the short-run effects of childcare expansions but also the impact in the long run. I probe the validity of the implicit assumption on homogeneous treatment effects over time, which is required for a causal interpretation of coefficients estimated in a two-way fixed effect model. Additionally, I consider whether the effect of childcare provision is heterogeneous across different levels of initial childcare provision. Using a novel difference-in-difference-in-differences approach based on arguably exogenous initial rates of childcare coverage, I compare the effect of childcare expansions across areas with initially high and low levels of coverage.

I estimate that a 10 percentage point increase in childcare provision is associated with a 2.8 percentage point increase in the probability of being employed for mothers with children between the ages of one and three. Most of the effect is driven by part-time employment. These results compare to Müller and Wrohlich (2020) investigating the effect of childcare expansions on mothers’ probability of employment in Germany in a generalised difference-in-differences framework. Using a 2x2 difference-in-differences strategy, Havnes and Mogstad (2011) find a 0.6 percentage point increase in employment for a 10 percentage point increase in coverage in Norway. In contrast to these previous studies assuming constant treatment effects, I show that mothers with young children have lower average treatment effects closer to the start of childcare expansions. As a higher propensity of employment in the short run might positively affect the long-run trajectories of mothers’ careers, I use the same sample of mothers to estimate the long-run impact. I find that mothers are more likely to switch from part-
time to full-time employment, while the overall probability of employment is unaffected in the long run. These results are similar to Kleven et al. (2020), who find neither short nor long-run effects of large and sudden expansions in childcare supply on mothers’ outcomes in Austria. Moreover, I find that the effect of childcare expansions differs between regions with initially higher and lower rates of childcare coverage. Using a novel difference-in-difference-in-differences approach, I first establish that the expansion in childcare was equally sized in expansive areas across counties with initially high and low childcare provision rates. I estimate a 4.4 percentage point higher probability of being employed for mothers in counties with initially low provision for a 10 percentage point increase in coverage. The effect on mothers living in counties where initial rates were higher, I find no impact on employment. However, the difference between these two estimates is statistically significant. These results contribute to the literature by investigating a new channel of heterogeneity: initial provision levels, in addition to mothers’ age and number of children (Nollenberger and Rodríguez-Planas, 2015), mothers’ skill levels (Müller and Wrohlich, 2020) and pre-birth earnings (Kleven et al., 2020).

Empirical evidence in the literature is often based on generalised difference-in-differences frameworks using two-way fixed effect models estimated via OLS. My results indicate that estimates from these models can only be interpreted as “the” causal effect with caution. Callaway et al. (2021) show that estimates from a two-way fixed effect model only recover a weighted average of reasonable treatment effects under strong homogeneity assumptions. As the effects of childcare on mothers’ outcomes differ between the short and the long run, vary by calendar year when fixing years since childbirth and are heterogeneous across pre-expansion levels of childcare, the estimated parameter of interest from a two-way fixed effect model could be biased.

The rest of the paper is structured as follows. Section 2 summarises key findings of previous literature on the effects of childcare on mothers’ labour market outcomes.
Section 3 describes the childcare reform in Germany. Section 4 describes the data and explains the sample selection. Section 5 presents the empirical framework. Section 6 shows the estimated effects of childcare expansions on labour market outcomes, including long-run and heterogeneous effects. Section 7 concludes.

### 3.2 Empirical Studies on the Effect of Childcare on Mothers’ Labour Market Outcomes

My paper adds to studies using micro-level data on the relationship between mothers’ labour market outcomes and the provision of publicly funded childcare as surveyed by Olivetti and Petrongolo (2017). Overall, the literature has found mixed results across countries. I first present papers related to Germany, then papers related to other similar contexts, followed by papers studying less similar contexts.

My paper takes the results in Müller and Wrohlich (2020) as a starting point. They investigate how the expansion in publicly subsidised childcare provision in Germany affected the employment of mothers with children between the ages of one and three using a generalised difference-in-differences framework. Estimating a two-way fixed effect model on individual-level data from the microcensus, they find that an increase in the childcare take-up rate by 10 percentage points increases mothers’ participation in the labour force by 2.2 percentage points. The authors show that most of the effect is driven by women working part-time, between 20 and 35 hours a week. When testing for heterogeneous effects across education groups, they find that the reform only increased participation for mothers with medium skill levels (A-levels or vocational training), but not for college-educated or untrained mothers. In contrast to my paper, their study differs along several dimensions. While they use the microcensus, I rely on a sample of mothers based on social security records resulting in a slightly different sample selection, which is discussed further in section 4. Additionally, while Müller and Wrohlich (2020) focus on the effect of childcare
during the first three years after childbirth, I show estimates for up to eight years after childbirth providing a longer-run view of the effect. I also provide a further robustness check on the effect of childcare during the first three years after childbirth by allowing for heterogeneity with respect to calendar year, as mothers with young children have less exposure if they gave birth closer to the start of childcare expansions. Lastly, I expand their framework to allow for the effect of childcare expansions to differ between regions with initially higher and lower rates of childcare coverage by using a novel difference-in-difference-in-differences approach.

Bauernschuster and Schlotter (2015) use an expansion of kindergarten places in post-reunification Germany. They present an intention-to-treat effect on employment of 6 percentage points of kindergarten coverage and a local average treatment effect of 3.7 percentage points for a 10 percentage point increase using age cut-offs as an instrument.

Papers related to other similar contexts include Havnes and Mogstad (2011) showing that mothers did not react to large childcare expansions in Norway during the 1970s using a 2x2 difference-in-differences framework. They argue that broadening universally accessible childcare crowded out existing informal childcare arrangements. Nevertheless, the authors point out that their results might only be valid in the short run as women and firms could adjust to the higher level of childcare, leading to improved labour market outcomes in the long run.

For Spain, Nollenberger and Rodríguez-Planas (2015) leverage the staggered funding of childcare expansions for three-year-old children across states to estimate the effect on mothers’ employment. They find that the reform increased maternal employment by 2.8 percentage points using a triple difference-in-differences framework comparing mothers of two-year-old and three-year-old children whose enrolment rates increased by 9.1 and 1.1 percentage points, respectively. They find that most of the effect stems from mothers over the age of 30 and mothers with two or more children.
Kleven et al. (2020) investigate the impact of parental leave reforms and childcare policies for the last 60 years in Austria using a double event-study approach, finding no effect in either the short or the long run. Their results also hold when repeating the estimation on a sample of mothers in the top pre-birth income quartile only.

My paper contributes to the literature by studying the long-run effects of childcare expansions on mothers’ labour market outcomes. I estimate the effect of the expansions up to eight years after childbirth. Therefore, my results connect the existing literature in the German context with long-run estimates.

Regarding less similar contexts, Schlosser (2018) exploits the staggered introduction of a universal preschool law in Israel in the early 2000s. Her setting is only of limited comparability as she compares two culturally different groups with vastly different levels of pre-reform childcare provision rates. She finds a positive effect on mothers’ labour supply with a larger impact on Arab mothers than for Israeli mothers.

Cascio (2009) employ the staggered introduction and age targeting of subsidised kindergarten - rather than early childcare of under three-year-olds as in my paper - in the US in the 1960s to study the effect on mothers’ labour supply. She finds that four mothers enter the labour force for every ten children who were enrolled in public kindergarten due to the policy change. She also shows a crowding out of private kindergarten enrolment due to subsidised public kindergarten.

Using data on preschool construction and enrolment rates in Argentina, Berlinski and Galiani (2007) find that the likelihood of maternal employment increases by between 0.7 and 1.4 percentage points for a 10 percentage point increase in childcare supply, which is broadly similar to a later study by Berlinski et al. (2011) using a regression discontinuity design. The expansion under investigation is these papers is based on constructing new childcare facilities, whereas the German expansion was mostly implemented by converting existing kindergarten places into early childcare slots.
3.3 Mothers and Childcare in Germany

3.3.1 Labour Market Participation of Mothers

Although women in economically advanced Western countries reached labour force participation rates up to 80% by the 2010s (Blau and Kahn, 2013), women with children in Germany, especially mothers of young children, have had low participation rates.

Women’s labour force participation differs between West and East Germany due to its divide for many decades. The Federal Republic of Germany, commonly referred to as West Germany, was leaning towards gender-conservative policies. These policies encouraged a single-earner model and included, for example, the income tax system based on income splitting inducing high marginal tax rates on second earners (Fink, 2020), limited parental leave until 1989 with little replacement of income (Schönberg and Ludsteck, 2014), and a limited supply of childcare (Müller and Wrohlich, 2020). The German Democratic Republic, known as East Germany, encouraged mothers to participate in the labour market by promoting the take-up of publicly provided childcare (Rosenfeld et al., 2004). Thus, mothers’ employment rates significantly differed between the two states just before the German reunification in 1990 (Boelmann et al., 2021) and were still different in the mid-2000s (Müller and Wrohlich, 2020).

In response to low participation rates among mothers and low fertility rates since the 1970s (Bauernschuster et al., 2014), the German federal government started a series of programmes to increase the supply of publicly subsidised childcare, allowing mothers with young children an easier return to employment. In this paper, I will focus on West German mothers due to historical differences between the two.
3.3.2 Childcare Reform and Expansions

In the early 2000s, childcare availability was scarce for children under the age of 
three. The average number of children in this age group enrolled in childcare was 
roughly 9 per 100 children across all counties in East and West Germany in 2002. 
Figure 3.1 presents childcare enrolment rates across West German counties for 2002 
in panel (a). Enrolment rates were low and varied across regions due to varying 
supply. On average, only 2 in 100 children under the age of three were enrolled in the 
former area of West Germany. Wrohlich (2008) estimates that 59% of children were 
rationed for formal childcare in West Germany in 2002. Furthermore, strict regulations 
regarding group size, staff-child ratios, opening hours, necessary qualifications and 
exams of staff led to no private market emerging (Felfe and Lalive, 2018). Therefore, 
the main providers of formal childcare were municipalities, religious and non-religious 
non-profit institutions or commercial providers, though the latter only accounted for 
about 2% in 2009 (Mühler, 2010).

The daycare expansion law, Tagesbetreuungsausbaugesetz, of 2005 was the first 
law intended to increase the supply of childcare, especially in the states of former 
West Germany. The law pursued an expansion aimed at creating nearly 160,000 
new places in childcare centres and about 70,000 in publicly financed daycare by 
2010. Most newly created places in childcare centres were converted kindergarten 
slots that became superfluous due to demographic change. Federal funds also 
subsidised the operation of childcare centres such that 70% to 80% of running costs 
are covered by public sources, whereas the remaining share is financed through 
fees which vary across counties and depend on parents’ income (Mühler, 2010; 
Bauernschuster et al., 2014). Two years after the law had been passed, a summit of 
federal, state and local representatives was organised to discuss the lack of financial 
backing from the federal level. By the end of the summit, the support for children law, 
Kinderförderungsgesetz, was passed. Besides defining more explicit guidelines for 
financial responsibilities, the law also stated that the new targeted childcare coverage
rate was 35% across Germany by 2013. Policymakers also agreed on introducing a legal claim for publicly subsidised childcare, which came into effect on 1 August 2013. By 2018, the federal government had subsidised the creation of more than 564,000 childcare slots (Bundesministerium für Familie, Senioren, Frauen und Jugend, 2020). Panel (b) of figure 3.1 shows enrolment rates for the year 2017. Childcare was still unequally distributed across counties but on a higher level of provision. 28 in 100 children were, on average, enrolled in the former area of West Germany. The number of children enrolled in childcare in East and West Germany increased by 532,410, implying a 94.4% take-up rate of newly created slots. Panel (c) of figure 3.1 illustrates that the intensity of the expansions considerably differed across counties.

3.4 Data

I combine the accuracy and size of administrative data from German social security records with local-level data on childcare enrolment of children under the age of three to investigate the relationship between childcare supply and mothers’ labour market outcomes.

3.4.1 Data on Labour Market Outcomes and Childcare

Sample of Integrated Labour Market Biographies - SIAB

I use the Sample of Integrated Labour Market Biographies (SIAB) as a source of individual-level data on labour market outcomes (Antoni et al., 2019). The SIAB is a 2% random sample derived from the Integrated Employment Biographies (IEB) of the Institute for Employment Research (IAB) at the German federal employment agency (Bundesagentur für Arbeit). The IEB follows a person’s employment status and records any notifications to the social security system on the exact day. The data includes any person in Germany who is employed subject to social security, in marginal part-time employment, receives benefits under the German Social Code III or II, is registered as job-seeking or planning on participating in programs of active
labour market policies (Antoni et al., 2019). The data excludes information on civil servants, self-employed, and individuals out of the labour force. The data spans all employment spells of individuals in the sample between 1975 and 2017. Besides employment, the SIAB includes a range of demographics such as gender, year of birth, nationality, and location data at the county, *Kreis*, level.

I follow the procedure by Müller and Strauch (2017) to identify mothers, using information from employers’ notifications due to employment interruptions. As the statutory health insurance provider pays maternity benefits, the notification entries in the IEB are identical for absence due to illness for more than six weeks and maternal leave. Therefore, this strategy of identifying mothers assumes that women of a certain age are less likely to interrupt employment due to long-term illness. The scheduled date of birth is reconciled by using the end date of the last employment spell before the notification of absence and adding 42 days as maternity leave starts six weeks before the estimated childbirth. Although the procedure is unlikely to determine the exact date of birth, the empirical strategy in my paper will not rely on the exact date. Another caveat is that this algorithm cannot account for multiple births, stillbirths, or infant mortality (Müller and Strauch, 2017). Gauging the validity of their method, the authors compare the number of births identified by their procedure with data from the statistical offices for mothers across age groups. As the SIAB is a 2% sample of the IEB, the resulting number of identified births using the SIAB should be multiplied by 50 for comparison with official statistics. Müller and Strauch (2017) show that their method identifies fewer births compared to official statistics, as the SIAB does not include data on civil servants, self-employed, and individuals out of the labour force as mothers need to be recorded in the SIAB before childbirth to be identified by the procedure. Additionally, they set the maximum age for childbearing at 40 after checking for other thresholds. To avoid misclassification of long-term sickness leave as childbirth, they consider all employment interruptions shorter than 98 days to be
long-term sickness.\footnote{\textsuperscript{2}} Using the tables provided by Müller and Strauch (2017), the number of scaled-up births identified by the procedure is about half of the number of births recorded in official statistics.

One key reason for the discrepancy between the scaled number of births recorded in the SIAB and official data is that only mothers who were employed or registered unemployed\footnote{\textsuperscript{3}} at the time maternity leave benefits started to pay out are recorded in the SIAB. Therefore, if a mother gave birth without having a record in the SIAB, but picked up employment after birth, I would not identify this mother as having had a child. Suppose mothers who were not recorded prior to childbirth are systematically more (less) likely to work after giving birth due to an increase in childcare supply, my estimates would be biased downward (upward). When interpreting my results - especially in comparison to other studies - it is important to keep this feature of the data in mind. Müller and Wrohlich (2020) use the microcensus in their study, which might include mothers with no record in the SIAB. They find a slightly lower point estimate of the effect of childcare expansions on mothers’ labour market outcomes for the years 2007 to 2014. This difference might suggest that mothers who were not recorded prior to childbirth could be less likely to work after giving birth due to an increase in childcare supply. Therefore, my results should be interpreted as an upper bound.

As mothers might leave the labour force after childbirth, I impute spells of no labour force participation by creating observations for years in which a previously recorded individual has not been registered as unemployed or employed. Location information is only systematically available for years after 1999 for employed and registered unemployed individuals. I impute any missing location by assuming that each individual continued to reside at the last observed location for subsequent years.

\footnote{\textsuperscript{2}}Paid maternity leave commences six weeks before the estimated date of childbirth and lasts for eight weeks after childbirth. Therefore, employment interruptions due to childbirth last for (6 weeks + 8 weeks) × 7 days = 98 days of paid maternity leave.

\footnote{\textsuperscript{3}}For readability, I use the terms "employed or registered unemployed" to refer to any labour force status that would result in a registered spell in the IEB. These include individuals who are employed (i.e. subject to social security, marginal part-time employment), receive benefits in accordance with the German Social Code III or II, are registered as job-seeking or planning on participating in programs of active labour market policies.
Childcare Data

Data on childcare enrolment for children up to the age of three is provided by the Federal Statistical Office and Statistical Offices of the States for 2006 until 2017 as well as 2002 (Statistische Ämter des Bundes und der Länder, 2020). The data contains information on the number of children up to the age of three, the number of children enrolled in childcare and the number of children enrolled in childcare for at least 7 hours a day for each county. I refer to the latter as full-time childcare.

Over the observed period, counties changed their borders or merged with neighbouring counties into larger ones due to administrative reorganisations. I use the definition of territorial units of 2020 to construct a panel data set for 324 time-consistent counties in West Germany from 2002 until 2017. The main measures of childcare coverage are constructed by dividing the number of children enrolled in childcare by the number of children living in each county. I create separate measures for overall and full-time enrolment and interpret the resulting measures as coverage quotas.

3.4.2 Sample Selection and Descriptive Statistics

For my main sample, I select mothers living in West Germany who were between the ages of 20 and 45 at the time of childbirth. Following Müller and Wrohlich (2020), the sample includes mothers with children between the ages of one and three for the years 2007 until 2014. Descriptive statistics of my main sample are shown in table 3.1. The sample consists of 35,616 distinct mothers and 84,010 calendar year-mother observations. The average childcare coverage rate was about 19% between 2007 and 2014. 95% of mothers were employed one year before childbirth, whereas nearly half were employed after childbirth. The very high employment share before childbirth is a result of the sample selection in the SIAB, as only individuals who were employed or registered unemployed at the time maternity leave benefits started to pay out are recorded in the sample. The share of mothers working part-time increases after childbirth by 12 percentage points. While more than three-quarters of women work full-time before giving birth, only 21% do so after birth. Nearly 15% of mothers have
a university degree. Two-thirds of mothers in the sample have vocational training, whereas 20% have no vocational training.

3.5 Empirical Framework

3.5.1 Generalised Difference-in-Differences

I follow Müller and Wrohlich (2020) and leverage variation in childcare coverage across counties by estimating a two-way fixed effect model as my baseline specification via OLS:

\[ y_{i,s,t} = \beta CCC_{s,t} + \kappa_s + \omega_t + X_{i,s,t} \gamma + v_{i,s,t}. \]  (3.1)

The outcome variable \( y_{i,s,t} \) denotes the labour market outcome of interest. \( \kappa_s \) and \( \omega_t \) are county and calendar year-fixed effects, respectively. \( CCC_{s,t} \) is the measure of childcare coverage, calculated as the ratio between the number of children enrolled in early childcare and the total number of children between the ages of one and three living in each county each year. \( X_{i,s,t} \) is a set of controls including age, age squared and dummies for three levels of education (no vocational training, vocational training, university degree). The parameter of interest is \( \beta \). This specification and choice of controls is similar to equation 1 in Müller and Wrohlich (2020) and shall serve as an established benchmark.

Equation 3.1 is estimated on my sample of mothers with children between the ages of one and three spanning the years 2007 until 2014. The identifying assumption for this generalised difference-in-differences approach is that changes in mothers’ labour market outcomes for areas with high provision rates of childcare would have been equal to those in areas with provision rates in the absence of the expansion. This parallel trend assumption is similar to Havnes and Mogstad (2011) and Müller and Wrohlich (2020). In addition to a parallel trend assumption, the interpretation of \( \beta \) as a sensibly weighted causal parameter likely involves invoking a stronger
set of assumptions regarding treatment effect homogeneity (Callaway et al., 2021),
which I probe below. Figure 3.2 provides supporting evidence for the validity of the
parallel trend assumption. Before the first law was passed in 2005, employment rates
and shares of part-time and full-time employment moved in parallel until the start of
expansions in 2005.4

3.5.2 Identification

Identification for the parameter of interest, $\beta$, stems from variation in the supply of
early childcare across counties over many years. Figure 3.1 illustrates that the size
of childcare expansions varied across counties. In this section, I provide additional
evidence beyond existing documentation in previous papers showing that at least
part of this variation can be regarded as exogenous. A key source of variation is that
counties do not follow standardised procedures when forecasting future demand for
childcare. Consequently, estimates for future demand differ across regions, even
under similar demand for childcare. The Kommunalverband für Jugend und Soziales
Baden-Württemberg (2020) provides excellent documentation about this topic.

The Bundesministerium für Familie, Senioren, Frauen und Jugend (2012) reports
that the vast majority of counties (88%) draws data from registration offices to inform
their forecast for future demand. The second and third most-used sources of infor-
mation are registration and waiting lists for facilities (68%) and data on past take-up
rates (50%), respectively. Although parents ought to inform the municipality about
their intended use of childcare at least six months before enrolment, it is common
practice for parents to inform the childcare facility directly. Consequently, municipal-
ities are often unaware of a large fraction of registered interest in using childcare,
even if municipalities operate these childcare facilities (Kommunalverband für Jugend
und Soziales Baden-Württemberg, 2020). Thus, the provision of childcare would

---

4Figure 3.2 shows a sudden spike in employment and full-time employment in 2004. This spike
occurs equally for above and below-median growth counties. Therefore, the spike does not pose a
threat to the identification strategy of my generalised difference-in-differences approach, as it occurs
for both the "control" and the "treatment" group.
be idiosyncratically different across counties, even if the demand for childcare was the same. However, if the provision was primarily based on surveys as a proxy for demand, then variation in childcare would be more endogenous to mothers’ activity in the labour market. Although 32% of counties conduct surveys among parents, responses often diverge from the actual demand for childcare (Hüsken, 2011). Thus, municipalities’ forecasts of the future need for childcare are unlikely to be driven primarily by demand.

Moreover, there is considerable heterogeneity in the planning procedure across counties. City counties, *Kreisfreie Städte*, are responsible for planning and executing the planning task. In rural counties, *Landkreise*, the local youth office is responsible for planning, whereas municipalities in these counties execute the planning task based on limited information on registered demand. While municipalities in some rural counties meet regularly to coordinate planning procedures, other counties have no coordination across municipalities. These bureaucratic differences on the local level induce some additional randomness in the variation of childcare coverage.

Finally, counties have differing expertise in dealing with the funding system, which involves various federal levels, presenting another source of variation likely uncorrelated with mothers’ labour market outcomes. Based on childcare demand forecasts, local authorities in counties submit applications to the non-legislative government between the state and local level\(^5\), which apply for funding from the state. States receive reimbursement from the federal government after deciding to fully fund, partially fund, or not fund new childcare places in the municipalities. The process could be lengthy, so municipalities often start building new childcare centres before receiving money from federal funds without knowing to what extent their costs will be reimbursed.\(^6\)

When building new childcare facilities, municipalities also face different constraints regarding adequate grounds and sourcing qualified staff to open and operate these

\(^5\) These non-legislative government bodies between the state and local level are called *Regierungspräsidien*, *Bezirksregierungen* or similar other names depending on the region.

\(^6\) I am grateful to Joachim Fiebig at the *Kommunalverband für Jugend und Soziales Baden-Württemberg* for his detailed explanations. All mistakes and inaccuracies are my own.
new centres (Bauernschuster et al., 2014).

Overall, at least some variation in childcare coverage rates is arguably exogenous. Similar to Müller and Wrohlich (2020), I add a set of controls to capture factors correlated with mothers’ labour market outcomes and childcare use. Nevertheless, a causal interpretation of the effect of childcare expansions is barred when some variation on childcare expansions remains endogenous.  

### 3.6 Results

#### 3.6.1 Baseline Results

Table 3.2 shows results from estimating equation 3.1 via OLS. Column 1 presents results for employment as the outcome. As each mother in the sample is observed up to three years after childbirth, the data allows me to observe a mother’s employment status for each year separately. The employment variable takes on the value one if the mother is employed in that particular year after childbirth and zero otherwise. The variable measures the extensive margin response of mothers averaged over the first three years after childbirth. The estimates suggest that a 10 percentage point increase in childcare coverage is associated with a 2.75 percentage point increase in the probability of being employed one to three years after childbirth.

The coefficient is statistically significant at the 5% level. Columns 2 and 3 show estimates for part-time and full-time employment, respectively. In my sample, most of the correlation between employment and childcare is driven by part-time employment rather than full-time. Table 3A.1 in the appendix shows results for full-time childcare and suggests that no measure of employment is associated with changes in full-time childcare.

---

7 In a previous version, I have exploited the arguably exogenous timing of creating new slots. Following Kleven et al. (2020), I based my identification on sudden large expansions of childcare within a county, as these sudden expansions can be regarded as exogenous in terms of timing and size. However, the results suffered from imprecision and an overall lack of power.
My estimates are close to the corresponding coefficients by Müller and Wrohlich (2020), who report a 2.19 percentage point increase in the probability of participating in the labour market, which is statically significant at the 10% level using the micro-census for the same years. The effect of childcare enrolments is also comparable to Havnes and Mogstad (2011), who find qualitatively similar results for Norway. Previous papers interpret similar estimates as mothers of young children reacting mostly at the extensive margin of labour supply rather than at the intensive margin of switching from part-time to full-time employment.

A caveat of the two-way fixed effect model with continuous treatment is that interpreting $\beta$ as a weighted average of sensible treatment effects requires strong assumptions regarding treatment effect homogeneity: 1) treatment effects do not vary over time, 2) homogeneous causal responses across groups that receive treatment at different times, and 3) homogeneous causal responses across doses (Callaway et al., 2021).

Robustness Check: Heterogeneous Treatment Effects Over Time

Responses to expansions in childcare might be heterogeneous across mothers who had their first child in different years relative to the start of the expansion. One source of this heterogeneity could be slow-moving gender norms concerning what mothers of very young children "ought" to do (Kleven et al., 2020; Boelmann et al., 2021). Another explanation might be that women who gave birth shortly before the start of the expansion were only partially exposed to the reform. For example, if a mother gave birth in 2004, she would have experienced higher childcare provision only in the third year after birth. In contrast, mothers who gave birth in 2007 experienced, on average, higher levels of childcare for three years. When estimating $\beta$ in equation 3.1, potential heterogeneity in treatment effects across birth cohorts could lead to a weighting of non-sensible treatment effect estimates.

One way to allow the effect of childcare on mothers’ outcomes to differ over
calendar years is to interact the level of provision with calendar year dummies. I estimate the following modified version of the two-way fixed effect specification:

\[ y_{i,s,t} = \sum_{k \neq 2007} \beta_k (CCC_{s,t} \times 1[t = k]) + \kappa_s + \omega_t + X_{i,s,t} \gamma + \nu_{i,s,t}. \]  

(3.2)

The main identifying assumption remains that mothers' outcomes would have evolved similarly in low and high-coverage counties in the absence of childcare expansions. \( CCC_{s,t} \) is the level of childcare in county \( s \) in calendar year \( t \), interacted with calendar year dummies. The parameters of interest, \( \beta_k \), capture heterogeneous effects of childcare coverage over calendar time. I follow Müller and Wrohlich (2020) in setting the baseline year to 2007. \( X_{i,s,t} \) is the same set of controls as in equation 3.1. I estimate equation 3.2 on my sample of mothers with children between the ages of one and three. Since data on childcare enrolment is also available for 2002 and 2006, I add these years to the sample to estimate coefficients for pre-expansion years. This approach of probing dynamic treatment effects is similar to Bauernschuster and Schlotter (2015).

Figure 3A.1 in the appendix shows estimates of \( \beta_k \) in equation 3.2. Panel (a) shows results for mothers' employment and suggests that treatment effects are smaller for the first two years after childbirth. The effect is marginally statistically significant at the 5% level. Panel (b) shows coefficients for part-time employment, implying a similar dynamic pattern, but the effects are imprecisely estimated. Similarly, none of the coefficients on full-time employment in panel (c) is statistically significant.

One interpretation of this robustness exercise is that mothers who experience expansions earlier have lower average treatment effects. These dynamics are indicative that the assumption of homogeneous causal responses across groups receiving treatment at different times could be violated.
### 3.6.2 Long-run Effects of Childcare on Labour Market Outcomes

Increases in childcare coverage are associated with a higher probability of mothers working part-time. Müller and Wrohlich (2020) find a higher probability of working between 20 to 30 hours a week for higher levels of childcare. If higher levels of childcare provision enable mothers to work longer and at certain core hours, which are disproportionately rewarded (Goldin, 2014), then early childcare provision could also impact the long-run trajectories of mothers’ careers.

I test whether childcare expansions persistently increase mothers’ employment by shifting the window of calendar years forward. Using the same sample of mothers as in my baseline results, I estimate mothers’ probability of employment six to eight years after childbirth using the same specification as equation 3.1. These results also serve as a test for dynamics in treatment effects.

Table 3.3 presents coefficients for the long-run effect of childcare coverage on mothers’ labour market outcomes from a specification as equation 3.1 estimated via OLS. Column 1 shows results for employment and suggests that a 10% percentage point increase in coverage is associated with a 0.99 percentage point higher probability of employment six to eight years after childbirth. The effect is not statistically significant on any conventional level. Columns 2 and 3 present results for part-time and full-time employment, respectively. An increase in provision is associated with a 2.6 percentage point lower probability of being employed part-time but a 3.6 percentage point higher probability of being employed full-time, where the latter effect is statistically significant at the 1% level.

One interpretation consistent with these estimates is that increases in childcare provision have no measurable effect on the probability of employment in the long run. Positive effects on the probability of full-time employment represent switches from part-time work. Furthermore, the effect of childcare fading-out in the long run suggests that treatment effects are dynamic with respect to years after childbirth.
Consequently, the estimate of $\beta$ in equation 3.1 might be a weighted average of unreasonable treatment effects should treatment effects be dynamic within the zero-to-three age bin. (Callaway et al., 2021).

3.6.3 Heterogeneous Effects by Levels of Childcare Coverage

Before the start of childcare expansions, provision levels differed across counties as shown in panel (a) of figure 3.1. Suppose that in counties with initially low coverage rates, mothers are constrained in their employment decisions due to the lack of formal childcare. Increasing formal childcare provision in these counties would induce mothers to take up employment and complement additional care needs with informal arrangements. However, in counties with high initial coverage rates, mothers with the highest propensity to work are likely already employed as they are less constrained by formal childcare provision. Additional childcare places in these counties would lead to fewer mothers taking up employment. Since mothers of young children react mostly at the extensive margin of labour supply rather than at the intensive margin of switching from part-time to full-time employment in the short run, an increase in provision in counties with initially high levels is unlikely to induce switches from part-time to full-time employment.

Using a difference-in-difference-in-differences approach, I explore potentially heterogeneous treatment effects across pre-expansion levels of childcare. If the effect of childcare expansions differed across areas with higher and lower initial provision rates, then the estimate of $\beta$ in equation 3.1 might be contaminated due to a violation of the assumption of homogeneous causal responses across doses (Callaway et al., 2021). Therefore, my heterogeneity exercise can be seen as a test for treatment effect heterogeneity with respect to pre-expansion levels of childcare.

Initial levels of childcare coverage were mostly determined by demographic factors as counties mainly use data from registration offices, registration and waiting lists for facilities, and data on past take-up rates to inform their provision of childcare.
(Bundesministerium für Familie, Senioren, Frauen und Jugend, 2012). I can exploit the differences in initial childcare provision rates combined with varying intensities of expansions across areas to investigate potential heterogeneous effects across levels of childcare provision.

I follow Hüsken (2011) and define four types of childcare expansion based on initial coverage rates and the intensity of expansions. I define "newcomers" as counties whose level of childcare coverage was below the 40th percentile in 2002 but had an expansion in childcare between 2002 and 2017 that was above the 60th percentile. "Latecomers" are counties with a provision below the 40th percentile in 2002 and an expansion below the 40th percentile between 2002 and 2017. Counties with coverage levels above the 60th percentile in 2002 are "forerunners" if the expansion was higher than the 60th percentile or "dropouts" if the expansion was below the 40th percentile.

Figure 3.3 provides an overview of the four types of expansion. Panel (a) illustrates the definition of the four different types of childcare expansions in a scatter plot, whereas panel (b) presents the level of childcare coverage over time for each expansion type. In 2002, initial levels of childcare provision were similarly low for newcomers and latecomers and similarly high for forerunners and dropouts. Between 2002 and 2017, childcare provision expanded more strongly among newcomers and forerunners but less for latecomers and dropouts.

I test for heterogeneous effects of childcare expansions across different pre-expansion levels of provision by estimating the following difference-in-difference-in-differences model via OLS (Gruber, 1994; Olden and Møen, 2022):

\[
y_{i,s,t} = \lambda_1 \text{after}_t + \lambda_2 \text{high}_s + \lambda_3 \text{expansive}_s + \lambda_4 (\text{after}_t \times \text{high}_s) + \\
\lambda_5 (\text{after}_t \times \text{expansive}_s) + \lambda_6 (\text{high}_s \times \text{expansive}_s) + \\
\lambda_7 (\text{after}_t \times \text{high}_s \times \text{expansive}_s) + \mathbf{X}_{i,s,t} \gamma + \xi_{i,s,t},
\]  

(3.3)
where the outcome variable $y_{i,s,t}$ denotes the labour market outcome of interest. $X_{i,s,t}$ is the same set of controls of age, age squared, and dummies for three levels of education (no vocational training, vocational training, university degree) as before. The dummy $after_t$ indicates post-expansion periods, which are years since 2006. $high_s$ captures differences in levels between types with low initial levels of childcare (newcomers and latecomers) and high initial levels (forerunners and dropouts) before the start of the expansion. The binary variable $expansive_s$ allows for different levels between expansive (newcomers and forerunners) and non-expansive (latecomers and dropouts) types before the expansion. 103 of 208 counties in this subsample are classified as having high initial coverage. The parameter of interest measuring heterogeneous effects of childcare across initial provision rates is $\lambda_7$. Interpreting the estimates as causal requires a parallel trend assumption stipulating that the difference in trends between high- and low-provision counties as measured before expansion is the same for expansive and non-expansive counties (Olden and Møen, 2022). Figure 3.4 plots measures of mothers’ outcomes over time to assess potential trends. The figure indicates that the unconditional averages of outcomes across groups evolve in parallel until the start of expansions.\(^8\)

Table 3.4 shows estimates of equation 3.3 for the three employment measures. Column 1 presents estimates for a specification in which childcare coverage is the dependent variable. The coefficient on the interaction term $after_t \times expansive_s$ measures the difference in the expansion’s intensity between newcomers and latecomers. Newcomers created 6.3 more childcare slots per 100 children compared to latecomers.\(^9\) The coefficient on $after_t \times high_s \times expansive_s$ measures whether forerunners expanded more relative to dropouts compared to newcomers relative to latecomers. As the estimate is small and statistically insignificant, the intensity of relative expans-

---

\(^8\)Figure 3.4 also shows a sudden spike in employment and full-time employment for the year 2004. Reassuringly, this spike occurs equally for all four expansion types. Therefore, the spike does not pose a threat to the identification strategy of my difference-in-difference-in-differences approach, as it occurs for all types.

\(^9\)The difference in intensity of the expansion between the four types does not differ for other cut-offs in the data. Therefore, my results are likely to hold for other cut-off points as well.
sions was similar across counties with high and low initial rates. This allows for a clean comparison of areas with different initial rates. Columns 2 to 4 of table 3.4 show results for employment measures as the left-hand side variable. Coefficients correspond to estimates of the average treatment effects of the treated. They are only comparable with the intention-to-treat estimates in table 3.2 when divided by the intensity of relative expansion from column 1. The estimate of \( \lambda_5 \) suggests that a 10 percentage point increase in childcare coverage is associated with a 4.4 percentage point higher probability of being employed for mothers in counties with initially low provision. The estimate is statistically significant at the 5% level. The estimated parameter of interest governing heterogeneous effects across initial provision rates, \( \lambda_7 \), is precisely estimated, suggesting a smaller effect of childcare expansions on employment in counties with initially high coverage rates. In these counties, an increase in childcare coverage is associated with a lower probability of employment, though the overall effect is statistically insignificant on all conventional levels. For part-time and full-time employment as the dependent variable, a 10 percentage point increase in coverage is associated with a 2.1 and 2.7 percentage point higher probability in low-coverage counties. Both effects are imprecisely measured. The effect is not statistically significantly different between high and low-provision counties.

Overall, this heterogeneity exercise suggests that mothers react more strongly to childcare expansions in counties where childcare provision was lower before the reform than mothers living in counties with higher initial rates. One explanation consistent with these estimates is that mothers in low-provision areas are more constrained in their employment decisions due to the lack of formal childcare. Mothers in high-provision areas are likely already employed as they would have been less constrained. As childcare coverage has increased much since 2006, these findings indicate that future childcare expansions could have less impact on mothers’ labour market activity.
3.7 Conclusion and Discussion

I study the effects of expansionary childcare policies on mothers’ labour market outcomes. In the early 2000s, the German federal government saw the potential of increasing labour market participation by expanding childcare to ease mothers’ transition back into employment after childbirth and lower the costs of having a child. Between 2007 and 2018, federal funds have subsidised the creation and operation of more than 564,000 childcare slots for children up to the age of three.

The intensity of expansions differed substantially across regions, driven by two key components. Firstly, counties used varying approaches to determine the demand for childcare. These approaches would have led to idiosyncratic differences across counties, even if the demand for childcare was the same. Secondly, applying for federal funding is lengthy and involves all levels of the federal system. Differing experiences in counties handling complex bureaucratic processes induced variation in childcare expansion on the local level.

Leveraging these idiosyncrasies, I estimate the effect of childcare expansions on mothers’ labour market outcomes after childbirth. Using administrative data derived from social security records for mothers in Germany, I find that an increase in childcare coverage is associated with a 2.8 percentage point increase in the probability of being employed for mothers with children between the ages of one and three. Most of the effect is driven by part-time employment. I test for heterogeneous treatment effects and find that mothers who experienced expansions in 2008 and 2009 have lower average treatment effects. Slow-moving gender norms or less exposure to childcare expansions for birth cohorts close to the start of the reform are possible explanations consistent with these estimates.

The increase in part-time employment might enable mothers to work longer and certain core hours, which are disproportionately rewarded (Goldin, 2014). Therefore,
I estimate the effect of childcare expansions in the long run using the same sample of mothers but measuring their outcomes six to eight years after childbirth. While I do not find any statistically significant effect on employment, an increase in childcare provision increases the probability of full-time employment by 3.6 percentage points for a 10 percentage point increase in coverage. The effect stems mainly from mothers switching to full-time employment rather than new employment activity.

Initial childcare rates differed substantially across counties before the expansion, mainly due to different demographic factors. Using a difference-in-difference-in-differences approach, I probe the assumption of treatment effect homogeneity by allowing the effect of childcare expansions to differ by initial levels of provision. In areas with initially low coverage, mothers are likely more constrained in taking up employment due to a lack of formal childcare. An increase in childcare provision in these areas is, therefore, more likely to induce mothers to take up employment and complement additional care needs with informal arrangements. This is consistent with my findings, where a 10 percentage point increase in coverage in low-provision areas is associated with a 4.4 percentage points higher probability of employment. In contrast, mothers living in counties with initially high levels of childcare provision are less likely to take up employment for additional childcare places. I find that the effect on mothers living in these areas is statistically significantly smaller at the 5% level, whereas the overall effect is not statistically significant.

My results also serve as a test of the treatment effect homogeneity assumptions in the two-way fixed effect model needed to recover a weighted average of reasonable treatment effects (Callaway et al., 2021). As the effects of childcare on mothers’ outcomes differ between the short and the long run, vary by calendar year when fixing years since childbirth and are heterogeneous across pre-expansion levels of childcare, the estimated parameter of interest from a two-way fixed effect model with continuous treatment could be biased and should only be interpreted as “the” causal effect with caution. My results suggest that the treatment effect homogeneity assumptions of the two-way fixed effect model are violated. Therefore, researchers
and policymakers should be concerned about the validity of previous estimates based on two-way fixed effect regression models with continuous treatment, particularly in the German context.

If policymakers aim to increase women’s labour force participation rates, then increasing childcare provision might be one tool, but with caveats. The effects are concentrated on the extensive margin of employment in the short run and induce no additional employment in the long run. Additionally, my estimates suggest that additional childcare expansions for areas with already high levels of provision might be unlikely to induce further employment of mothers.

At the time of writing, some state youth offices are linking registered demand for childcare of parents to the actual received provision, which presents promising avenues for future research. Linking these differences to labour market outcomes could be a way of estimating the average treatment effects of the treated. It could also provide a better understanding of the characteristics of mothers who benefit the most from childcare to guide future policy decisions.
References - Chapter 3


Figure 3.1: Childcare Enrolment Rates of Under Three-year-olds in West Germany Between 2002 and 2017

(a) Childcare Enrolment Rates of Under Three-year-olds in 2002

(b) Childcare Enrolment Rates of Under Three-year-olds in 2017

(c) Change in Childcare Enrolment Rates of Under Three-year-olds Between 2006 and 2017

Note: The figure presents childcare enrolment rates across counties. The measures of childcare coverage are constructed by dividing the number of children enrolled in childcare by the number of children living in each county. I create separate measures for overall and full-time enrolment and interpret the resulting measures as coverage quotas. Panel (a) shows enrolment rates for the year 2002. The average number of children enrolled in early childcare was about 2 in 100 children in West Germany. Panel (b) shows enrolment rates for the year 2017. On average, 28 in 100 children were enrolled in the former area of West Germany. Panel (c) shows the intensity of the expansions across counties as the total change in the number of children enrolled per 100.
Figure 3.2: Trends in Labour Market Outcomes Around the Years of Expansion in Childcare by Counties with Above and Below-median Growth

Note: The figure presents time series for mothers’ labour market outcomes. Counties in the above-median group had a total expansion in childcare that was equal to or larger than the median growth in coverage between 2002 and 2017. Counties in the below-median group had total expansions lower than the median growth in coverage. Panel (a) presents trends for employment. Panel (b) and (c) present trends for part-time and full-time employment, respectively. The sample includes mothers living in West Germany who were between the ages of 20 and 45 at the time of childbirth and whose first child is between the ages of one and three.
Figure 3.3: Types of Childcare Expansions

(a) Definition of Childcare Expansion Type

(b) Time-series of Childcare Provision by Expansion Type

Note: The figure shows measures of childcare between 2002 and 2017 for four expansion types following Hüsken (2011). The four childcare expansion types are defined as follows: "Newcomers" (circles, blue) are counties with a level of childcare coverage below the 40th percentile in 2002 and an expansion in childcare between 2002 and 2017, which was above the 60th percentile. "Latecomers" (diamonds, red) are counties with a provision below the 40th percentile in 2002 and an expansion below the 40th percentile between 2002 and 2017. "Forerunners" (triangles, green) are counties with childcare coverage levels above the 60th percentile in 2002 and had an expansion above the 60th percentile between 2002 and 2017. "Dropouts" (crosses, yellow) are counties with childcare coverage levels above the 60th percentile in 2002 and had an expansion below the 40th percentile between 2002 and 2017. Panel (a) is a scatter plot of the total size of childcare expansion between 2002 and 2007 over childcare coverage rates in 2002, illustrating the definition of each expansion type. Panel (b) shows the time series of childcare coverage separately for each expansion type.
Figure 3.4: Trends in Labour Market Outcomes Before and After Years of Expansions in Childcare by Expansion Type

Note: The figure presents time series for mothers' labour market outcomes separately for the four expansion types. Panel (a) presents trends for employment. Panel (b) and (c) present trends for part-time and full-time employment, respectively. The sample includes mothers living in West Germany whose first child is between the ages of one and three.
## Tables - Chapter 3

### Table 3.1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Childcare Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolment Rate</td>
<td>84,010</td>
<td>18.87</td>
<td>8.02</td>
</tr>
<tr>
<td><strong>Labour Market Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Year Before Birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td>84,010</td>
<td>0.951</td>
<td>0.216</td>
</tr>
<tr>
<td>Part-time Employment</td>
<td>84,010</td>
<td>0.185</td>
<td>0.389</td>
</tr>
<tr>
<td>Full-time Employment</td>
<td>84,010</td>
<td>0.765</td>
<td>0.424</td>
</tr>
<tr>
<td>One to Three Years After Birth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td>84,010</td>
<td>0.519</td>
<td>0.500</td>
</tr>
<tr>
<td>Part-time Employment</td>
<td>84,010</td>
<td>0.305</td>
<td>0.460</td>
</tr>
<tr>
<td>Full-time Employment</td>
<td>84,010</td>
<td>0.213</td>
<td>0.410</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at First Birth</td>
<td>84,010</td>
<td>29.71</td>
<td>4.52</td>
</tr>
<tr>
<td>No Vocational Training</td>
<td>84,010</td>
<td>0.192</td>
<td>0.394</td>
</tr>
<tr>
<td>Vocational Training</td>
<td>84,010</td>
<td>0.659</td>
<td>0.474</td>
</tr>
<tr>
<td>University Degree</td>
<td>84,010</td>
<td>0.149</td>
<td>0.356</td>
</tr>
<tr>
<td>Number of Distinct Mothers:</td>
<td>35,616</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Counties:</td>
<td>324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years Included:</td>
<td>2007-2014</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The table presents summary statistics for the main labour market outcomes and covariates from the SIAB and measures of childcare enrolment for children aged up to the age of three based on data provided by the Federal Statistical Office and Statistical Offices of the States (Statistische Ämter des Bundes und der Länder, 2020). Mothers are identified using the procedure proposed by Müller and Strauch (2017) using information from employer’s notifications due to employment interruptions. The sample includes mothers living in West Germany who were between the ages of 20 and 45 at the time of childbirth. Following Müller and Wrohlich (2020), the sample spans the years 2007 until 2014 and only includes mothers whose first child is between the ages of one and three.
Table 3.2: Baseline Results - Generalised Difference-in-Differences Specification

<table>
<thead>
<tr>
<th>Effect of Childcare Coverage on Mothers’ Labour Market Outcomes</th>
<th>Employment</th>
<th>Part-time Employment</th>
<th>Full-time Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Generalised DiD: Mothers with Children Aged 1 to 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CCC_{s,t}$</td>
<td>0.275</td>
<td>0.236</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td>(0.119)</td>
<td>(0.103)</td>
</tr>
<tr>
<td>Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.109</td>
<td>0.051</td>
<td>0.037</td>
</tr>
<tr>
<td>Observations</td>
<td>84,010</td>
<td>84,010</td>
<td>84,010</td>
</tr>
<tr>
<td>Counties Included</td>
<td>324</td>
<td>324</td>
<td>324</td>
</tr>
</tbody>
</table>

Note: The table shows estimates for the relationship between childcare expansions and mothers’ labour market outcomes ($\beta$ from equation 3.1) using OLS. The sample includes mothers with children between the ages of one and three and spans the years 2007 to 2014. Column 1 presents results for employment as the outcome. Columns 2 and 3 show estimates for part-time and full-time employment, respectively. All results are based on mothers living in West Germany only. All specifications include county-fixed effects, year-fixed effects, controls for age, age squared, dummies for three levels of education (no vocational training, vocational training, university degree) and the level of childcare coverage on the right-hand side. Standard errors in parentheses are clustered at the county level.
Table 3.3: Long-run Results - Generalised Difference-in-Differences Specification

<table>
<thead>
<tr>
<th></th>
<th>Employment</th>
<th>Part-time Employment</th>
<th>Full-time Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Generalised DiD: Mothers with Children Aged 6 to 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CCC_{s,t}$</td>
<td>0.099</td>
<td>-0.264</td>
<td>0.363</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
<td>(0.206)</td>
<td>(0.149)</td>
</tr>
<tr>
<td>Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.178</td>
<td>0.094</td>
<td>0.033</td>
</tr>
<tr>
<td>Observations</td>
<td>51,175</td>
<td>51,175</td>
<td>51,175</td>
</tr>
<tr>
<td>Counties Included</td>
<td>324</td>
<td>324</td>
<td>324</td>
</tr>
</tbody>
</table>

Note: The table shows estimates for the relationship between childcare expansions and mothers’ labour market outcomes in the long run ($\beta$ from equation 3.1) using OLS. The sample includes mothers with children between the ages of six and eight and spans the years 2013 to 2017. Column 1 presents results for employment as the outcome. Columns 2 and 3 show estimates for part-time and full-time employment, respectively. All results are based on mothers living in West Germany only. All specifications include county-fixed effects, year-fixed effects, controls for age, age squared, dummies for three levels of education (no vocational training, vocational training, university degree) and the level of childcare coverage on the right-hand side. Standard errors in parentheses are clustered at the county level.
Table 3.4: Difference-in-Difference-in-Differences Results - Effect Heterogeneity Across Initial Levels of Childcare Coverage

<table>
<thead>
<tr>
<th></th>
<th>Coverage (1)</th>
<th>Employment (2)</th>
<th>Part-time Employment (3)</th>
<th>Full-time Employment (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference-in-Difference-in-Differences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(after_t)</td>
<td>0.130</td>
<td>0.038</td>
<td>0.047</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>(high_s)</td>
<td>0.040</td>
<td>0.005</td>
<td>-0.014</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>(expansive_s)</td>
<td>0.021</td>
<td>0.004</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>(after_t \times high_s)</td>
<td>-0.002</td>
<td>0.035</td>
<td>0.013</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>(after_t \times expansive_s)</td>
<td>0.063</td>
<td>0.028</td>
<td>0.013</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>(high_s \times expansive_s)</td>
<td>0.014</td>
<td>0.009</td>
<td>-0.002</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>(after_t \times high_s \times expansive_s)</td>
<td>-0.007</td>
<td>-0.044</td>
<td>-0.024</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R²</td>
<td>0.514</td>
<td></td>
<td>0.102</td>
<td>0.042</td>
</tr>
<tr>
<td>Observations</td>
<td>80,403</td>
<td>102,653</td>
<td>102,653</td>
<td>102,653</td>
</tr>
<tr>
<td>Counties Included</td>
<td>208</td>
<td>208</td>
<td>208</td>
<td>208</td>
</tr>
</tbody>
</table>

Note: The table shows coefficients for the relationship between childcare expansions and mothers’ labour market outcomes when estimated via OLS when allowing for heterogeneity in effects across levels of initial childcare coverage as in equation 3.3. The sample includes mothers living in 208 counties who are defined as one of the four expansion types and includes the years 2000 to 2017. Column 1 presents estimates for childcare coverage as the dependent variable, which is unavailable for years 2000, 2001 and 2003 until 2005. Column 2 presents results for employment as the outcome. Columns 3 and 4 present results for part-time and full-time employment as the outcome, respectively. Included controls are age, age squared, and dummies for three levels of education (no vocational training, vocational training, university degree). Standard errors in parentheses are clustered at the county level.
Appendix 3A: Additional Figures and Tables

Figure 3A.1: Robustness Check - Dynamic Treatment Effects of Childcare Expansions on Labour Market Outcomes

(a) Dynamic Treatment Effects on Employment
(b) Dynamic Treatment Effects on Part-time Employment
(c) Dynamic Treatment Effects on Full-time Employment

Note: The figure presents estimates of the effect of childcare expansions on mothers’ labour market outcomes over time. Coefficients are the interaction terms between levels of childcare provision and calendar year dummies as in equation 3.2 to probe the heterogeneity of treatment effects over calendar time estimated via OLS. The outcome variable is employment in panel (a) and part-time and full-time employment in panels (b) and (c), respectively. The sample spans the years 2002 and 2006 until 2014 and includes mothers with children between the ages of one and three. The interaction between childcare coverage and the year dummy for 2007 (marked by the dashed vertical line) is omitted and serves as the benchmark. All regressions include controls for age, age squared, dummies for three levels of education (no vocational training, vocational training, university degree), year-fixed effects and county-fixed effects. Bars represent 95 percent confidence intervals. Standard errors in parentheses are clustered at the county level.
Table 3A.1: Additional Results - Generalised Difference-in-Differences Specification Using Full-time Care Coverage

<table>
<thead>
<tr>
<th>Effect of Full-time Childcare Coverage on Mothers’ Labour Market Outcomes</th>
<th>Employment (1)</th>
<th>Part-time Employment (2)</th>
<th>Full-time Employment (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalised DiD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$FT - CCC_{a,t}$</td>
<td>0.168</td>
<td>0.060</td>
<td>0.111</td>
</tr>
<tr>
<td>(0.126)</td>
<td>(0.118)</td>
<td>(0.100)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>County fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.109</td>
<td>0.051</td>
<td>0.037</td>
</tr>
<tr>
<td>Observations</td>
<td>83,918</td>
<td>83,918</td>
<td>83,918</td>
</tr>
<tr>
<td>Counties Included</td>
<td>324</td>
<td>324</td>
<td>324</td>
</tr>
</tbody>
</table>

Note: The table shows estimates for the relationship between expansions in full-time childcare coverage and mothers’ labour market outcomes ($\beta$ from equation 3.1) using OLS. The sample includes mothers with children between the ages of one and three and spans the years 2007 to 2014. Column 1 presents results for employment as the outcome. Columns 2 and 3 show estimates for part-time and full-time employment, respectively. All results are based on mothers living in West Germany only. All specifications include controls for age, age squared, dummies for three levels of education (no vocational training, vocational training, university degree), year-fixed effects, county-fixed effects, and the level of full-time childcare coverage on the right-hand side. Standard errors in parentheses are clustered at the county level.