TI 2026-014/IV
Tinbergen Institute Discussion Paper

# Work Requirements in Subsidized Child Care and Maternal Labor Supply

*Hans van Kippersluis*[1]
*Hongliang Zhang*[2]

1 Erasmus University Rotterdam, Tinbergen Institute

2 Zhejiang University

# Work Requirements in Subsidized Child Care and Maternal Labor Supply

Hans van Kippersluis[1,2] and Hongliang Zhang[3]

[1]Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands

[2]Tinbergen Institute, The Netherlands

[3]School of Economics, Zhejiang University, China

20th March 2026

**Abstract**

This paper investigates how tightening work requirements in subsidized child care affects parental employment. Using Dutch administrative data, we examine a 2012 reform that capped subsidized child care hours at 140% of the hours worked by the lesser-working parent. We identify affected households by creating proxy treatment and control groups based on older siblings' pre-reform child care usage relative to the lesser-working parent's work hours, then employ a triple-difference framework comparing virtually constrained and non-constrained families with and without younger siblings before and after the reform. Our findings reveal a pronounced gender difference in responses to stricter work requirements: fathers showed negligible changes, while mothers responded at both employment margins. At the extensive margin, many mothers exited the labor force, resulting in a persistent 12 percentage point decrease in maternal employment among affected families. At the intensive margin, some mothers increased their work hours to retain subsidies; these effects also persist over time. Child care use fell sharply, raising concerns about child development, especially for disadvantaged families. These findings demonstrate that intensifying work requirements can have unintended negative consequences on mothers' workforce participation and family welfare, informing debates on welfare conditionality and child care policy.

# 1 Introduction

A central challenge in designing social welfare programs is balancing income support with work incentives. Work requirements — policies that tie benefits to employment — are frequently adopted to counter potential disincentives to labor force participation (e.g., Moffitt, 2002; Blundell et al., 2016). For example, the U.S. Earned Income Tax Credit (EITC) requires earned income, while the U.K. Working Tax Credit specifies explicit work hour requirements. However, work requirements in social welfare programs remain contentious (e.g., Besley and Coate, 1992; Corinth et al., 2021), as exemplified by U.S. programs like the Supplementary Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF) and Medicaid (e.g., Hahn and Haskins, 2018; Sommers et al., 2020). These debates have intensified following President Trump's One Big Beautiful Bill Act, which mandates 80 hours per month of "community engagement" (work, education, or volunteer activities) for able-bodied Medicaid recipients aged 19 to 64. Although there is substantial evidence on the introduction or elimination of work requirements in welfare programs (e.g., Enriquez et al., 2023; Gray et al., 2023; Goldin et al., 2024), the effects of *intensifying* these requirements remain largely unexplored.

In this paper, we fill this gap by examining how intensifying work requirements in subsidized child care affects parental employment, focusing on a 2012 Dutch reform that tightened work requirements for subsidized child care. Under the pre-reform system, children aged 0-4 were eligible for subsidized child care for up to 230 hours per month if both parents were employed, regardless of their work hours. The 2012 reform, introduced as part of austerity measures following the 2008-2010 financial crisis, imposed stricter work requirements by capping the maximum number of subsidized child care hours at 140% of lesser-working parent's work hours.[1] We refer to families who, absent the reform, would have used more child care hours than their entitlement as being affected by the "hours constraint." Our main objective is to examine how exposure to this constraint affects parental labor supply decisions.

Stricter work requirements can affect labor supply in opposing ways for affected families.

[1] These measures were taken to maintain the financial sustainability of the child care system (Staatsblad, 2011), and to discourage improper use of child care subsidies (Parliament, 2011). In 2011, approximately two-thirds of Dutch children aged 0-4 used child care, of which about one in ten used more child care hours than 140% of the hours worked of the lesser-working parent.

At the intensive margin, some parents may increase their work hours to meet the new threshold. At the extensive margin, however, reduced child care benefits could discourage labor force participation in the presence of a fixed cost of working (see e.g., Moffitt, 1992; Blundell et al., 1998; Ziliak, 2015). In cases where families previously relied heavily on child care with one parent working only a few hours per week under the original rules, the lesser-working parent may find it optimal to exit the workforce entirely and assume child care responsibilities under the new regulations.[2] Existing evidence suggests that these requirements are especially relevant for women, as their labor supply tends to be more responsive to tax and transfer policies than men's, making women central in both policy design and evaluation (e.g., Killingsworth and Heckman, 1986; Hotz and Miller, 1988; Attanasio et al., 2008; Blau and Kahn, 2013). Moreover, given traditional gender norms in the Netherlands (Fortin, 2005; Rabaté and Rellstab, 2022), the added child care responsibilities — and the resulting negative impact on labor force participation — are likely to disproportionately affect mothers. Therefore, while we analyze both parents, our main focus is maternal employment.

Since parents may adjust their work and child care hours in response to the 2012 rule changes, we cannot observe which families were *actually* affected by the hours constraint. This challenge resembles that of identifying *directly* affected groups in other contexts where regulatory changes induce behavioral adjustments, such as minimum wage increases (Neumark et al., 2014), bans on gender-preferential job ads (Card et al., 2025), or restrictions on outsourced labor (Jiménez and Rendon, 2025). In these settings, researchers have employed demographics-based (e.g., Allegretto et al., 2011; Guiliano, 2013; Jiménez and Rendon, 2025) or prediction-based approaches (e.g., Cengiz et al., 2022; Card et al., 2025) to analyze effects on specific, exogenously determined groups highly likely to be affected by the regulatory change.

Our baseline analysis uses a demographics-based approach, identifying groups highly exposed to the 2012 hours constraint through observable predetermined family characteristics. For firstborn children, predicting whether the hours constraint will bind turns out to be extremely difficult, as child care use cannot be reliably predicted from parents' pre-parenthood characteristics alone. However, for later-born children, an older sibling's child

[2] Additionally, the reform means that for families with one parent out of the labor force, the costs of re-entering the job market have increased significantly.

care usage accurately predicts the younger sibling's child care patterns and, consequently, their exposure to the hours constraint. We therefore focus our analysis on families with multiple children where both parents' work hours and older siblings' child care hours were observed before 2012. We establish a *proxy* treatment group comprising families where the hours constraint was virtually binding for an older sibling before 2012 (indicating the younger sibling is highly exposed to the constraint), referred to as "virtually constrained families." Similarly, we define a *proxy* control group of families where the constraint was not virtually binding for an older sibling (indicating the younger sibling is minimally exposed), referred to as "virtually non-constrained families." In our robustness analysis, we also employ a prediction-based approach using machine learning techniques to estimate the likelihood that the constraint will bind for younger siblings, based on the virtual binding status of their older siblings and predetermined family-level socio-demographic characteristics.

Figure 1 intuitively illustrates our empirical strategy. Panel 1a displays employment trends for mothers with a toddler-age younger child, distinguishing between virtually constrained and unconstrained families. Mothers in virtually constrained families experienced a decline in employment of over 4 percentage points after 2012, while those in non-constrained families saw only a slight reduction of less than 1 percentage point. However, a conventional difference-in-differences (DD) approach may yield unreliable estimates of the reform's impact if labor market dynamics vary between virtually constrained and non-constrained families with younger children. To address potential non-parallel trends, we adopt a triple-difference (DDD) approach, adding families without younger children as an additional control group. Panel 1b shows that mothers without younger children did not experience significant employment changes after 2012, regardless of the older sibling's virtual constraint status before 2012. Panels 1c and 1d present similar diagrams for fathers, indicating that fathers in virtually constrained families with younger children did not experience a comparable employment decline as mothers. Taken together, the findings in Figure 1 preview our main result: stricter work requirements for subsidized child care reduced employment rates specifically among affected mothers while having minimal impact on fathers.

Figure 2 compares the distribution of mothers' weekly work hours across five intervals in 2011 (one year before the reform) and 2014 (two years after). We again categorize families by whether they were virtually constrained and whether they had a younger child. Comparing virtually constrained mothers with younger children (top-left panel) with the other panels, we observe notable increases in two categories: not working at all and working 28+ hours per week. The increase in non-working mothers aligns with the extensive-margin maternal employment response shown in Figure 1. Meanwhile, the growth in the 28+ hours category suggests that the reform incentivized some mothers to work more hours to meet the eligibility requirements for full-time child care subsidies.

In our empirical analysis, we integrate triple-difference strategies with distribution regressions (e.g., Chernozhukov et al., 2013; Dube, 2019; Biewen et al., 2022) to evaluate the reform's effects on the distribution of parents' work hours, earnings, and taxable income. The DDD estimator measures the intent-to-treat (ITT) effect of the reform as the change in outcomes between virtually constrained and non-constrained families with younger children. However, the setting is more nuanced: not all virtually constrained families are *actually* bound by the hours constraint for their younger children after the reform, and a small portion of virtually non-constrained families become bound. Thus, the difference in binding rates between the two groups is substantial but not sharp. For fuzzy DD settings, De Chaisemartin and d'Haultfoeuille (2018) propose a Wald estimator derived as the ratio of the DD in outcomes to the DD in treatment rates. Building on this fuzzy DD framework and incorporating a methodology proposed by Botosaru and Gutierrez (2018) to address unobservable post-reform treatment status, we introduce a Wald-DDD estimator. Specifically, we compute this Wald-DDD estimator by dividing the DDD estimate of the ITT effect by the *pre-reform* difference in binding rates between virtually constrained and non-constrained families with younger children. Under certain assumptions, we show that this Wald-DDD estimator equals a weighted difference of the average treatment effect on the treated (ATT) in virtually constrained families and the ATT in virtually non-constrained families. Moreover, if the ATT is homogeneous across treated families in both groups, the estimator recovers the ATT.

Our results show that tightening work requirements for subsidized child care led to a sizable 12 percentage point drop in employment among affected mothers. This finding is remarkable given that the government justified the 2012 reform as an austerity measure with no negative impact on labor supply (Parliament, 2011). Quitting the labor force reduces the societal tax base and can cause irreversible damage to young mothers' human capital development, economic self-sufficiency, and household bargaining power (Goldin and Mitchell, 2017). Indeed, we find evidence that the negative labor supply effects persist even after children enter primary school. Our findings also show significant reductions in child care use among affected families, raising concerns about children's development, especially for those from disadvantaged backgrounds (e.g., Elango et al., 2015; Hotz and Wiswall, 2019). At the same time, the reform increased the likelihood that mothers work more than three days a week, effects that also persist over time. These contrasting effects — reductions in employment at the extensive margin, increases in the upper end of the working hours distribution, and decreases in child care use — were likely unintended, offering valuable insights for future policy discussions on work requirements in social welfare programs.

Our paper contributes to three literatures. First, we build on research examining the labor supply effects of work requirements in social welfare programs (e.g., Blank, 2002; Grogger and Karoly, 2005). In the U.S., programs like the EITC and Child Tax Credit (CTC) encourage workforce participation by limiting benefits to taxpayers with earned income. Similarly, TANF and SNAP have federal work requirements, though implementation varies by states. While there is consensus that introducing work requirements reduces program participation, findings on employment effects are mixed. Although earlier studies find modest increases in employment from imposing work requirements (e.g., Fang and Keane, 2004; Chan, 2013; Chan and Moffitt, 2018; Guner et al., 2020; Harris, 2021), recent research reports no employment effects of introducing work requirements in SNAP (Gray et al., 2023; Cook and East, 2024) and Medicaid (Sommers et al., 2019). In contrast, studies on eliminating work requirements — such as in the CTC in some states or exemptions from work requirements in SNAP — generally show small negative labor supply effects (e.g., Han, 2022; Enriquez et al., 2023; Goldin et al., 2024).

Second, our paper relates to the literature on the effects of child care policy on maternal labor force participation [3,4] In countries with relatively low child care costs (e.g., Havnes

[3]For examples, see, e.g., Black et al. (2014); Bettendorf et al. (2015); Givord and Marbot (2015); Schlosser (2005); Lefebvre and Merrigan (2008); Lundin et al. (2008); Baker et al. (2008); Cascio (2009); Havnes and Mogstad (2011); Kosonen (2014); Nollenberger and Rodríguez-Planas (2015); Andresen and Havnes (2019); Pepin (2024)

[4]A related literature studies child home care subsidies — payments to parents who stay home — and their effects on maternal labor supply. Evidence from Nordic countries and Germany shows that such subsidies reduce maternal employment (e.g., Naz, 2004; Schøne, 2004; Giuliani and Duvander, 2017; Gathmann and Sass, 2018; Collischon et al., 2022; Gruber et al., 2025; Kosonen, 2014). In contrast,

and Mogstad, 2011; de Boer et al., 2022, for Norway and the Netherlands, respectively) and countries where other family members commonly provide child care (e.g., Attanasio et al., 2022, for Brazil), increasing subsidized child care may simply replace informal arrangements with minimal impact on maternal labor supply. Effects are more positive in countries where child care costs are high or labor force participation is low (Hotz and Wiswall, 2019; Albanesi et al., 2023; Harris and Patacchini, 2024). Several recent studies have evaluated how improving the affordability and access to child care affects gender inequality (Cortés and Pan, 2019; Kleven et al., 2024; Andresen and Nix, 2025; Hermes et al., 2024; Huber and Rolvering, 2023). The evidence generally suggests that expanding subsidized child care access can reduce labor market gender inequality in the short term, though long-term effects remain uncertain.[5]

At the intersection of the work requirement and child care policy literatures, Kuka and Shenhav (2024) find that increased EITC work incentives at first birth encourage mothers to return to work sooner, accumulate more work experience, and earn higher long-term wages. Our paper investigates the same nexus of work requirements and child care policy, but through a reverse lens: the impact of restricting child care access via stricter work requirements. The 2012 Dutch reform provides a distinct contrast to previous research in both domains. While existing empirical literature on work requirements in social welfare programs typically examines imposing or eliminating requirements at the extensive margin or increasing program benefits, we focus on tightening work requirements at the intensive margin.[6] Similarly, while prior child care policy research concentrates on expanding access, our study uniquely investigates the impact of reducing child care subsidy generosity.

Finally, we provide evidence that short-term job incentives can have persistent effects on maternal labor supply. Mothers who leave their jobs in response to a tighter work requirement for childcare subsidies are significantly less likely to be employed even after

subsidies for part-time work increase post-birth labor supply (Zimmert and Zimmert, 2024), and access to affordable live-in domestic helpers raises maternal employment in Hong Kong relative to Taiwan (Cortés and Pan, 2013).

[5] An exception is Kleven et al. (2024), who find a precisely estimated zero effect of expanding child care access in Austria on maternal employment.

[6] The UK did introduce a tighter work requirement in April 2012 for working tax credit eligibility: the minimum work requirement for one parent remained at 16 hours, yet couples with children had to work 24 hours between them. While the Institute for Fiscal Studies (IFS) published some policy reports regarding the ex-ante predicted impacts on household income (Browne, 2011; Joyce, 2012), to the best of our knowledge there is no ex-post evaluation of the work requirement reform.

their children enter primary school. Conversely, mothers who increased work hours in the immediate aftermath of the reform exhibit a persistent increase in the probability of working more than 28 hours per week. These results suggest that short-run incentives can shift long-run labor supply trajectories. Our findings complement existing evidence on persistent labor supply responses to policy incentives and labor market shocks, including work incentives in the Earned Income Tax Credit (e.g., Chetty et al., 2013; Kuka and Shenhav, 2024), tax reforms (e.g., Blundell et al., 1998), welfare reforms (Card and Hyslop, 2009; Fang and Keane, 2004), and early-career shocks such as graduating during a recession (e.g., Oreopoulos et al., 2012; Von Wachter, 2020).

The rest of the paper is organized as follows. Section 2 introduces the institutional setting and child care reforms in the Netherlands. Sections 3 and 4 describe the data and reduced-form identification strategy for estimating the intention-to-treat (ITT) effect. Section 5 presents the ITT results alongside a battery of robustness checks. Section 6 develops the identification strategy and reports the ATT estimates. Section 7 provides a brief discussion and concludes.

# 2 Institutional Background

## 2.1 Labor market and child penalty

The Dutch labor market is characterized by high participation rates and widespread part-time work, a pattern most evident among women who have higher initial employment rates than men upon entering the labor market but work fewer hours.[7] Consistent with evidence from other countries (e.g., Angelov et al., 2016; Lundborg et al., 2017; Kleven et al., 2019a; Sieppi and Pehkonen, 2019; Andresen and Nix, 2022), women experience a significant, immediate, and lasting motherhood penalty after childbirth, while men remain largely unaffected (Rabaté and Rellstab, 2022). Online Appendix Figure A.1 replicates Rabaté and Rellstab (2022), showing that substantial gender inequalities in labor supply and earnings arise in the Netherlands around the birth of the first child. The size of this

[7]Part-time work is much more common in the Netherlands than in other OECD countries. About 60% of Dutch women and 20% of Dutch men work part time – almost three times and twice the OECD averages, respectively (OECD, 2019) This stems from the wide availability of quality part-time jobs, comparable conditions to full-time roles, and the financial viability of living on 1.5 incomes (Merens and Bucx, 2018).

penalty in the Netherlands is comparable to those in Austria and Germany, which have the largest child penalties among the countries studied by Kleven et al. (2019b).[8]

## 2.2 Child care

In 2005, the Netherlands established a Child Care Act (*Wet Kinderopvang*) with two main aims: (i) to help parents balance work and child care responsibilities, and (ii) to provide high-quality child care that meets both parents' needs and supports children's development (SZW, 2015). Under the law, parents pay the full fee of child care but receive government subsidies for formal child care when both parents are working.[9] Formal child care operates as a regulated private market with limited competition (Bansraj and Xu, 2024), and includes day care for children up to age 4 in daycare centers, and after-school care for children aged 4 to 12.[10] Children are subject to compulsory schooling laws from age 5, though virtually all children enter school on their 4th birthday. In this paper, we focus on children up to age 4 and thus exclusively on daycare services. In the Netherlands, daycare access is nearly universal, with more than 99% of children having access to one or more facilities within a 3km radius of home (Van der Wiel and Van 't Riet, 2011; CBS, 2016), and this accessibility does not vary systematically with income (SZW, 2012).

The child care subsidy $S_{ct}$ for child $c$ in year $t$ can be summarized using the following formula:

$$S_{ct} = \min\left(HC_{ct}, HC_{ct}^{max}\right) \times \min\left(P_{ct}, P_t^{max}\right) \times \tau_{ct}\left(I_{it}, I_{pt}\right), \tag{1}$$

where $HC_{ct}$ is the hours of child care for child $c$ in year $t$, $HC_{ct}^{max}$ is the maximum hours of care subsidized in year $t$, $P_{ct}$ is the price per hour for child $c$ in year $t$, $P_t^{max}$ is the maximum hourly price in year $t$, and $0 \leq \tau_{ct} < 1$ is the fraction subsidized, which depends on the mother's taxable income $I_{it}$ and the subsidy partner's taxable income $I_{pt}$. The subsidy partner is defined as the person you are married to, your registered partner,

[8]See Online Appendix B.2 for a discussion of parental leave policies around childbirth.

[9]Parents who become unemployed remain eligible for child care subsidies for a period of three months. Parents participating in active labor market policies are also eligible for child care subsidies. However, we do not observe this in our data.

[10]An alternative form of formal child care is "host parent care", where someone other than the parent cares for up to 6 children, either in their own home or the children's home. This small-scale, flexible type of care is typically provided by women aged 50-59 (Van der Wiel and Van 't Riet, 2011). To qualify as formal child care, the host parent must be registered in the municipality's child care supplier registry. Grandparents served as host parents in 25-30% of cases, but this percentage decreased in 2010 due to tighter regulations (Intomart/GfK, 2011), see Online Appendix B.1 for details.

or the person you live with and have a child together. The subsidy rate $\tau_{ct}$ additionally varies by the child's rank within the family in terms of child care hours. That is, a lower subsidy rate applies to the child who uses the most child care hours compared to other children in the same household (see Online Appendix B.1 for details). Since the child with the most child care hours within the family is typically the youngest child, we refer to it as the youngest child in the remainder of the paper.

### 2.2.1 The 2012 work hours requirement reform

Before 2012, the maximum number of yearly subsidized hours $HC_{ct}^{max}$ was very generous at 2,760 hours (i.e., 230 hours per month × 12 months), which in practice was not binding for any family.[11] After 2012, the maximum number of hours eligible for subsidies $HC_{ct}^{max}$ was reduced to 140% of the hours worked by the lesser-working parent. Applying the 2012 rules to earlier years, we find that around 6% of all parents — or 10% of the parents who were using child care — were virtually affected by the 2012 reform. This reform was announced in June 2011 and implemented on January 1, 2012. Online Appendix Figure A.2 shows a clear, immediate drop in maternal employment rates among the families who were most affected by the reform starting in early 2012, suggesting that parents understood the reform and responded immediately.

### 2.2.2 Other changes in the subsidy formula

Since 2005, the Dutch government made several additional adjustments to elements of the subsidy formula in Equation (1).[12] The government reduced child care subsidy rates for higher-parity children in 2012 and extended these reductions to the youngest child in 2013 (see Online Appendix Figure A.4 and Mari and Marie (2026) for an evaluation of this reform).[13] These subsidy adjustments affected a population distinct from that exposed to our primary treatment. The cuts were concentrated among high-income households: in 2013, subsidy rates fell most sharply for families earning above €118,189 (6% of all

[11] In our sample, only 0.01% used more than 2,760 hours of care per year. These families were using host parent care in addition to daycare, which was in turn also subsidized up to a maximum of an additional 2,760 hours yearly. Hence in practice, not a single family was effectively bound by this constraint.

[12] Online Appendix Figure A.3 shows the maximum subsidizable hourly price for daycare between 2007 and 2014, which was usually adjusted for inflation.

[13] Earlier subsidy expansions (2005–2007) increased labor force participation by about 5% and formal child care use by roughly 10%, largely reflecting substitution from informal care (CPB, 2008; Berden and Kok, 2009; Bettendorf et al., 2015)

families; see Online Appendix B.1). By contrast, the work requirement affects families with relatively low parental work hours compared to child care hours. We further limit potential confounding in two ways. First, the main analysis restricts the sample to households with joint incomes below €90,000, with a robustness check below €60,000, thereby excluding the income ranges most affected by the subsidy reforms. Second, directly controlling for exposure to contemporaneous subsidy changes leaves the results essentially unchanged (see Online Appendix E).

### 2.2.3 Evolution of child care over time

Figure 3 displays time series of government expenditures on child care (panel 3a), number of child care workers (panel 3b), subsidized child care use (panel 3c), and subsidized hours of child care (panel 3d). The year 2012 marked a striking trend break in child care policy and usage. Total government expenditures dropped from approximately €3 billion to less than €2 billion between 2011 and 2014. This substantial cut in government spending was associated with sharp reductions in subsidized child care use at both extensive and intensive margins. For example, the average hours of subsidized care decreased from 950 in 2011 to less than 800 in 2014. Part of this decline may be mechanical — the 2012 reform directly capped the maximum hours of subsidized child care to 140% of the hours worked by the lesser-working parent. However, in Online Appendix C.2, we confirm through survey data that actual child care use closely tracks subsidized child care use and also experienced a substantial drop after 2012. Moreover, Online Appendix C.1 presents a DD analysis showing that the reduction in child care use was significantly larger among virtually constrained families, consistent with greater exposure to the policy. The reduced demand for child care subsequently led to a contraction in supply, with the number of child care workers declining by roughly 20,000 (or 13%) between 2011 and 2014. Taken together, these developments underscore the magnitude of the 2012 reform and motivate our analysis of its employment effects.

# 3 Data and descriptive statistics

## 3.1 Data

The data for this paper consist of a series of population-wide administrative registers collected by Statistics Netherlands. These registers can be linked through an encrypted social security number and encompass all inhabitants in the Netherlands, totaling approximately 17 million observations annually. We use 8 distinct registers covering the period from 2007 to 2014, which can be categorized into 4 groups:

**Socio-demographics:** year and month of birth, gender, country of birth, current partner and marital status from the municipality register, and educational attainment from a separate education register.

**Intergenerational linkages:** all children are linked to their parents.

**Work and income:** current employment status (working, unemployed, disabled, retired, etc.), hours of work, earnings, and taxable income derived from tax authority registers.

**Child care:** subsidized hours of child care for each child, type of care for each child, total costs incurred by each family, and the total subsidy received by each family.

## 3.2 Sample construction

Our identification strategy exploits variation from a pre-reform exposure proxy — whether a family was virtually constrained by the 2012 hours requirement for an older sibling at age 2 — constructed using data on that sibling's subsidized child care hours and both parents' work hours. To ensure data availability and a clean identification strategy, we apply the following sample restrictions. First, we include only families with a singleton child born between 2005 and 2009 (hereafter the "focal child"), whose second birthday fell between 2007 (when our subsidy records begin) and 2011 (the year before the reform). Second, both parents must have resided in the Netherlands and been employed as wage earners when the focal child aged 2, ensuring work hours information is available.[14] Third, to isolate the effect of the 2012 hours constraint from concurrent subsidy rate changes for high-income families, we limit to families with a joint annual income below €90,000 when the focal child was two — a threshold covering more than 90% of all families.[15] Finally, we restrict our analysis to 2009–2014. This window begins after the earliest cohort of focal children in our sample — those born in 2005 — turned 4 and enrolled in primary school, and ends before universal subsidy rate increases implemented in 2015.

We divide families into two groups based on whether the focal child has a younger sibling. The first group, the "YC" sample, consists of families with an additional younger child (21.6% of the whole sample). To avoid contamination while retaining the relevance of the proxy, we include only families where the age difference between the younger and older sibling was 3 years. This ensures that the child care arrangements for the focal child at age 2 were not influenced by having another younger child. We focus on periods when younger children aged 1–3 (with the corresponding focal child aged 4–6 and enrolled in primary school) when child care needs were most pressing. The second group, the "NYC" sample, comprises families where the focal child does *not* have a younger sibling, meaning these families were not exposed to the 2012 reform through any younger child. As with the "YC" sample, we restrict observations to when the focal child aged 4 to 6. In defining these two samples, we assume that fertility decisions are not influenced by child care reforms — an assumption we empirically verify in Online Appendix Section E.1.

Applying these sample selection criteria yields a final sample of 224,442 families — defined by focal child — observed during 2009–2014. Since each family contributes up to three yearly observations (when the focal child was aged 4 to 6), our total sample comprises 602,035 family-year observations.

[14] The reason for excluding self-employed parents is that we do not observe their hours of work, and hence cannot determine the virtual binding status of the 2012 hours constraint to their focal child at age 2. Some parents not in the labor force might still be eligible for child care subsidies if they were enrolled in labor market training or other types of education, but since we do not observe these types of training, we also cannot determine their eligibility for child care subsidies.

[15] In Online Appendix C.3, we compare our main analysis sample to the full sample of families with children born between 2005 and 2009. The results show that our sample restrictions create predictable and relatively small differences. Our focus on families where both parents were working when the focal child was 2 years old makes our sample somewhat more selective compared to the full sample, where 82% of mothers and 96% of fathers were working. Additionally, our main analysis sample also excludes the 22% of fathers and 11% of mothers who were partially or fully self-employed when the focal child was 2. While average education levels and earnings are slightly lower in our sample, demographic characteristics remain very similar to those in the full sample.

### 3.3 Variables and descriptive statistics

**Dependent variables:** Our main outcomes of interest focus on maternal employment. For the extensive margin, we use binary indicators for mother's employment status, which encompasses both regular employment and self-employment. For the intensive margin, we examine hours worked and earnings, measured as annual salaried hours and annual labor earnings summed across all jobs. Since work hours and earnings are only recorded for employees and not for the self-employed, we also analyze taxable personal income (including income from self-employment and other sources) to account for this data limitation.

**Treatment and control variables:** As we will elaborate in Section 4, we proxy treatment status based on whether the older sibling faced a virtually binding hours constraint at age 2. Online Appendix Figure A.5 illustrates the predictive value of the child care use of an older sibling for a younger sibling. The figure shows that before 2012, in families with multiple children, the older sibling's child care hours reliably predicted a younger child's child care hours at age 2. We will present a formal first stage in Section 6.

Control variables include the focal child's rank and age, mother's and father's age and education, and a list of variables measured when the focal child was aged 2: whether the parents used child care, child care hours, mother's and father's work hours, earnings, and taxable income.

**Descriptive statistics:** As discussed in Section 1, Figures 1 and 2 present the evolution of employment outcomes over time, stratified by whether families were virtually constrained. Table 1 presents descriptive statistics of the control variables separately for the YC sample (columns 1–3) and NYC sample (columns 4–6). Each sample is divided into two subgroups: (i) virtually constrained families where the work hours requirement was binding for the focal child at age 2 pre-2012 (columns 1 and 4), and (ii) virtually non-constrained families where the requirement was not binding (columns 2 and 5). The differences between constrained and non-constrained families for each sample are shown in columns 3 and 6.

In the YC sample, families virtually constrained for the older sibling are 27 percentage points more likely to be constrained for the younger sibling than those not virtually constrained, confirming that our proxy for treatment status is highly predictive. By construction, virtually constrained YC families have significantly higher hours of care and lower maternal work hours and earnings than non-constrained families. Importantly, these predictable differences between virtually constrained and non-constrained YC families closely mirror those between virtually constrained and non-constrained NYC families (see columns 3 and 6 of Table 1). This pattern indicates that while virtually constrained and non-constrained families have different characteristics — and may experience different labor market dynamics — these differences remain consistent across both the YC and NYC samples. This consistency provides the foundation for our triple-difference strategy outlined in Section 4. Furthermore, it is important to note that our virtually constrained group is not exclusively composed of families with very low working hours for the lesser-working parent and very high child care hours. As Online Appendix Figure A.6 shows, virtually constrained families span the entire spectrum of work and child care hours.

# 4 Reduced-form Identification

This section outlines our reduced-form identification strategy. Section 4.1 introduces a DD estimator in a simple two-period framework, which compares outcome changes for YC families based on whether an older sibling was virtually constrained by the hours requirement. Section 4.2 introduces a DDD estimator that leverages NYC families as an additional control group to help account for potential non-parallel trends in outcomes related to the older sibling's virtual constraint status. Section 4.3 generalizes the DDD strategy to a multi-period framework with covariates and extends it to evaluate the reform's effect on the distribution of outcomes.

## 4.1 DD estimator

To evaluate the impact of the 2012 reform, a straightforward approach would be to compare affected and unaffected families before and after implementation. However, treatment status — whether a family would have used child care hours exceeding 140% of the lesser-working parent's work hours absent the reform — becomes unobservable after 2012, as families adjust their work and child care choices in response to the policy change. As noted in Section 1, this observational challenge resembles those encountered in studies of other regulatory changes that induce behavioral adjustments relevant for determining treatment eligibility (e.g., Neumark et al., 2014; Card et al., 2025; Jiménez and Rendon, 2025). In such settings, researchers address this challenge by focusing on groups defined by characteristics predetermined before the policy change, which remain exogenous to behavioral responses. Following this approach, we restrict our DD analysis to families where both parents' work hours and the older sibling's child care hours were observed prior to 2012, allowing us to classify families as virtually constrained or non-constrained based on characteristics observed before the reform.

For ease of illustration, we first present our identification strategy using a simplified two-period, cross-sectional framework without covariates. The full empirical specification with multiple periods and covariates is developed in Section 4.3. Let $Y$ denote the outcome, $T$ the time period ($T = 0$ pre-reform, $T = 1$ post-reform), and $G$ the group indicator ($G = 1$ for YC families, $G = 0$ for NYC families). To address unobservable treatment status after the reform, we use whether the older sibling was virtually constrained by the hours requirement before the reform ($Z$) as a proxy assignment variable for the younger sibling's treatment status ($D$).

Our benchmark DD estimator measures the difference in outcome changes between virtually constrained ($Z = 1$) and non-constrained ($Z = 0$) families, among those with a younger child ($G = 1$):

$$\widehat{\rho}_{DD} = \left(\bar{Y}_{G=1,Z=1,T=1} - \bar{Y}_{G=1,Z=1,T=0}\right) - \left(\bar{Y}_{G=1,Z=0,T=1} - \bar{Y}_{G=1,Z=0,T=0}\right), \tag{2}$$

where $\bar{Y}_{G=g,Z=z,T=t}$ denotes the conditional mean outcome for families with group status $g$ and proxy value $z$ in time period $t$, for $g, t, z \in \{0, 1\}$. For the DD estimator in Equation (2) to identify the reform's ITT effect on the $Z = 1$ subgroup of YC families against the $Z = 0$ subgroup, it relies on the standard DD parallel trends assumption: in the absence of the reform, outcome changes would have been identical between virtually constrained and non-constrained YC families.

## 4.2 DDD estimator

Table 1 reveals significant differences between the $Z = 1$ and $Z = 0$ subgroups of YC families across several characteristics, including immigration background, education, work hours, and child care hours. Since these subgroups consist of fundamentally different families, the standard parallel trends assumption may not hold. To address this concern, we introduce an additional control group — NYC families ($G = 0$) — to account for potential differences in outcome dynamics between virtually constrained and non-constrained families. This allows us to construct a DDD estimator that measures the difference in relative outcome changes based on the focal child's virtual binding status between YC and NYC families:

$$\begin{aligned}\widehat{\rho}_{DDD} =& \Big[\big(\bar{Y}_{G=1,Z=1,T=1} - \bar{Y}_{G=1,Z=1,T=0}\big) - \big(\bar{Y}_{G=1,Z=0,T=1} - \bar{Y}_{G=1,Z=0,T=0}\big)\Big] \\ &- \Big[\big(\bar{Y}_{G=0,Z=1,T=1} - \bar{Y}_{G=0,Z=1,T=0}\big) - \big(\bar{Y}_{G=0,Z=0,T=1} - \bar{Y}_{G=0,Z=0,T=0}\big)\Big].\end{aligned} \tag{3}$$

The key advantage of the DDD estimator is that it relaxes the parallel trends assumption required in the DD framework (Gruber, 1994; Olden and Møen, 2022). By incorporating NYC families as an additional control group, we allow outcome dynamics to differ between the $Z = 1$ and $Z = 0$ subgroups, as long as this variation is identical between YC and NYC families (Ortiz-Villavicencio and Sant'Anna, 2025).

## 4.3 Generalized DDD estimation

The empirical specification we use to estimate $\widehat{\rho}_{DDD}$, incorporating multiple periods of repeated cross-sectional data and covariates, is as follows:

$$\begin{aligned}Y_i =& \sum_{t=2009}^{2014} \delta_t \mathbf{1}(T_i = t) + \sum_{t=2009}^{2014} \kappa_t\big(G_i \times \mathbf{1}(T_i = t)\big) + \sum_{t=2009}^{2014} \theta_t\big(Z_i \times \mathbf{1}(T_i = t)\big) + \tau\big(G_i \times Z_i\big) \\ &+ \rho\big(G_i \times Z_i \times \mathbf{1}(T_i \geq 2012)\big) + f\big(X_i; Z_i, G_i, T_i\big) + \varepsilon_i\end{aligned} \tag{4}$$

where the coefficients $\delta_t$, $\kappa_t$, $\theta_t$ represent the calendar year fixed effect and its interactions with the group dummy $G_i$ and proxy treatment assignment dummy $Z_i$, and $\tau$ represents the two-way interactive fixed effect between $G_i$ and $Z_i$. The coefficient $\rho$ on the three-way interaction term $G_i \times Z_i \times \mathbf{1}(t \geq 2012)$ identifies the reform's ITT effect on the $Z = 1$ subgroup of YC families ($G = 1$) relative to the $Z = 0$ subgroup of YC families. Finally, $f\big(X_i; Z_i, G_i, t\big)$ denotes all main effects of covariates $X_i$, as well as their two-way interactions with calendar year, group, and proxy treatment assignment dummies.[16]

To examine how the reform affects the distribution of family outcomes (e.g., maternal work hours, earnings), we integrate the DDD strategy with the distribution regression approach developed by Chernozhukov et al. (2013) to analyze the effect of the reform on the cumulative distribution function (CDF) of the respective outcomes, $F(y) = p\,(Y \leq y)$, calculated as the proportion of samples below threshold $y$.[17] The DDD estimates of these distributional regressions jointly reveal how the reform shifts the overall distribution of a certain outcome. For mothers' work hours specifically, this approach helps us determine whether the extensive margin response in labor force participation shown in Figure 1 is offset elsewhere in the work hours distribution.[18]

# 5 ITT Effects

## 5.1 Immediate effects

Table 2 presents our analysis in three panels. We first report separate DD estimates for families with a younger child (Panel A) and families without a younger child (Panel B), followed by our preferred DDD estimates using the pooled sample (Panel C). Columns 1–3 examine maternal employment at the extensive margin. For the YC sample, the basic DD estimate without control variables (Column 1) shows a 3.8 percentage point decline in the probability of maternal employment. Adding controls to address potential differential

[16]In Online Appendix E.3 we show that our results are robust to including control variables using the doubly robust DDD estimator proposed by Ortiz-Villavicencio and Sant'Anna (2025). They show that, with covariates, the conventional triple-differences regression is generally no longer equivalent to the simple difference of two difference-in-differences estimators, and propose doubly robust estimators that remain consistent if either the outcome or propensity model is correctly specified.

[17]The primary focus of our distributional treatment effect analysis is on the nominal levels of the outcome measures (Biewen et al., 2022), rather than the quantiles as in Firpo et al. (2009), Havnes and Mogstad (2015), or Dube (2019), since the 2012 reform is itself targeted at nominal levels of child care hours and work hours and weekly working hours tends to concentrate on particular spikes. Following Dube (2019), we convert the estimates on the CDF to unconditional quantile partial effects (UQPE) through scaling by the probability density at a particular cut-off. Results are available upon request.

[18]When applying the DDD estimator to analyze distributional treatment effects, the identification assumption becomes that the relative change in population shares at a given outcome level comparing the $Z = 1$ and $Z = 0$ subgroups would be the same between YC and NYC families. Take work hours as an example: while pre- and post-reform changes in the share of individuals with specific work hours may differ between virtually constrained and non-constrained families, this difference should remain stable regardless of whether these families have a younger child. A nice property of the distribution regression is that the identification assumption is invariant to monotonic transformations of the outcome, see e.g., Havnes and Mogstad (2015) for an example of a logarithm transformation of earnings.

dynamics yields similar estimates: −3.8 percentage points with a saturated set of controls (Column 2) and −4.0 percentage points using post-double selection lasso (Belloni et al., 2014) to select controls (Column 3). Panel B reports the corresponding DD estimates for the NYC sample, which serves as a placebo test since families without a younger child are by definition unaffected by the 2012 reform.[19] Whereas the most basic specification without control variables shows a small but statistically significant reduction in maternal employment of 0.8 percentage points, this effect disappears when adding a saturated or lasso-selected set of control variables (Columns 2 and 3). The DDD estimates using the pooled YC and NYC samples in Panel C suggest an ITT effect of the reform of around 3.0–3.5 percentage point reduction in maternal employment, robust across specifications regardless of the inclusion and form of controls.

Columns 4–6 of Table 2 report the corresponding results for mothers' weekly working hours. The statistically significant negative employment effect estimated for the extensive margin disappears when the extensive and intensive margins are combined. Although the point estimates remain negative for weekly working hours, they are never statistically significant, suggesting that positive effects at certain parts of the working hours distribution counteract the negative effect on the extensive margin. To analyze effects across the entire distribution, we present DDD estimates on the cumulative distribution of weekly working hours. Figure 4a shows the corresponding estimates and 95% confidence intervals. Consistent with the extensive margin results, the probability of working zero hours per week increased by 3 percentage points. Significant positive effects on the cumulative probability persist up to 10 hours per week, indicating that the reform increased the likelihood of working less than 10 hours per week. Interestingly, the point estimates decline to 0 at 16 hours per week and turn negative hereafter, with statistically significant estimates at around 24 and 32 hours per week. This pattern reveals that the reform increased the probability of mothers working more than 24 hours (approximately 3 days) and 32 hours (approximately 4 days) per week. This shift from positive to negative point estimates explains why we find no effect on average weekly working hours.

Consistent with the pattern for weekly working hours, we find no effect on mean maternal

[19] Although the 2012 reform also imposed an hours constraint on after-school care subsidy for school-age children (see Section 2), school-age children are likely to be minimally affected by such constraints. Our preferred DDD estimation also accounts for the influence of the after-school care hours constraint through two-way interactions between the virtual constraint indicator $Z = 1$ and calendar year dummies.

earnings (Columns 7–9, Table 2), but a clear divergent response when examining the cumulative distribution (Figure 4b). Specifically, the reform increased the probability of mothers having zero annual earnings by approximately 3 percentage points, while simultaneously reducing the probability of earning less than €20,000 — in other words, increasing the likelihood of mothers earning more than €20,000 per year. A remarkably similar distributional shift is evident for mothers' taxable income (Online Appendix Figure A.7), suggesting that mothers exited the labor market voluntarily; otherwise we would expect taxable income to be cushioned by unemployment benefits or offset by moves into self-employment — a channel we rule out in Online Appendix E.1.

Taken together, the evidence points to a dichotomous ITT effect: the reform induced some mothers to withdraw from the labor force while prompting others to increase their work hours and earnings. In the following sections, we first assess the robustness of these findings (Section 5.2) and then examine how this heterogeneity relates to observable family and maternal characteristics (Section 5.3). We reserve the conversion from the ITT effect estimated above to the ATT for Section 6, where we address the imperfect treatment compliance.

## 5.2 Robustness

**Parallel trends assumption:** Our DDD estimator identifies the reform's ITT effect on the $Z = 1$ subgroup of YC families relative to the $Z = 0$ subgroup of YC families under the assumption that, absent the reform, differences in maternal labor supply trends between virtually constrained ($Z = 1$) and unconstrained ($Z = 0$) families would have evolved in parallel across YC and NYC groups. While this counterfactual cannot be directly tested, we probe the assumption's plausibility by examining the pre-reform trends in an event-study specification. We replace the three-way interaction term $G_i \times Z_i \times \mathbf{1}(t \geq 2012)$ in Equation (4) with interactions between $G_i \times Z_i$ and calendar year dummies, omitting 2011 as the baseline. Figure 5 presents the results for maternal employment. The estimated three-way interaction coefficients are insignificant in all pre-reform years, supporting the parallel trends assumption. After 2012, the coefficients become significant, consistent with our main maternal employment results. To further assess the parallel trends assumption across the full distribution of maternal work hours and earnings, we conduct a placebo test. We artificially shift the reform's start date from 2012 to 2011 and re-estimate

Equation (4) using only data from 2007 to 2011 (i.e., excluding all observations from 2012 onward). Online Appendix Table A.1 and Figure A.8 present the results. None of the placebo estimates are statistically significant, indicating that no spurious effect is detected when the reform is falsely pre-dated. This finding lends support to the plausibility of the parallel trends assumption on both the extensive and intensive margins of maternal labor supply.

**Alternative definitions of treatment and machine learning prediction** The virtual treatment status of an older sibling is the strongest single predictor for treatment status for the younger sibling, and it allows for a transparent dichotomous distinction between proxy treated and control families. To assess whether this discretization is justified, we construct a continuous measure of "excess hours", defined as 1.4× the hours of the least working parent minus the childcare hours used for the older sibling at age two. Negative values indicate virtually constrained families. We partition this measure into 50 percentile bins, with the bottom three bins (i.e., the bottom 6 percentiles) corresponding to our binary treatment indicator. In turn, we estimate the intent-to-treat triple-difference specification interacting these bins with indicators for having a younger child and the post-2012 period. Online Appendix Figure A.9 plots the coefficients relative to the top bin. Estimates are flat and insignificant above the threshold, while the lowest three bins show clear reductions in maternal labor supply. The similarity of these estimates across the treated bins suggests that the binary treatment definition is an adequate approximation.

However, we can improve precision in identifying cleaner treatment and control groups by incorporating multiple predictors. We trained three machine learning algorithms — namely, random forest, gradient boosted model, and lasso — to predict virtual treatment status. These predictions are based on whether the older sibling was virtually constrained, along with a host of other socio-demographic predictors measured when the older sibling was aged 2 (see Online Appendix D for details).

Following Cengiz et al. (2022), Online Appendix Figure A.10 shows the possible levels of precision and recall for our three algorithms as well as our dichotomous indicator of whether the older sibling was virtually constrained. Given a baseline prevalence of 6% in the treated group, our baseline precision of 32% and recall of 26% imply meaningful signal relative to chance, though classification error remains substantial. The machine learning models attain higher levels of precision of treatment status when adding other predictors, and the random forest algorithm performs best (i.e., higher levels of precision and recall, as well as highest Area under the Curve — AUC, see Online Appendix D for details).

We employ the random forest predictions to identify a more precise treatment group (top 6 percentiles of the predicted probability distribution – corresponding to the empirical frequency of being affected by the reform) as well as a more precise control group (bottom 60 percentiles of the predicted probability distribution).[20] Online Appendix Table A.2 shows that the ML-predicted treatment leads to a very similar ITT effect of around 3.0–3.5 percentage points reduction in maternal employment at the extensive margin. In line with our baseline findings, Figure 6 similarly shows a very clear divergent response in working hours that is estimated with more precision. The reform led to a reduction in employment at the extensive margin, but shows a clear increase in the probability of working more than 16 hours per week. These findings are echoed in the analysis of maternal earnings. Overall, these analyses are reassuring in that the main findings are robust to an alternative and more precise definition of proxy treatment and control groups.

**Specification checks:** In Online Appendix E, we conduct a broad set of additional robustness checks that leave our main results unchanged. We find no evidence of endogenous selection through fertility or maternal self-employment responses. To rule out confounding from concurrent subsidy and tax reforms, we follow Gruber and Saez's (2002) approach and construct predicted child care subsidy rates and parental tax credits holding pre-reform household income fixed; augmenting the DDD specification with both levels and changes in these predicted policies leaves the estimates virtually unchanged. Results are robust to alternative income cutoffs, different sample restrictions, exclusion of the transition year, richer control specifications and interactions, and the use of a doubly robust DDD estimator (Ortiz-Villavicencio and Sant'Anna, 2025).

**Effects on first child at age 3** Our main analysis focuses on families with at least two children, raising the question whether the effects are specific to later-born children.

[20] Our results are robust against using other cutoffs, see Online Appendix D for details.

To address this, we study firstborns and proxy treatment status based on whether the child was virtually constrained at age 2 pre-reform. This limits the analysis to age-3 outcomes and to a difference-in-differences design. Nevertheless, the results are similar. Online Appendix Table A.3 reports a DD estimate of $-0.028$ for maternal employment. This is slightly smaller than our main ITT estimate but still statistically significant at the 1 percent level and comparable in magnitude. Child care use also responds at both the extensive and intensive margins. These findings indicate that the reform's effects are not specific to later-born children.

### 5.3 Heterogeneity

As noted in Section 3, virtually constrained ($Z = 1$) and non-constrained families ($Z = 0$) differ significantly in characteristics and prior labor market attachment. Given these compositional differences, the overall ITT estimates may obscure important heterogeneity in responses. To investigate this variation, we estimates ITT effects separately across quartiles of three key baseline characteristics: the mother's weekly work hours, her hourly wage, and her partner's income (all measured when the focal child was aged 2).[21]

**By work hours** Figure 7a presents the ITT coefficients from distribution regressions of mother's weekly working hours, estimated separately for each quartile of mother's baseline work hours. Although estimates are less precise than the full-sample results in Figure 4a, they reveal important heterogeneity: the extensive-margin reduction in employment is driven entirely by mothers in the bottom quartile of baseline work hours; the intensive-margin increases in work hours are most pronounced for mothers in the third quartile.

**By hourly wage** Figure 7b presents the ITT coefficients estimated separately for each quartile of mother's baseline hourly wage. The strongest responses appear among mothers in the lowest hourly wage quartile. They exhibited the largest decline in employment probability but also the largest increase in the likelihood of working more than 24 hours

[21] In Online Appendix Table A.4, we show that heterogeneity in ITT effects across these quartile groups is not driven by differences in the first-stage relationship between the virtual constraint on older siblings versus younger children. The first stage coefficients are very similar across quartiles of hourly wages and partner's income. For maternal hours of work, we find the largest ITT effects among the first quartile, and this is the quartile with a slightly lower first stage effect. Consequently, the observed heterogeneity in ITT effects carries over to ATT effects, and if anything is more pronounced for the ATT.

per week — a pattern consistent with the full sample. In contrast, point estimates for the highest quartile are predominantly negative (significant at 3 to 4 days per week), suggesting a rightward shift in the hours distribution. The two middle quartiles display more muted responses. These patterns suggest that the reform most strongly incentivized labor market exit for mothers with the weakest earnings potential, while prompting increased work hours among those with the strongest earnings potential.

**By partner's income** Since men are typically the primary earners in Dutch families, mother's employment responses may depend on father's income. Figure 7c presents responses by quartiles of the partner's baseline income. For mothers with partners in the lowest two income quartiles, we find no extensive-margin effect but significant increases in the likelihood of working more than 16 hours (second quartile) and 24 hours (first quartile) per week. Conversely, for mothers with partners in the top two income quartiles, we observe a strong decline in maternal employment at the extensive margin with no intensive-margin response.

## 5.4 Longer-run effects

In this subsection, we further extend the analysis through 2019 — the last pre-pandemic year — to assess the persistence of these effects. Because the youngest affected cohort consists of children born in 2012, this allows us to examine maternal labor supply outcomes up to at least age 7, well beyond the period of early childhood. Figure 8 reports DDD estimates for maternal employment. The estimates indicate no subsequent recovery in participation: the negative effect on employment remains statistically significant through 2019.

We also examine longer-run effects on the intensive margin of work hours and on maternal earnings. To maintain statistical power, we pool the post-2016 coefficients. The results are presented in Figure 9. While the estimates are imprecise, they are consistent with a persistent increase in hours worked conditional on employment, as well as higher earnings. In particular, we find suggestive evidence of sustained increases in the share of mothers working more than three days per week.

### 5.5 Effect on fathers

Online Appendix Table A.5 and Figure A.11 report our DDD estimates for fathers' employment at the extensive and intensive margins. At the extensive margin, effects are precisely estimated and close to zero, ruling out changes smaller than $-0.019$ or larger than 0.009. At the intensive margin, few distributional effects are statistically significant. We find only a marginal increase in the likelihood that wage-employed fathers work fewer than 35 hours per week and a reduction in the probability of earning less than €50,000 annually. These intensive-margin patterns should be interpreted with caution: as Online Appendix Table A.5 shows a decline in self-employment with no change in overall employment, the estimates likely reflect compositional changes rather than behavioral responses among continuously employed fathers.

The reform led to a clear widening of gender inequality at the extensive margin. To assess whether this reflects a general reduction in women's labor supply or an amplification of pre-existing household specialization, Online Table A.5 (bottom panel) reports estimates for families in which the father was the lesser-working partner when the older child was age 2. In this subsample, we find no reduction in mothers' employment and negative (though imprecisely estimated) effects for fathers, consistent with specialization-driven adjustment: the lesser-working partner is more likely to exit employment. Because mothers were the lesser-working partner for 92% families, the reform nonetheless increased aggregate gender inequality.

## 6 From ITT to ATT

So far, we have addressed the unobservability of post-reform treatment status by estimating an ITT effect that reflects the difference in the reform's effects between two subgroups of YC families categorized by the proxy treatment variable $Z$ (i.e., whether the older sibling was virtually constrained). In this section, we propose a methodology to convert the ITT estimate to the ATT and discuss the estimation results. Section 6.1 first introduces a Wald-DDD estimator in a simplified two-period, cross-sectional framework without covariates analogous to Section 4.1. Section 6.2 then extends the Wald-DDD estimator to a linear regression framework and develops a Generalized Method of Moments (GMM) estimator that incorporates multiple periods of cross-sectional data. Section 6.3 discusses the GMM estimation results.

### 6.1 Wald-DDD Estimator

The DDD approach estimates an ITT effect rather than an ATT effect because the actual treatment status $D$ does not align perfectly with the observed proxy variable $Z$. Specifically, some YC families with a virtually binding older sibling are not constrained by the hours requirement (where $Z = 1$ but $D = 0$), while a small number of YC families without a virtually binding older sibling are constrained (where $Z = 0$ but $D = 1$). In standard fuzzy DD settings, where the actual binary treatment $D$ is observed but imperfectly predicted by the proxy $Z$, De Chaisemartin and d'Haultfoeuille (2018) propose a Wald-DD estimator — the ratio of the DD estimate for the outcome to the difference in treatment take-up between groups defined by the proxy. Our setting, however, presents a further complication: we can accurately observe the younger sibling's exposure to the hours constraint ($D$) only in the pre-reform period, but not after the reform. The limitation of partial observability is formally stated in Assumption 1.

**Assumption 1** *(Partial Observability) Observability of proxy treatment assignment $Z$, actual treatment status $D$, and outcome $Y$ varies across groups and time periods as follows:*

(a) ***NYC families:*** *For both the pre-reform period ($t = 0$) and the post-reform period ($t = 1$), the joint distribution of $(Z_t, Y_t)$ is observed.*[22]

(b) ***YC families:*** *In the pre-reform period ($t = 0$), the joint distribution of $(Z_0, D_0, Y_0)$ is observed. In the post-reform period, only the joint distribution of $(Z_1, Y_1)$ is observed, whereas actual treatment $D_1$ is latent and unobservable.*

Under Assumption 1, $D_1$ is unobserved for YC families post-reform due to endogenous behavioral adjustments. To address this observational limitation, we adopt an approach from Botosaru and Gutierrez (2018), applying the conditional distribution $P(D_0|Z_0)$ for YC families from period 0 to their marginal distribution $P(Z_1)$ in period 1. This enables us to estimate their probability of treatment $P(D_1)$ in period 1 without the influence of endogenous behavioral adjustments under the following stationarity assumption.

[22] Note that $D_t$ is undefined in both periods for NYC families since they have no younger child.

**Assumption 2** *(Stationarity) For YC families, the propensity score for a younger sibling's treatment, conditional on the older sibling's proxy treatment assignment, remains constant over time:*

$$P(D_t = 1|Z_t = z) = \pi_z \quad \text{for} \quad z, t \in \{0, 1\}.$$

For $Z_t$ to serve as an effective proxy for $D_t$, the propensity score for a younger sibling's treatment must differ between subgroups of YC families classified by the older sibling's virtual binding status — a requirement we formalize in Assumption 3 below.

**Assumption 3** *(Relevance) $Z_t$ is an effective proxy for $D_t$, that is, $\pi_1 > \pi_0$.*

Moreover, identification of the DDD estimator requires that, in the absence of the reform, the difference in outcome dynamics between virtually constrained ($Z = 1$) and unconstrained ($Z = 0$) families is identical across YC and NYC families. Assumption 4 below formally states this identification assumption in potential outcome notation:

**Assumption 4** *(Parallel Trends in Relative Changes)*

$$\begin{aligned}
&\big(E[Y_1(0)|Z_1 = 1, G_1 = 1] - E[Y_0|Z_0 = 1, G_0 = 1]\big) - \big(E[Y_1(0)|Z_1 = 0, G_1 = 1] - E[Y_0|Z_0 = 0, G_0 = 1]\big) \\
&\quad = \big(E[Y_1|Z_1 = 1, G_1 = 0] - E[Y_0|Z_0 = 1, G_0 = 0]\big) - \big(E[Y_1|Z_1 = 0, G_1 = 0] - E[Y_0|Z_0 = 0, G_0 = 0]\big),
\end{aligned} \tag{5}$$

*where $Y_1(0)$ denotes the potential outcome for YC families in the post-reform period if the hours constraint were not binding for the younger child (i.e., $D_1 = 0$).*

For YC families, let $Y_1(1)$ denote the potential outcome in the post-reform period if the hours constraint were binding (i.e., $D_1 = 1$). The observed post-reform outcome $Y_1$ is determined by the latent treatment status $D_1$ and the corresponding potential outcomes as follows:

$$Y_1 = D_1 Y_1(1) + (1 - D_1)Y_1(0).$$

For YC families after the reform, we define the conditional average treatment effect on the treated (CATT), given the older sibling's virtual binding status:

$$\Delta_z = E[Y_1(1) - Y_1(0)|D_1 = 1, Z_1 = z, G_1 = 1] \quad \text{for} \quad z \in \{0, 1\}.$$

Under Assumptions 1-4, the DDD estimator in Equation (3) can be expressed as a weighted difference between $\Delta_1$ and $\Delta_0$:

$$E[\widehat{\rho}_{DDD}] = \pi_1 \Delta_1 - \pi_0 \Delta_0.$$

Following De Chaisemartin and d'Haultfoeuille (2018), we develop a Wald-DDD estimator by dividing the DDD estimate for the outcome by the treatment propensity score difference between virtually constrained and non-constrained YC families:

$$\widehat{W}_{DDD} = \frac{\widehat{\rho}_{DDD}}{\widehat{\pi}_1 - \widehat{\pi}_0}. \tag{6}$$

**Theorem 1** *If Assumptions 1-4 are satisfied, then*

$$E[\widehat{W}_{DDD}] = \Delta_1 + \alpha(\Delta_1 - \Delta_0),$$

*where* $\alpha = \frac{\pi_0}{\pi_1 - \pi_0}$.

If we further assume that the CATTs are identical across subgroups of treated YC families regardless of their proxy treatment assignment (i.e., $\Delta_1 = \Delta_0$), then the Wald-DDD estimator identifies the average treatment effect on the treated (ATT) among all YC families.

**Assumption 5** *(Homogeneity in the CATT between Subgroups)*

$$\Delta_1 = \Delta_0 = \Delta,$$

*where* $\Delta = E[Y_1(1) - Y_1(0)|D_1 = 1, G_1 = 1]$, *the ATT among all YC families.*

**Theorem 2** *If Assumptions 1-5 are satisfied, then*

$$E[\widehat{W}_{DDD}] = \Delta.$$

Online Appendix F explores a partial identification strategy that constructs informative bounds on the CATT parameter $\Delta_1$ using the Wald-DDD estimator $\widehat{W}_{DDD}$ without imposing Assumption 5. However, since $\pi_1 \gg \pi_0$ in our context,[23] the bias term in Theorem

[23]Specifically, $\hat{\pi}_1 = 0.31$ and $\hat{\pi}_0 = 0.04$ (row 1, Table 1), which implies $\hat{\alpha} = \frac{\hat{\pi}_0}{\hat{\pi}_1 - \hat{\pi}_0} = \frac{0.04}{0.31 - 0.04} \approx 0.15$.

1 remains quite small even without imposing Assumption 5. Moreover, using our random forest algorithm from Section 5.2 to select a control group with $\pi_0 < 0.01$ yields very similar results.

## 6.2 GMM estimator

In this subsection, we extend the Wald-DDD estimator in Equation (6) to a GMM estimator within a linear regression framework. This extension allows us to leverage multiple periods of repeated cross-sectional data, incorporate covariates, and facilitate formal statistical inference. Consider a stacked data set comprising two groups observed over $T$ periods, where each observation is characterized by the variables $(G_i, T_i, Z_i, Y_i, D_i)$ for $i = 1, \ldots, N$. In line with Assumption 1, the actual treatment status $D_i$ is observed only for observations in the YC group ($G_i = 1$) during pre-reform periods ($T_i < 2012$).

We begin with a linear treatment probability model for all YC families before the reform (i.e., $G_i = 1$ and $T_i < 2012$):

$$D_i = \alpha + \beta Z_i + \mu_i. \tag{7}$$

For outcome analysis, we specify a generalized linear regression model:

$$\begin{aligned} Y_i = \sum_{t=2009}^{2014} \delta_t \times \mathbf{1}(T_i = t) + \sum_{t=2009}^{2014} \kappa_t \big(G_i \times \mathbf{1}(T_i = t)\big) + \sum_{t=2009}^{2014} \theta_t \big(Z_i \times \mathbf{1}(T_i = t)\big) + \tau(Z_i \times G_i) \\ + \lambda\big(G_i \times D_i \times \mathbf{1}(T_i \geq 2012)\big) + \varepsilon_i, \end{aligned} \tag{8}$$

where all one-way and two-way fixed effects between calendar year, group, and proxy dummies are accounted for in the same way as the generalized DDD specification in Equation (4), and the coefficient $\lambda$ on the three-way interaction term $G_i \times D_i \times \mathbf{1}(T_i \geq 2012)$ represents the ATT parameter of interest.

Let $\Lambda \equiv \big(\lambda, \alpha, \beta, \tau, \{\delta_t, \kappa_t, \theta_t\}_{t \in \{2009, \ldots, 2014\}}\big)$ denote the vector of population parameters. Online Appendix G derives the moment conditions for the joint estimation of Equations (7) and (8) under Assumption 2 (i.e., stationarity in the first-stage relationship), and also discusses incorporation of covariates in this GMM framework.

### 6.3 GMM Results

Table 3 reports the GMM estimates for the extensive margin of maternal employment. Under Assumption 5, this estimate corresponds to the average effect of the 2012 reform for families with a younger child that are affected by the policy change. The specification strategy parallels the ITT analysis in Section 5. Column 1 includes two-way fixed effects, while Column 2 augments the baseline specification with three-way interaction terms between covariates and the two-way fixed effects, chosen by the post-double selection procedure.

The estimated effects are similar across specifications. The first-stage estimates in the top panel indicate a baseline exposure probability of approximately 4 percentage points among families whose older child was not constrained by the reform. The exposure probability increases by more than 26 percentage points among families whose older child was constrained, indicating that the proxy provides substantial predictive power for treatment status. The bottom panel reports the corresponding treatment effect estimates: the reform reduces maternal employment by approximately 11–12 percentage points among affected families. The magnitude is stable across specifications and is consistent with scaling the ITT estimates in Table 2 (approximately 3 percentage points) by the inverse of the first-stage coefficient ($\frac{1}{0.26}$).

Figure 10 reports the intensive-margin estimates. The reform raises the likelihood of very low hours of work, with positive effects up to roughly 10 hours per week. Beyond that point, the estimates turn negative after 20 weekly hours, consistent with a reallocation toward higher hours among employed mothers. Maternal earnings display a comparable pattern. Although the intensive-margin and earnings estimates are somewhat imprecise, the point estimates at 26 and 32 hours per week are negative and statistically significant, indicating increased probabilities of working beyond these thresholds.

## 7 Discussion

This paper studies the labor supply effects of tightening work requirements embedded in a subsidized child care system. Unlike reforms that introduce or eliminate work requirements at the extensive margin, the 2012 Dutch reform increased the intensity of work required to maintain access to subsidized child care. We show that this tightening led to a substantial 12 percentage point decline in maternal employment rates among families close to the new eligibility threshold, alongside increases in work hours among a subset of mothers who remained employed. These findings illustrate how intensive-margin work requirements can generate sharply non-monotonic labor supply responses.

This pattern reflects a reallocation of labor supply across the distribution of work hours. By tying child care subsidy eligibility to parental work intensity, the reform raised the minimum scale of market work required to maintain prior child care arrangements. For some mothers, this created incentives to increase work hours in order to satisfy the new eligibility threshold. For others, the increase in required work intensity made labor force participation less attractive, leading to exits from employment. This pattern is consistent with the reform introducing a nonconvexity in the maternal labor supply budget set by creating a discrete threshold for maintaining subsidized child care.

Our estimates pertain to families whose pre-reform child care usage placed them near the eligibility margin created by the hours constraint. Using a Wald-DDD framework, we identify the effect of the reform for mothers whose access to subsidized child care was sensitive to changes in work hours. As such, the results should not be interpreted as the average effect of child care subsidies or work requirements more broadly, but rather as the impact of tightening work requirements at the intensive margin for families close to the constraint. Considering there were around 500,000 families with a child in the age category 1–3 in this period, and that around 6% of them were bound by the 2012 hours constraint, an effect size of 12 percentage points implies that at least 3,600 maternal jobs were lost in the aftermath of the 2012 reform. We also find that these employment effects persist beyond early childhood, suggesting that temporary exits induced by tighter work requirements may have longer-lasting consequences due to re-entry frictions, even after direct child care needs diminish.

The findings help reconcile mixed evidence in the existing literature on work requirements in social welfare programs. Much of the prior evidence, particularly from the U.S., focuses on requirements that operate at the extensive margin and primarily screen non-participation. In contrast, the reform studied here conditions access to a key complementary input to work — subsidized child care — on meeting a minimum intensity threshold. When work requirements interact with fixed costs of employment, such as child care and within-household coordination, tightening these requirements can induce employment exits even as some workers increase hours.

From a policy perspective, the labor supply response to the reform is concentrated among mothers, with little evidence of adjustment among fathers. This asymmetry is consistent with gender differences in labor supply elasticities and patterns of household specialization in the presence of young children. The extensive margin labor supply response is further concentrated among mothers who worked relatively few hours before the reform, with below-median hourly wages, and above-median partner income. This pattern, combined with the persistence of the labor supply effects, is consistent with the reform deepening gender inequality.

Finally, the results speak to an active policy debate in the Netherlands regarding the role of child care subsidies. If subsidized child care is primarily viewed as a labor market instrument, conditioning access on work intensity may be justified on incentive grounds. However, if the objective is to ensure access to high-quality child care — particularly for children from disadvantaged backgrounds — then intensive work requirements may undermine that goal. Our findings show that tighter requirements reduce child care use among mothers with lower work intensity and lower hourly wages. Since early childhood education yields especially high returns for disadvantaged children (Elango et al., 2015; Cornelissen et al., 2018; Gruber et al., 2025), reductions in access among these families may conflict with broader distributional and developmental objectives.

# Acknowledgements

Van Kippersluis and Zhang graciously acknowledge funding from the Hong Kong Research Grant Council (RGC) General Research Fund (GRF) Projects 12500820 and 14500815. The authors acknowledge access to several linked data resources by Statistics Netherlands (CBS). We thank Ruby Doeleman, Andy Moose Lee, and Cheuk Fai Ng for research assistance, and Gordon Dahl, Peter Eibich, Mery Ferrando, Chao Fu, Shoya Ishimaru, Egbert Jongen, Lance Lochner, Sara Rellstab, Christopher Taber, Junjian Yi, Jia Wu and seminar participants at Chinese University of Hong Kong, Peking University, Sun Yat-sen University, University of Maastricht, the Netherlands Bureau for Economic Policy Analysis, the Essen Health conference, Erasmus University, University of Duisburg-Essen, Chinese Economist Society Annual Meeting, City University of Hong Kong Microeconomics Conference, the Dutch Economists Day, the Dutch Health Econometrics Workshop, and the Econometric Society European Meeting for useful comments.

# References


**Albanesi, Stefania, Claudia Olivetti, and Barbara Petrongolo**, "Families, labor markets, and policy," in "Handbook of the Economics of the Family," Vol. 1, Elsevier, 2023, pp. 255–326.

**Allegretto, Sylvia A, Arindrajit Dube, and Michael Reich**, "Do Minimum Wages Really Reduce Teen Employment? Accounting for Heterogeneity and Selectivity in State Panel Data," *Industrial Relations*, 2011, *50* (2).

**Andresen, Martin Eckhoff and Emily Nix**, "You Cannot Force Me Into Caregiving: Paternity Leave and the Child Penalty," *The Economic Journal*, 07 2025, *136* (674), 780–797.

_ **and Tarjei Havnes**, "Child care, parental labor supply and tax revenue," *Labour Economics*, 2019, *61*, 101762.

**Andresen, Martin Eckhoo and Emily Nix**, "What Causes the Child Penalty? Evidence from Adopting and Same Sex Couples," *Journal of Labor Economics*, 2022, *40* (4), 971–1004.

**Angelov, Nikolay, Per Johansson, and Erica Lindahl**, "Parenthood and the Gender Gap in Pay," *Journal of Labor Economics*, 2016, *34* (3), 545–579.

**Attanasio, Orazio, Hamish Low, and Virginia Sánchez-Marcos**, "Explaining changes in female labor supply in a life-cycle model," *American Economic Review*, 2008, *98* (4), 1517–1552.

_ **, Ricardo Paes de Barros, Pedro Carneiro, David K Evans, Lycia Lima, Pedro Olinto, and Norbert Schady**, "Public childcare, labor market outcomes of caregivers, and child development: Experimental evidence from Brazil," Technical Report, National Bureau of Economic Research 2022.

**Averett, Susan L and Yang Wang**, "Assessing the Fertility Effects of Childcare Cost Subsidies: Evidence from the Child and Dependent Care Tax Credit," 2023.

**Baker, Michael, Jonathan Gruber, and Kevin Milligan**, "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *Journal of Political Economy*, 2008, *116* (4), 709–745.

**Bansraj, Dyaran and Dong Xu**, "How Does Private Equity Shape Childcare? Implications for Competition, Pricing, and Quality," Technical Report, SSRN report number: 5045089 2024.

**Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen**, "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 2014, *81* (2), 608–650.

**Berden, C. and L. Kok**, "Participatie-effect kinderopvangtoeslag," *SEO rapport 2009-073*, 2009.

**Besley, Timothy and Stephen Coate**, "Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs," *The American Economic Review*, 1992, *82* (1), 249–261.

**Bettendorf, Leon J H, Egbert L W Jongen, and Paul Muller**, "Childcare subsidies and labour supply — Evidence from a large Dutch reform," *Labour Economics*, 2015, *36*, 112–123.

**Biewen, Martin, Bernd Fitzenberger, and Marian Rümmele**, "Using Distribution Regression Difference-in-Differences to Evaluate the Effects of a Minimum Wage Introduction on the Distribution of Hourly Wages and Hours Worked," 2022.

**Black, Sandra E, Paul J Devereux, Katrine V Løken, and Kjell G Salvanes**, "Care or cash? The effect of child care subsidies on student performance," *Review of Economics and Statistics*, 2014, *96* (5), 824–837.

**Blank, Rebecca M**, "Evaluating welfare reform in the United States," *Journal of Economic Literature*, 2002, *40* (4), 1105–1166.

**Blau, Francine D and Lawrence M Kahn**, "Female labor supply: Why is the United States falling behind?," *American Economic Review*, 2013, *103* (3), 251–256.

**Blundell, Richard, Alan Duncan, and Costas Meghir**, "Estimating labor supply responses using tax reforms," *Econometrica*, 1998, pp. 827–861.

**_ , Monica Costa Dias, Costas Meghir, and Jonathan Shaw**, "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 2016, *84* (5), 1705–1753.

**Boer, Henk Wim De, Egbert Jongen, and Patrick Koot**, "Optimal Taxation of Secondary Earners in the Netherlands: Has Equity Lost Ground?," CPB Discussion Paper, Centraal Planbureau, The Hague 2018.

**Botosaru, Irene and Federico H. Gutierrez**, "Difference-in-differences when the treatment status is observed in only one period," *Journal of Applied Econometrics*, 2018, *33* (1), 73=90.

**Browne, James**, "The impact of tax and benefit reforms by sex: some simple analysis," *IFS Briefing Note 118*, 2011.

**Card, David and Dean R. Hyslop**, "The dynamic effects of an earnings subsidy for long-term welfare recipients: Evidence from the self sufficiency project applicant experiment," *Journal of Econometrics*, 2009, *153* (1), 1–20.

**_ , Fabrizio Colella, and Rafael Lalive Lalive**, "Gender Preferences in Job Vacancies and Workplace Gender Diversity," *Review of Economic Studies*, 2025, *92* (4).

**Cascio, Elizabeth U**, "Maternal labor supply and the introduction of kindergartens into American public schools," *Journal of Human Resources*, 2009, *44* (1), 140–170.

**CBS**, "Sociaaleconomische trends," Technical Report, Centraal Bureau voor de Statistiek, Den Haag/Heerlen 2011.

**CBS**, "Nabijheid voorzieningen," 2016. Statline.

**Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and David Zentler-Munro**, "Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes," *Journal of Labor Economics*, 2022, *40* (S1), S203–S247.

**Chaisemartin, Clément De and Xavier d'Haultfoeuille**, "Fuzzy differences-in-differences," *The Review of Economic Studies*, 2018, *85* (2), 999–1028.

**Chan, Marc K**, "A dynamic model of welfare reform," *Econometrica*, 2013, *81* (3), 941–1001.

_ **and Robert Moffitt**, "Welfare reform and the labor market," *Annual Review of Economics*, 2018, *10* (1), 347–381.

**Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly**, "Inference on counterfactual distributions," *Econometrica*, 2013, *81* (6), 2205–2268.

**Chetty, Raj, John N. Friedman, and Emmanuel Saez**, "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, December 2013, *103* (7), 2683–2721.

**Collischon, Matthias, Daniel Kuehnle, and Michael Oberfichtner**, "Who benefits from cash-for-care? The effects of a home care subsidy on maternal employment, childcare choices, and children's development," *Journal of Human Resources*, 2022, pp. 0720–11051R1.

**Cook, Jason B and Chloe N East**, "The Disenrollment and Labor Supply Effects of SNAP General Work Requirements," Technical Report, National Bureau of Economic Research 2024.

**Corinth, Kevin, Bruce D Meyer, Matthew Stadnicki, and Derek Wu**, "The anti-poverty, targeting, and labor supply effects of replacing a Child Tax Credit with a child allowance," Technical Report, National Bureau of Economic Research 2021.

**Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg**, "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance," *Journal of Political Economy*, 2018, *126* (6), 2356–2409.

**Cortés, Patricia and Jessica Pan**, "Outsourcing Household Production: Foreign Domestic Workers and Native Labor Supply in Hong Kong," *Journal of Labor Economics*, 2013, *31* (2), 328–371.

_ **and** _ , "When Time Binds: Substitutes for Household Production, Returns to Working Long Hours, and the Skilled Gender Wage Gap," *Journal of Labor Economics*, 2019, *39* (2), 351–398.

**CPB**, "Een analyse van de groei van de formele kinderopvang in het recente verleden en in de nabije toekomst," *CPB (Centraal Planbureau) Notitie, 14 april*, 2008.

**Dahl, Gordon B and Lance Lochner**, "The impact of family income on child achievement: Evidence from the earned income tax credit," *The American Economic Review*, 2012, *102* (5), 1927–1956.

**de Boer, Henk Wim, Egbert LW Jongen, and Jan Kabatek**, "The effectiveness of fiscal stimuli for working parents," *Labour Economics*, 2022, *76*, 102152.

**de la Croix, J., A. Houtkoop, E.C. Junger van Hoorn, A.J. Machiels van Es, W. Smits, and G. Wiggers**, "Arbeid en (Mantel)zorg," Technical Report, Ministerie van Sociale Zaken en Werkgelegenheid, Den Haag 2014.

**den Dulk, Laura**, "The Netherlands - country brief," Technical Report July, 13th International Review of Leave Policies and Related Research 2017.

**Dube, Arindrajit**, "Minimum wages and the distribution of family incomes," *American Economic Journal: Applied Economics*, 2019, *11* (4), 268–304.

**Einav, Liran, Amy Finkelstein, Sendhil Mullainathan, and Ziad Obermeyer**, "Employed in a SNAP? The impact of work requirements on program participation and labor supply," *Science*, 2018, *360*, 1462–1465.

**Elango, Sneha, Jorge Luis García, James J Heckman, and Andrés Hojman**, "Early childhood education," in "Economics of Means-Tested Transfer Programs in the United States, Volume 2," University of Chicago Press, 2015, pp. 235–297.

**Enriquez, Brandon, Damon Jones, and Ernie Tedeschi**, "The Short-Term Labor Supply Response to the Expanded Child Tax Credit," *AEA Papers and Proceedings*, May 2023, *113*, 401–05.

**Fang, Hanming and Michael P Keane**, "Assessing the impact of welfare reform on single mothers," *Brookings Papers on Economic Activity*, 2004, *2004* (1), 1–116.

**Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux**, "Unconditional quantile regressions," *Econometrica*, 2009, *77* (3), 953–973.

**Fortin, Nicole M**, "Gender role attitudes and the labour-market outcomes of women across OECD countries," *Oxford Review of Economic Policy*, 2005, *21* (3), 416–438.

**Gathmann, Christina and Björn Sass**, "Taxing Childcare: Effects on Childcare Choices, Family Labor Supply, and Children," *Journal of Labor Economics*, 2018, *36* (3), 665–709.

**Giuliani, Giuliana and Ann Zofie Duvander**, "Cash-for-care policy in Sweden: An appraisal of its consequences on female employment," *International Journal of Social Welfare*, 2017, *26* (1), 49–62.

**Givord, Pauline and Claire Marbot**, "Does the cost of child care affect female labor market participation? An evaluation of a French reform of childcare subsidies," *Labour Economics*, 2015, *36*, 99–111.

**Goldin, Claudia and Joshua Mitchell**, "The new life cycle of women's employment: Disappearing humps, sagging middles, expanding tops," *Journal of Economic Perspectives*, 2017, *31* (1), 161–182.

**Goldin, Jacob, Tatiana Homonoff, Neel A Lal, Ithai Lurie, and Katherine Michelmore**, "Child Tax Benefits and Labor Supply: Evidence from California," Technical Report, National Bureau of Economic Research 2024.

**Gray, Colin, Adam Leive, Elena Prager, Kelsey Pukelis, and Mary Zaki**, "Employed in a SNAP? The impact of work requirements on program participation and labor supply," *American Economic Journal: Economic Policy*, 2023, *15* (1), 306–341.

**Grogger, Jeff and Lynn A Karoly**, *Welfare reform: Effects of a decade of change*, Harvard University Press, 2005.

**Gruber, Jon and Emmanuel Saez**, "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 2002, *84* (1), 1–32.

**Gruber, Jonathan**, "The incidence of mandated maternity benefits," *The American Economic Review*, 1994, pp. 622–641.

**_ , Tuomas Kosonen, and Kristiina Huttunen**, "Paying Moms to Stay Home: Short and Long Run Effects on Parents and Children," *Journal of Public Economics*, 2025, *251*, 105496.

**Guiliano, Laura**, "Minimum Wage Effects on Employment, Substitution, and the Teenage Labor Supply: Evidence from Personnel Data," *Journal of Labor Economics*, 2013, *31* (1).

**Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura**, "Child-related transfers, household labour supply, and welfare," *The Review of Economic Studies*, 2020, *87* (5), 2290–2321.

**Hahn, Heather and Ron Haskins**, "Federal Work Requirements Debate," Technical Report, American Enterprise Institute Poverty and Social Policy Debate Series, June 18, 2018. 2018.

**Han, Jeehoon**, "The impact of SNAP work requirements on labor supply," *Labour Economics*, 2022, *74*, 102089.

**Harris, J. M. and E. Patacchini**, "Closing the Gender Gap: Promoting Labour Market Participation," in "Handbook on Labour Markets in Transition," Edward Elgar Publishing Limited, 2024, pp. 463–482.

**Harris, Timothy F**, "Do SNAP work requirements work?," *Economic Inquiry*, 2021, *59* (1), 72–94.

**Havnes, Tarjei and Magne Mogstad**, "Money for nothing? Universal child care and maternal employment," *Journal of Public Economics*, 2011, *95* (11), 1455–1465.

_ **and** _ , "Is universal child care leveling the playing field?," *Journal of Public Economics*, 2015, *127*, 100–114.

**Hermes, Henning, Marina Krauß, Philipp Lergetporer, Frauke Peter, and Simon Wiederhold**, "Early child care, maternal labor supply, and gender equality: A randomized controlled trial," *IWH Discussion Papers, No. 14/2024*, 2024.

**Hotz, V Joseph and Matthew Wiswall**, "Child care and child care policy: Existing policies, their effects, and reforms," *The ANNALS of the American Academy of Political and Social Science*, 2019, *686* (1), 310–338.

_ **and Robert A Miller**, "An empirical analysis of life cycle fertility and female labor supply," *Econometrica: Journal of the Econometric Society*, 1988, pp. 91–118.

**Huber, Katrin and Geske Rolvering**, "Public child care and mothers' career trajectories," Technical Report, Center for Economic Policy Analysis 2023.

**Intomart/GfK**, "Motieven voor gastouderopvang," 12 2011.

**Jiménez, Bruno and Silvio Rendon**, "Labor Market Effects of Bounds on Domestic Outsourcing," *Journal of Development Economics*, 2025, *173.*

**Joyce, Robert**, "Tax and benefit reforms due in 2012-13, and the outlook for household incomes," *IFS Briefing Note BN126*, 2012.

**Killingsworth, Mark R and James J Heckman**, "Female labor supply: A survey," *Handbook of Labor Economics*, 1986, *1*, 103–204.

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

**_ , _ , Johanna Posch, Andreas Steinhauer, and Josef Zweimüller**, "Child penalties across countries: Evidence and explanations," in "AEA Papers and Proceedings," Vol. 109 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2019, pp. 122–126.

**_ , _ , _ , _ , and _** , "Do family policies reduce gender inequality? Evidence from 60 years of policy experimentation," *American Economic Journal: Economic Policy*, 2024, *16* (2), 110–149.

**Korvorst, Marjolein**, "De werk-privébalans van vaders met jonge kinderen," Technical Report, Centraal Bureau voor de Statistiek, Den Haag 2019.

**Kosonen, Tuomas**, "To Work or Not to Work? The Effect of Childcare Subsidies on the Labour Supply of Parents," *The BE Journal of Economic Analysis & Policy*, 2014, *14* (3), 817–848.

**Kuka, Elira and Na'amma Shenhav**, "'Long-Run Effects of Incentivizing Work after Childbirth," *American Economic Review*, 2024, *114* (6), 1692–1722.

**Lefebvre, Pierre and Philip Merrigan**, "Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada," *Journal of Labor Economics*, 2008, *26* (3), 519–548.

**Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen**, "Can Women Have Children and a Career?," *American Economic Review*, 2017, *107* (6), 1611–1637.

**Lundin, Daniela, Eva Mörk, and Björn Öckert**, "How far can reduced childcare prices push female labour supply?," *Labour Economics*, 2008, *15* (4), 647–659.

**Mari, Gabriele and Olivier Marie**, "Poor Rich Women: Losing Childcare Subsidies and Female Labour Market Outcomes," Technical Report, Mimeo 2026.

**Merens, Ans and Freek Bucx**, "Werken aan de start - Jonge vrouwen en mannen op de arbeidsmarkt," Technical Report, Sociaal en Cultureel Planbureau SCP, Den Haag 2018.

**Moffitt, Robert**, "Incentive effects of the US welfare system: A review," *Journal of Economic Literature*, 1992, *30* (1), 1–61.

**Moffitt, Robert A**, "Welfare programs and labor supply," *Handbook of Public Economics*, 2002, *4*, 2393–2430.

**Naz, Ghazala**, "The impact of cash-benefit reform on parents' labour force participation," *Journal of Population Economics*, 2004, *17*, 369–383.

**Neumark, David, J.M. Ian Salas, and William Wascher**, "Revisiting the Minimum Wage-Employment Debate: Throwing Out the Baby with the Bathwater?," *Industrial and Labor Relations Review*, 2014, *67* (suppl.).

**Nollenberger, Natalia and Núria Rodríguez-Planas**, "Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from Spain," *Labour Economics*, 2015, *36*, 124–136.

**OECD**, "Part-time employment rate (indicator).," Technical Report 2019.

**Olden, Andreas and Jarle Møen**, "The triple difference estimator," *The Econometrics Journal*, 2022, *25* (3), 531–553.

**Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz**, "The short-and long-term career effects of graduating in a recession," *American Economic Journal: Applied Economics*, 2012, *4* (1), 1–29.

**Ortiz-Villavicencio, Marcelo and Pedro HC Sant'Anna**, "Better Understanding Triple Differences Estimators," *arXiv preprint arXiv:2505.09942*, 2025.

**Parliament**, "Letter from the Minister of Social Affairs and State Secretary of Finance to the Parliament, 31322, nr. 123," 2011. Parliamentary document.

**Pepin, Gabrielle**, "The Effects of Child Care Subsidies on Paid Child Care Participation and Labor Market Outcomes: Evidence from the Child and Dependent Care Credit," Technical Report, IZA Discussion Paper No. 17238 2024.

**Rabaté, Simon and Sara Rellstab**, "What Determines the Child Penalty in the Netherlands? The Role of Policy and Norms," *De Economist*, 2022, *170* (2), 195–229.

**Schlosser, Analia**, "Public preschool and the labor supply of Arab mothers: Evidence from a natural experiment," *Manuscript, The Hebrew University of Jerusalem*, 2005.

**Schøne, Pål**, "Labour supply effects of a cash-for-care subsidy," *Journal of Population Economics*, 2004, *17*, 703–727.

**Sieppi, Antti and Jaakko Pehkonen**, "Parenthood and gender inequality: Population-based evidence on the child penalty in Finland," *Economics Letters*, 2019, *182*, 5–9.

**Sommers, Benjamin D, Anna L Goldman, Robert J Blendon, E John Orav, and Arnold M Epstein**, "Medicaid work requirements—results from the first year in Arkansas," *New England Journal of Medicine*, 2019, *381* (11), 1073–1082.

**Sommers, Benjamin D., Lucy Chen, Robert J. Blendon, E. John Orav, and Arnold M. Epstein**, "Medicaid Work Requirements In Arkansas: Two-Year Impacts On Coverage, Employment, And Affordability Of Care," *Health Affairs*, 2020, *39* (9), 1522–1530. PMID: 32897784.

**Staatsblad**, "Besluit van 16 september 2011, Staatsblad van het Koninkrijk der Nederlanden, Jaargang 2011, Nr. 424," 2011. Retrieved online.

**SZW**, "Evaluatie van de Wet Kinderopvang," *Ministry of Social Affairs report 25 May 2012*, 2012.

—, "Beleidsdoorlichting Kinderopvang," *Ministry of Social Affairs report 16 November 2015*, 2015.

**Van der Wiel, Karen and Maarten Van 't Riet**, "Kinderopvang in kaart," *CPB Notitie, 27 oktober*, 2011.

**Von Wachter, Till**, "The persistent effects of initial labor market conditions for young adults and their sources," *Journal of Economic Perspectives*, 2020, *34* (4), 168–194.

**Ziliak, James P**, "Temporary assistance for needy families," in "Economics of Means-Tested Transfer Programs in the United States, Volume 1," University of Chicago Press, 2015, pp. 303–393.

**Zimmert, Franziska and Michael Zimmert**, "Part-time subsidies and maternal reemployment: Evidence from a difference-in-differences analysis," *Journal of Applied Econometrics*, 2024, *39* (6), 1149–1171.

# Tables

Table 1: Descriptive statistics

| Variable | (1)<br>VC-YC | (2)<br>VNC-YC | (3)<br>$\Delta_{YC}$ | (4)<br>VC-NYC | (5)<br>VNC-NYC | (6)<br>$\Delta_{NYC}$ |
|---|---|---|---|---|---|---|
| Binded younger child | 0.31 | 0.04 | −0.27*** | | | |
| Boy focal child | 0.53 | 0.52 | −0.01* | 0.51 | 0.50 | 0.00 |
| Rank focal child | 1.28 | 1.26 | −0.03*** | 2.01 | 2.05 | 0.04*** |
| | (0.54) | (0.54) | | (0.71) | (0.74) | |
| Born in the Netherlands | 0.82 | 0.93 | 0.11*** | 0.79 | 0.89 | 0.10*** |
| Education of the mother | | | | | | |
| Missing | 0.15 | 0.26 | 0.11*** | 0.21 | 0.32 | 0.11*** |
| Lower | 0.07 | 0.05 | −0.01*** | 0.09 | 0.08 | −0.01*** |
| Lower vocational | 0.28 | 0.27 | −0.01* | 0.32 | 0.28 | −0.05*** |
| Higher vocational | 0.30 | 0.28 | −0.02*** | 0.23 | 0.22 | −0.01*** |
| University | 0.20 | 0.13 | −0.07*** | 0.14 | 0.10 | −0.04*** |
| Education of the father | | | | | | |
| Missing | 0.21 | 0.30 | 0.10*** | 0.27 | 0.36 | 0.09*** |
| Lower | 0.08 | 0.06 | −0.01*** | 0.10 | 0.08 | −0.01*** |
| Lower vocational | 0.24 | 0.26 | 0.02*** | 0.26 | 0.26 | −0.00 |
| Higher vocational | 0.25 | 0.24 | −0.01 | 0.23 | 0.20 | −0.03*** |
| University | 0.23 | 0.13 | −0.09*** | 0.15 | 0.10 | −0.05*** |
| Birthyear mother | 1977 | 1977 | 0.35*** | 1974 | 1974 | 0.19*** |
| | (4) | (4) | | (5) | (4) | |
| Birthyear father | 1974 | 1975 | 0.46*** | 1971 | 1972 | 0.36*** |
| | (5) | (4) | | (5) | (5) | |
| Hours work mother $s_2$ | 953 | 1210 | 256.57*** | 889 | 1118 | 229.02*** |
| | (488) | (387) | | (459) | (418) | |
| Hours work father $s_2$ | 1748 | 1964 | 216.08*** | 1803 | 1953 | 150.20*** |
| | (616) | (299) | | (581) | (346) | |
| Earnings mother $s_2$ | 17,853 | 21,321 | 3,468*** | 16,260 | 19,836 | 3,575*** |
| | (12,069) | (10,060) | | (11,207) | (10,567) | |
| Earnings father $s_2$ | 38,913 | 39,785 | 872*** | 40,376 | 40,553 | 176 |
| | (19,101) | (12,934) | | (19,147) | (14,017) | |
| Tax. inc. mother $s_2$ | 18,186 | 20,224 | 2,038*** | 17,047 | 19,066 | 2,019*** |
| | (10,106) | (8,690) | | (9,824) | (9,281) | |
| Tax. inc. father $s_2$ | 33,592 | 32,277 | -1,314*** | 34,996 | 33,703 | -1,292*** |
| | (14,661) | (11,116) | | (14,704) | (11,895) | |
| Child care use $s_2$ | 1.00 | 0.67 | −0.33*** | 1.00 | 0.59 | −0.41*** |
| Hours of care $s_2$ | 1,494 | 963 | -530*** | 1,474 | 955 | -519*** |
| | (546) | (443) | | (549) | (448) | |
| Mother's hourly wage | | | | | | |
| Q1 | 0.26 | 0.22 | −0.05*** | 0.31 | 0.26 | −0.05*** |
| Q2 | 0.25 | 0.26 | 0.00 | 0.26 | 0.25 | −0.01*** |
| Q3 | 0.22 | 0.28 | 0.06*** | 0.20 | 0.24 | 0.04*** |
| Q4 | 0.26 | 0.25 | −0.02** | 0.23 | 0.25 | 0.02*** |
| Father's income | | | | | | |
| Q1 | 0.29 | 0.28 | −0.01* | 0.25 | 0.24 | −0.01*** |
| Q2 | 0.20 | 0.27 | 0.07*** | 0.20 | 0.25 | 0.05*** |
| Q3 | 0.21 | 0.24 | 0.03*** | 0.23 | 0.25 | 0.03*** |
| Q4 | 0.30 | 0.21 | −0.09*** | 0.32 | 0.25 | −0.07*** |
| Number of observations | 6,533 | 124,178 | 130,711 | 30,719 | 440,605 | 471,324 |
| Number of unique families | 2,414 | 45,239 | 47,653 | 11,612 | 164,421 | 176,033 |

*Notes*: Descriptive statistics (means, standard deviations in parentheses for non-binary variables) for the sample who was virtually constrained (VC) for the older sibling by the 2012 hours requirement and has a younger child (VC-YC; column 1), the sample who was not virtually constrained (NVC) but has a younger child (NVC-YC; column 2); the sample who was virtually constrained for the older sibling but did not have a younger child (VC-NYC; column 4), and the sample who was not virtually constrained and did not have a younger child (NVC-NYC; column 5). The $\Delta$ symbol indicates the difference across virtually constrained and non-constrained families. $s_2$ denotes that the variable is measured when the focal child (i.e., the older sibling) was aged 2. *, **, and *** denote the statistical significance of the corresponding $t$-test of the difference across groups at 10%, 5%, and 1%, respectively.

Table 2: DD and DDD results for maternal employment outcomes

| | **Working** | | | **Weekly working hours** | | | **Yearly earnings** | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | Basic | Saturated | PDS | Basic | Saturated | PDS | Basic | Saturated | PDS |
| *Panel A. YC sample (DD)* | | $N = 130,711$ | | | $N = 125,335$ | | | $N = 125,335$ | |
| Focal child virtually constrained $s_2$ × post-2012 | $-0.038^{***}$ | $-0.038^{***}$ | $-0.040^{***}$ | $-0.494$ | 0.022 | $-0.034$ | $-353.808$ | 115.038 | 26.331 |
| | (0.009) | (0.011) | (0.011) | (0.333) | (0.270) | (0.270) | (509.682) | (304.241) | (303.805) |
| *Panel B. NYC sample (DD)* | | $N = 471,324$ | | | $N = 449,636$ | | | $N = 449,636$ | |
| Focal child virtually constrained $s_2$ × post-2012 | $-0.008^{**}$ | $-0.004$ | $-0.004$ | $-0.127$ | 0.152 | 0.141 | $-507.217^{**}$ | $-155.532$ | $-158.204$ |
| | (0.004) | (0.004) | (0.004) | (0.147) | (0.127) | (0.127) | (226.189) | (150.893) | (150.245) |
| *Panel C. Pooled sample (DDD)* | | $N = 602,035$ | | | $N = 574,971$ | | | $N = 574,971$ | |
| Focal child virtually constrained $s_2$ × YC × post-2012 | $-0.030^{***}$ | $-0.035^{**}$ | $-0.034^{**}$ | $-0.288$ | $-0.095$ | $-0.140$ | 152.163 | 372.648 | 290.132 |
| | (0.010) | (0.013) | (0.013) | (0.368) | (0.347) | (0.347) | (564.855) | (396.334) | (395.416) |
| *Lagged dep. var. mean $s_2$* | | 1 | | | 21.61 | | | 19,938.62 | |

*Notes*: Each cell reports the estimated coefficient on the interaction term indicated by the row heading from a separate regression. Panels A and B estimate difference-in-differences (DD) regressions among families with a younger child (YC sample) and without a younger child (NYC sample), respectively; Panel C estimates triple-difference (DDD) regressions using the pooled sample of YC and NYC families. For the DD regressions (Panels A and B), the columns 'Basic' control for the standalone fixed effects for dummy indicator for the focal child being virtually constrained by the 2012 hours constraint at age 2 (hereafter virtual constraint dummy) and calendar year dummies; the columns 'Saturated' add to the 'Basic' columns all the two-way interaction terms between control variables and the virtual constraint or year dummies; the columns 'PDS' add to the 'Basic' columns only the two-way interaction terms between control variables and the virtual constraint or year dummies selected based on a post-double selection procedure. For the DDD regressions (Panel C), the columns 'Basic' control for all the two-way interactive fixed effects between the virtual constraint dummy, YC dummy, and year dummies; the columns 'Saturated' add to the 'Basic' columns all the three-way interaction terms between control variables and two-way interactive fixed effects between the virtual constraint dummy, YC dummy, and year dummies; the columns 'PDS' add to the 'Basic' columns only the three-way interaction terms between control variables and and two-way interactive fixed effects between the virtual constraint dummy, YC dummy, and year dummies selected based on a post-double selection procedure. Robust standard errors, clustered at the household level, are reported in parentheses. The lagged dependent variable means when the focal child aged 2 are reported at the bottom of each panel. * p<0.1, ** p<0.05, *** p<0.01.

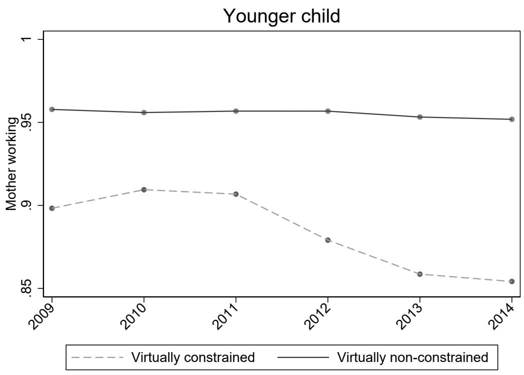
Table 3: ATT results for maternal employment outcomes

| | **Working** | |
|---|---|---|
| | Basic | PDS |
| *Panel A. First stage* | | |
| $\alpha$ | $0.043^{***}$ | $0.042^{***}$ |
| | (0.001) | (0.001) |
| $\beta$ | $0.267^{***}$ | $0.267^{***}$ |
| | (0.011) | (0.011) |
| *Panel B. ATT* | | |
| $\lambda$ | $-0.114^{***}$ | $-0.126^{***}$ |
| | (0.037) | (0.049) |
| $N$ | 602,035 | 602,035 |

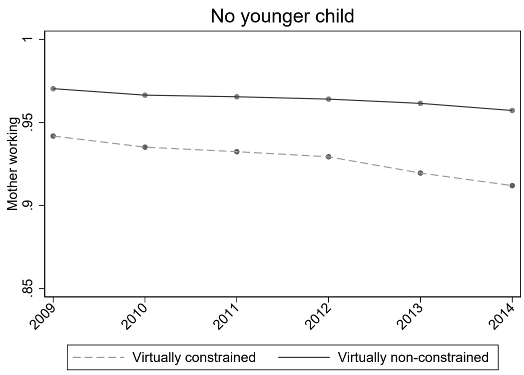
*Notes*: Panels A and B report the GMM estimates for the first-stage relationship (equation (7)) and ATT (equation (8)), respectively. In Panel B, the column 'Basic' controls for all the two-way interactive fixed effects between the virtual constraint dummy, YC dummy, and year dummies; the columns 'PDS' adds to the 'Basic' column the three-way interaction terms between control variables and and two-way interactive fixed effects between the virtual constraint dummy, YC dummy, and year dummies selected based on a post-double selection procedure. Robust standard errors, clustered at the household level, are reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

# Figures

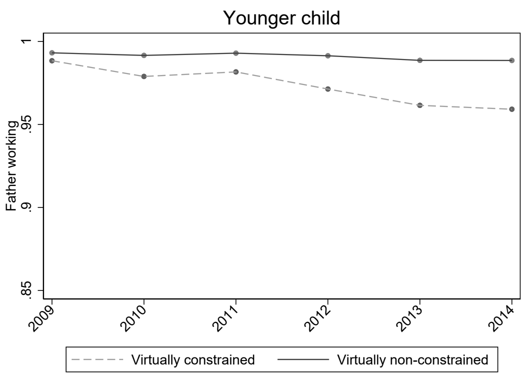
Figure 1: Labor supply for mothers and fathers



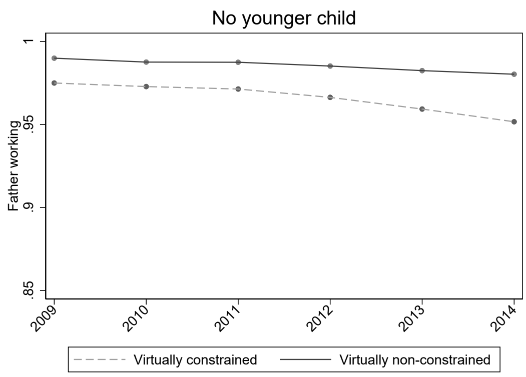

(a)




(b)



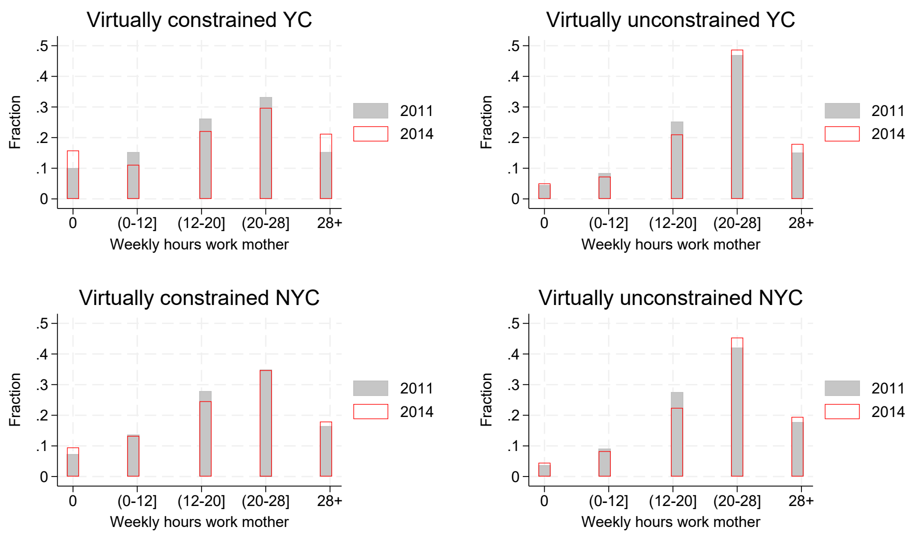

(c)




(d)

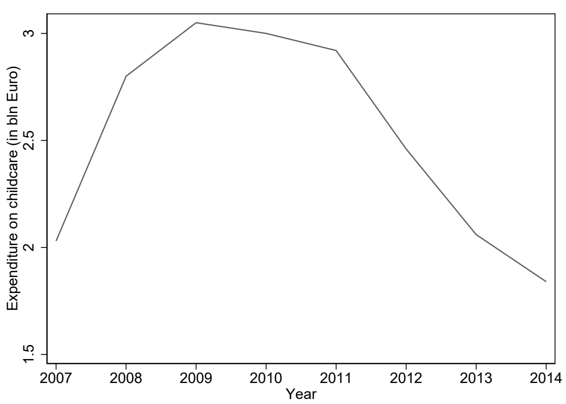
*Notes*: All families have a child born between 2005-2009. Panels (a) and (c) refer to families with a younger child, while panels (b) and (d) refer to families without a younger child. The dashed line corresponds to families who were virtually constrained by the 2012 hours requirement when the focal child was aged 2. The solid line corresponds to families who were not virtually constrained.

Figure 2: Labor supply for mothers at the intensive margin

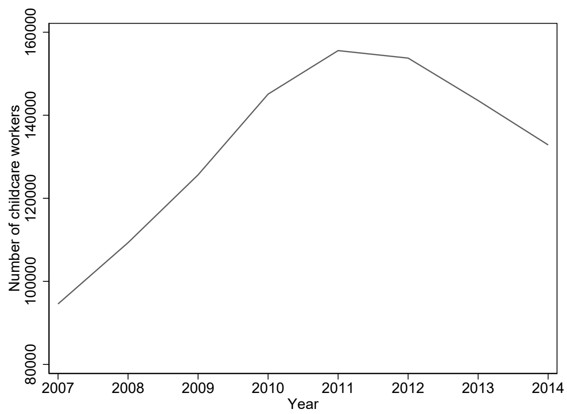



*Notes*: Annual work hours are converted to weekly work hours (dividing by 52) and binned into five categories: 0 for not working at all, $(0-12]$ for working less than or equal to 12 hours per week, $(12-20$ for working more than 12 but less than or equal to 20 hours per week, $(20-28]$ for working more than 20 but less than or equal to 28 hours per week, and 28+ for working more than 28 hours per week. All families have a child born between 2005-2009. The top-left panel refers to our proxy treatment group, families with a younger child (YC) who were virtually constrained by the 2012 hours requirement when the focal child aged 2. The top-right panel refers to families with a younger child (YC) but were not virtually constrained when the focal child aged 2. The bottom-left panel refers to families who did not have a younger child (NYC) but were virtually constrained for the focal child, while the bottom-right panel shows families without a younger child who were not virtually binded for the focal child.

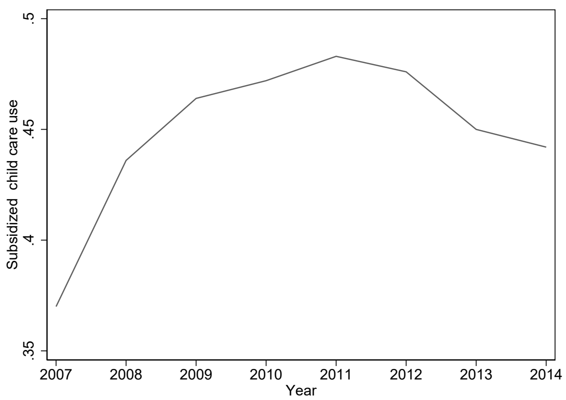
Figure 3: Time series over the period 2007–2014



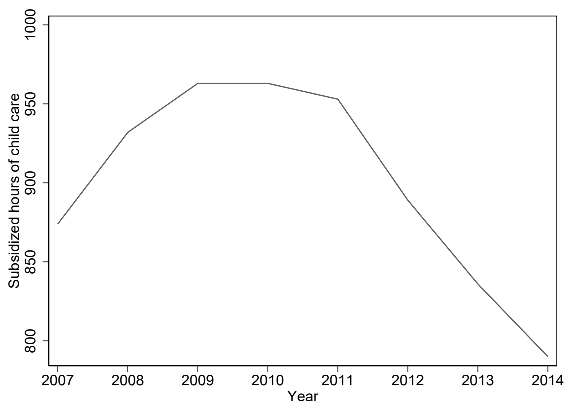

(a)




(b)



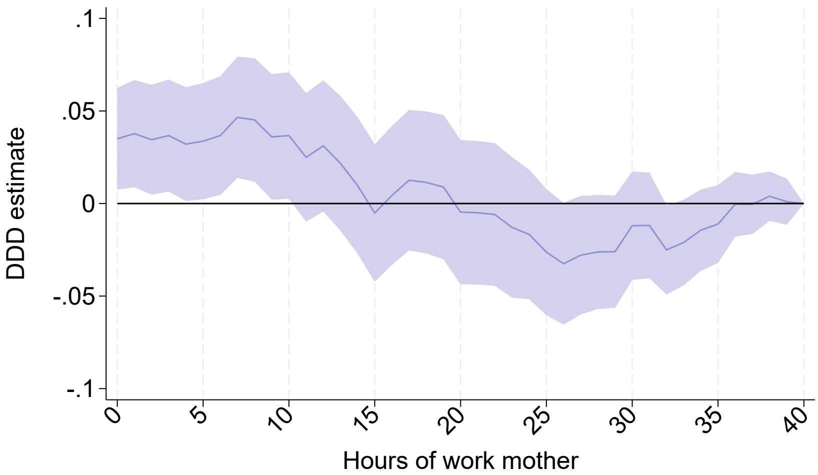

(c)



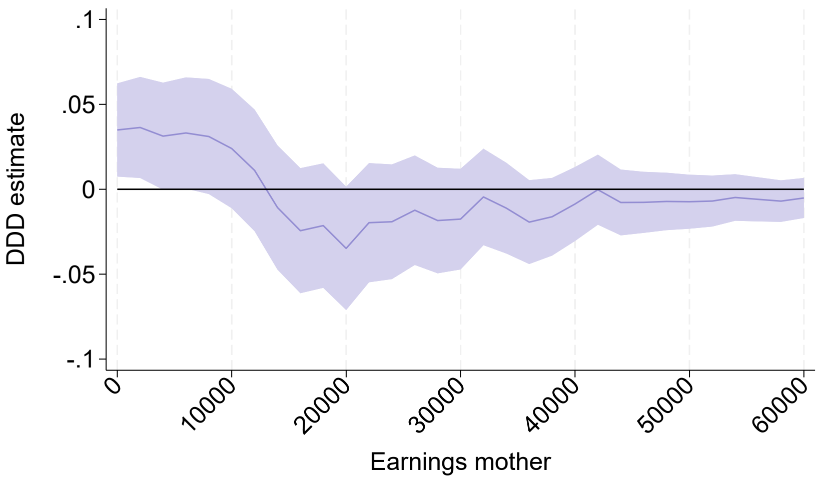

(d)

*Notes*: (a) Government expenditures on child care; (b) Number of child care workers; (c) Subsidized child care use; (d) Hours of child care among mothers with at least one child aged 0–4.

Figure 4: DDD estimates on the intensive margin



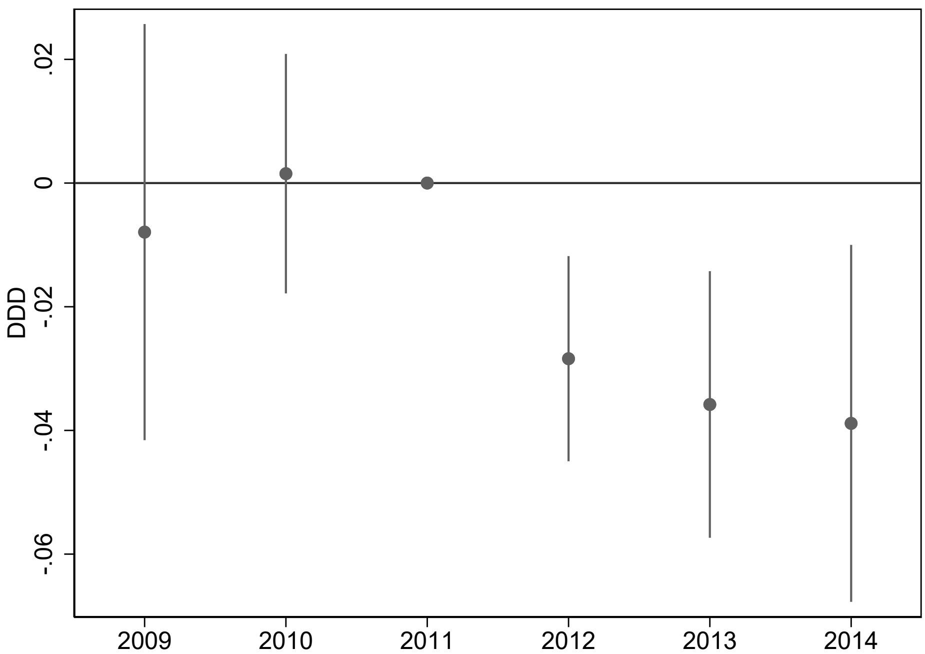

(a) Average weekly hours of work of the mother




(b) Yearly earnings of the mother

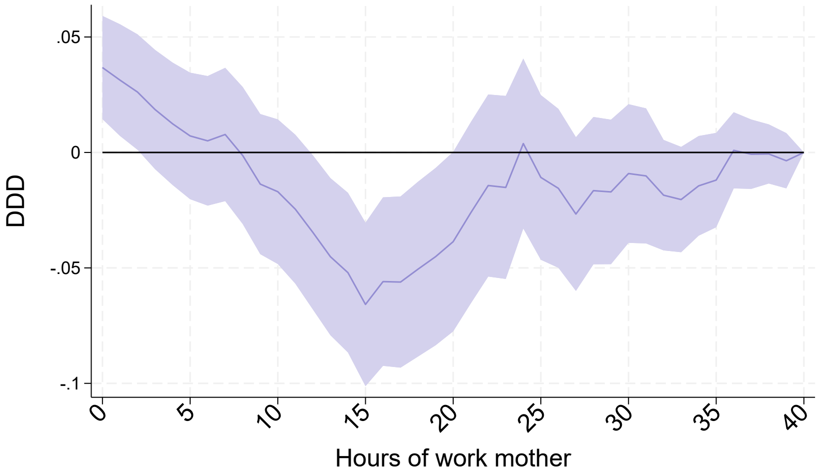
*Notes*: DDD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on maternal hours of work (panel 4a) and maternal earnings (panel 4b). The presented coefficients $\rho$ and associated confidence intervals stem from equation (4) where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for maternal weekly work hours, and bin size €1,000 for maternal earnings.

Figure 5: Event study for maternal labor supply.

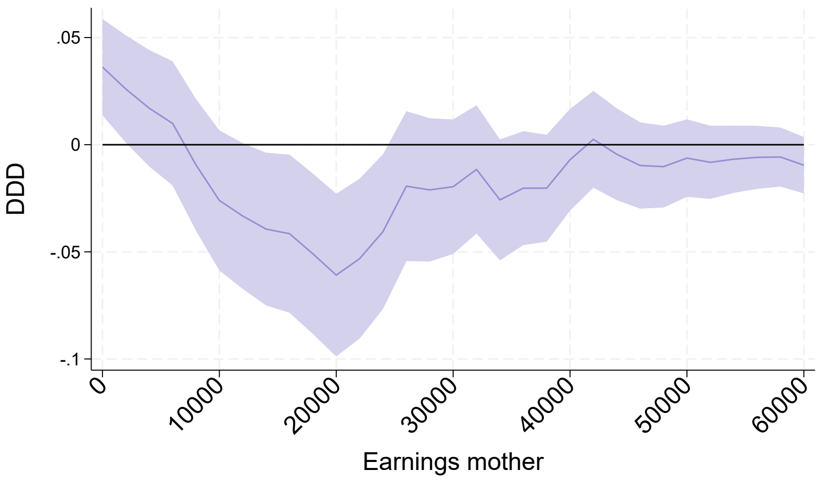



*Notes*: DDD estimates presented in an event study framework, with three-way interactions between virtually constrained, younger-child and the individual year dummies. 2011 (i.e., one year pre-reform) is the baseline year.

Figure 6: DDD estimates on maternal work hours and earnings — ML prediction



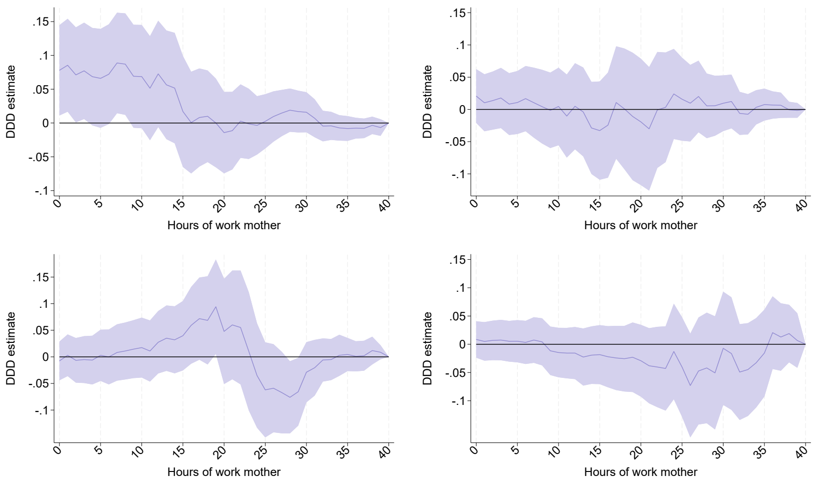

(a) Average weekly hours of work of the mother



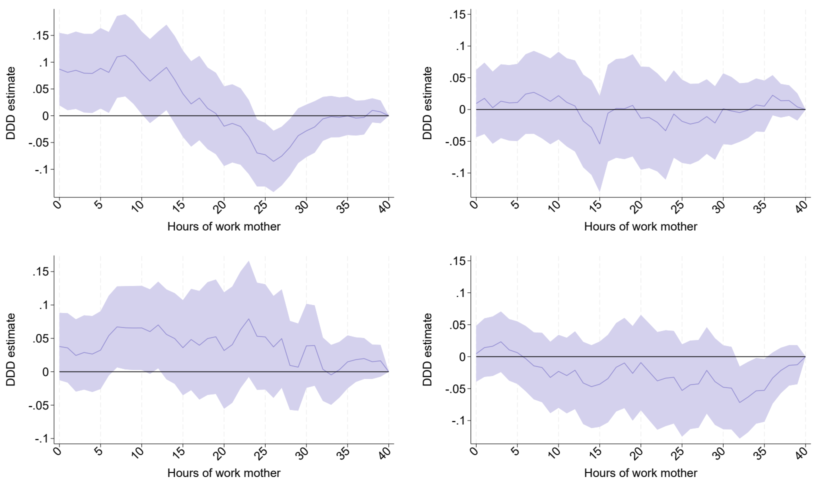

(b) Yearly earnings of the mother

*Notes*: DDD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on maternal hours of work (panel 6a) and maternal earnings (panel 6b). The presented coefficients $\rho$ and associated confidence intervals stem from equation (4) where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for maternal weekly work hours, and bin size €1,000 for maternal earnings. The proxy treatment and proxy control groups are defined here based on a Random Forest prediction algorithm with the top 6 percentiles constituting the proxy treatment group and the bottom 60 percentiles the proxy control group (see Appendix D for details).

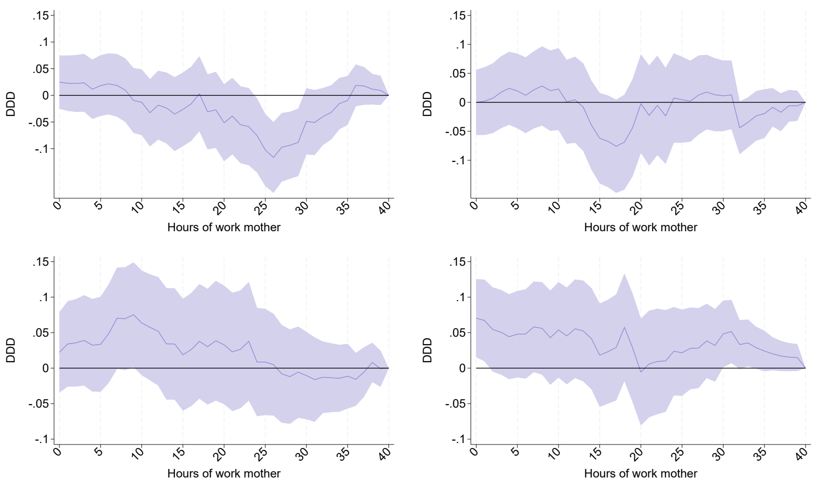
Figure 7: Heterogeneity in DDD estimates on maternal hours of work




(a) By quartiles of hours worked when older sibling was aged 2



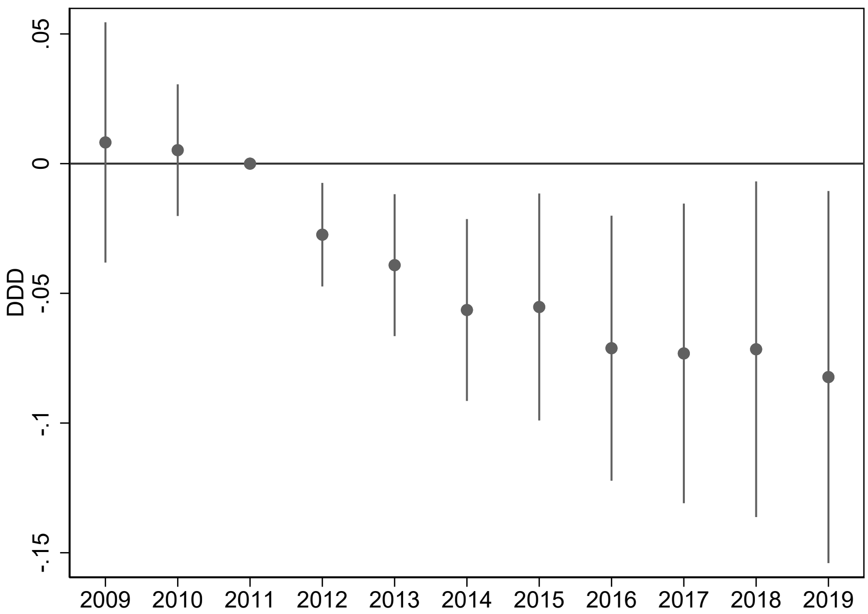

(b) By quartiles of hours worked when older sibling was aged 2




(c) By quartiles of father's taxable income when older sibling was aged 2

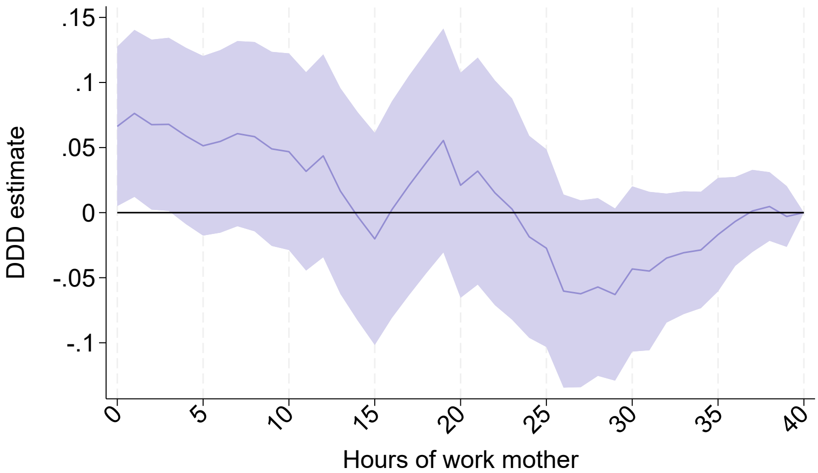
*Notes*: As in Figure 4.

Figure 8: *Coefficient plots of a triple difference specification by year.*

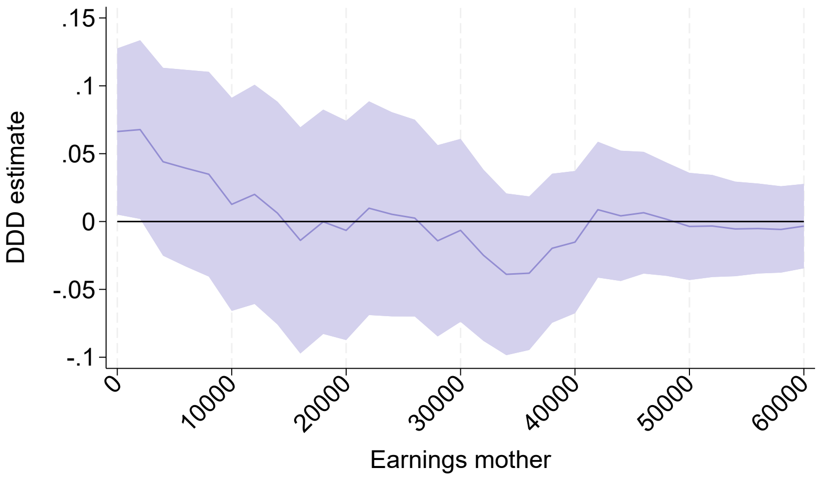



*Notes*: DDD estimates presented in an event study framework, with three-way interactions between virtually constrained, younger-child and the individual year dummies. 2011 (i.e., one year pre-reform) is the baseline year. The NYC sample is restricted not to have any further children until the final observation year 2019, such that the sample composition slightly differs from the baseline sample event study shown in Figure 5.

Figure 9: DDD estimates on the intensive margin in the long run (i.e., pooled 2016-2019 coefficients)



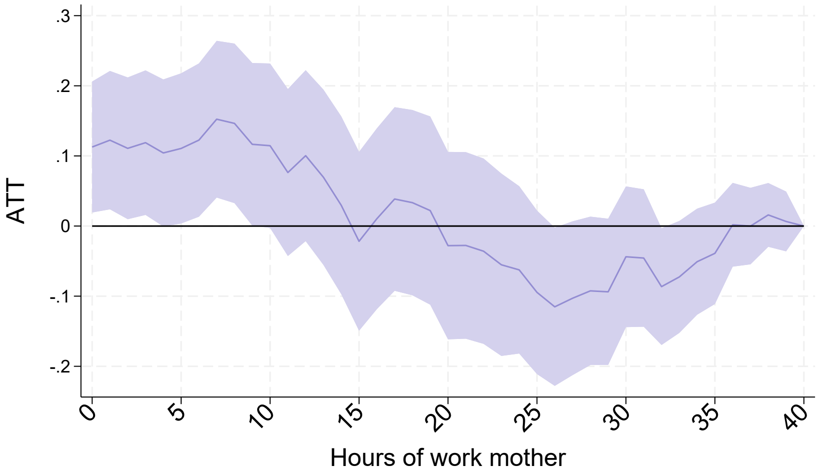

(a) Average weekly hours of work of the mother



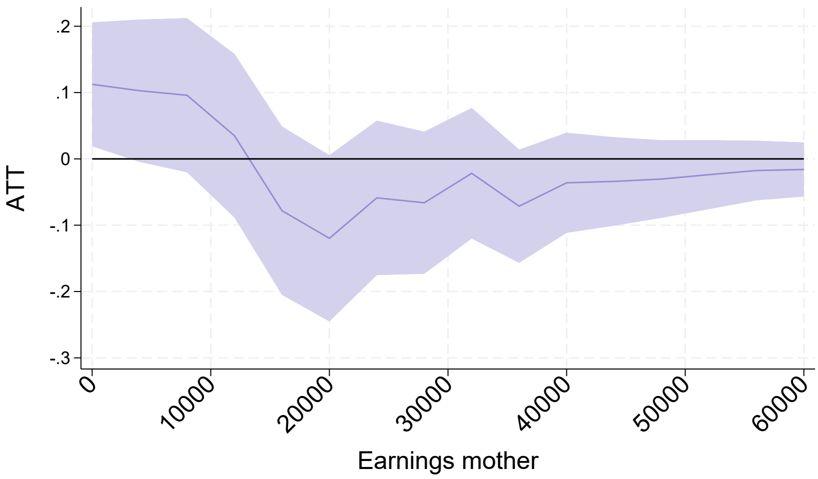

(b) Yearly earnings of the mother

*Notes*: DDD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on maternal hours of work (panel 9a) and maternal earnings (panel 9b). The presented coefficients $\rho$ and associated confidence intervals stem from equation (4), but with two modifications. First, we include three-way interactions between virtually constrained, younger-child and the pooled year dummies 2016-2019 instead of the binary indicator 'post' to study the long-run response. Second, the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for maternal weekly work hours, and bin size €1,000 for maternal earnings as in Figure 4.

Figure 10: ATT estimates on maternal work hours and earnings




(a) Weekly hours of work of the mother




(b) Yearly earnings of the mother

*Notes*: GMM estimates of the ATT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on maternal hours of work (panel 10a) and maternal earnings (panel 10b). The presented coefficients $\lambda$ and associated confidence intervals stem from GMM joint estimation of equations (7) and (8), where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for maternal weekly work hours, and bin size €1,000 for maternal earnings as in Figure 4. The three-way interaction terms between control variables and and two-way interactive fixed effects between the virtual constraint dummy, YC dummy, and year dummies selected based on a post-double selection procedure.

# *Online Appendix*

## A Supplementary Tables and Figures

Table A.1: Placebo tests on maternal employment

| | Basic | Saturated | PDS |
|---|---|---|---|
| YC | | $N = 61,278$ | |
| DD | 0.000 | −0.010 | −0.009 |
| | (0.010) | (0.013) | (0.012) |
| NYC | | $N = 208,650$ | |
| DD | −0.002 | −0.004 | −0.003 |
| | (0.004) | (0.005) | (0.005) |
| Pooled | | $N = 269,928$ | |
| DDD | 0.003 | −0.010 | −0.009 |
| | (0.011) | (0.015) | (0.015) |

*Notes*: Placebo DD and DDD Estimates where we restrict the data to before 2012 only, and the placebo 'post' period is 2011. Rest as in Table 2. * p<0.1, ** p<0.05, *** p<0.01.

Table A.2: Effects on maternal employment - Machine learning prediction

| | Basic | Saturated | PDS |
|---|---|---|---|
| Mother working | −0.035*** | −0.034*** | −0.032*** |
| | (0.007) | (0.011) | (0.011) |
| $N$ | 403,365 | 403,365 | 403,365 |

*Notes*: DDD estimates where proxy treatment status is based on Random Forest prediction. Proxy treatment group is defined as top 6 percentiles of the predicted probability of being constrained, proxy control group is defined as the bottom 60 percentiles. Rest as in Table 2. * p<0.1, ** p<0.05, *** p<0.01.

Table A.3: Effects on employment and child care among first children

| | Mother working | | | Child care use | | | Hours of care | | |
|---|---|---|---|---|---|---|---|---|---|
| | Basic | Saturated | PDS | Basic | Saturated | PDS | Basic | Saturated | PDS |
| DD | −0.034*** | −0.029*** | −0.028*** | −0.103*** | −0.028** | −0.031*** | −241.464*** | −114.960*** | −124.030*** |
| | (0.010) | (0.010) | (0.010) | (0.012) | (0.012) | (0.012) | (26.531) | (20.271) | (20.415) |
| $N$ | | 54,213 | | | 54,213 | | | 54,213 | |

*Notes*: DD regression based on the sample of firstborn children. Proxy treatment families are identified based on virtual constraint status at age 2 pre-reform, and outcomes are evaluated at age 3 post-reform. Rest as in Table 2. * p<0.1, ** p<0.05, *** p<0.01.

Table A.4: First stage effects among subgroups

| Maternal hours of work | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| Proxy treatment | 0.158*** | 0.236*** | 0.240*** | 0.148*** |
| | (0.007) | (0.011) | (0.015) | (0.011) |
| Constant | 0.030*** | 0.033*** | 0.033*** | 0.031*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $N$ | 23,344 | 29,609 | 38,505 | 39,253 |
| Maternal hourly wage | Q1 | Q2 | Q3 | Q4 |
| Proxy treatment | 0.185*** | 0.179*** | 0.196*** | 0.184*** |
| | (0.010) | (0.010) | (0.011) | (0.009) |
| Constant | 0.038*** | 0.033*** | 0.029*** | 0.028*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $N$ | 29,218 | 34,442 | 36,234 | 30,817 |
| Partner's income | Q1 | Q2 | Q3 | Q4 |
| Proxy treatment | 0.175*** | 0.182*** | 0.192*** | 0.192*** |
| | (0.009) | (0.011) | (0.011) | (0.010) |
| Constant | 0.032*** | 0.031*** | 0.027*** | 0.037*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $N$ | 36,280 | 34,361 | 31,613 | 28,457 |

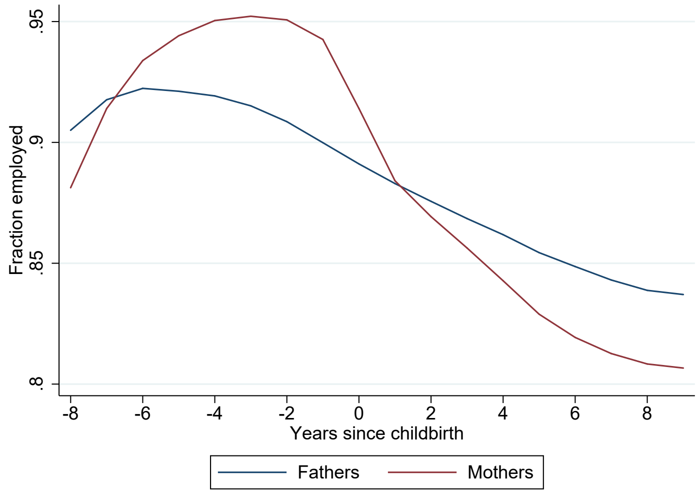
*Notes*: First stage regression based on data pre-reform of whether the younger child was virtually constrained on the proxy treatment (whether the older sibling was virtually constrained). First panel is for four quartiles of hours of work of the mothers when the older sibling was aged 2; second panel is for four quartiles of maternal hourly wage when the older sibling was aged 2; third panel is for four quartiles of partner's income. * p<0.1, ** p<0.05, *** p<0.01.

Table A.5: DDD results for father's employment

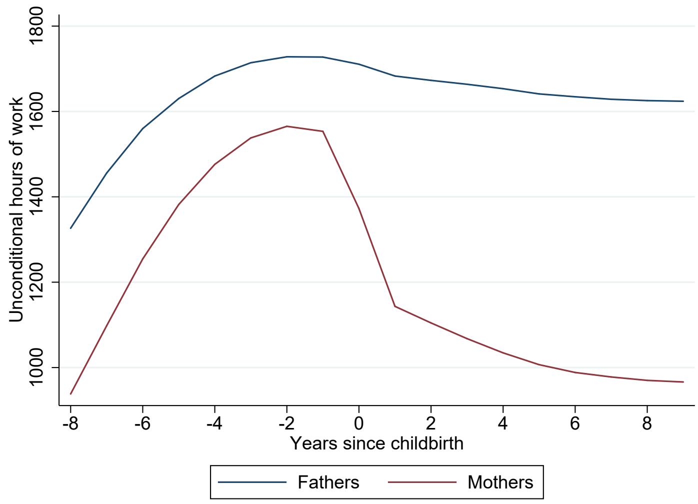
| | **Father working** | | | **Father self-employed** | | |
|---|---|---|---|---|---|---|
| | Basic | Saturated | PDS | Basic | Saturated | PDS |
| Pooled - $N = 602,035$ | | | | | | |
| *DDD* | −0.007 | −0.005 | −0.005 | −0.02** | −0.029** | −0.031** |
| | (0.005) | (0.007) | (0.007) | (0.010) | (0.013) | (0.014) |
| | **Father working** | | | **Mother working** | | |
| | Basic | Saturated | PDS | Basic | Saturated | PDS |
| Families with father least working partner - $N = 44,836$ | | | | | | |
| *DDD* | −0.045*** | −0.028 | −0.032 | 0.005 | 0.031 | 0.032 |
| | (0.023) | (0.034) | (0.034) | (0.016) | (0.025) | (0.024) |

*Notes*: Estimates for $\rho$ from the triple-difference regression. The columns 'Basic' includes only standard DDD fixed effects; 'Saturated' includes all main effects, one-way and two-way interactions between the control variables and DDD fixed effects; 'PDS' refers to a model where the main effects, one-way and two-way interactions with the control variables are selected on basis of post-double selection. Robust standard errors, clustered at the household level, are reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

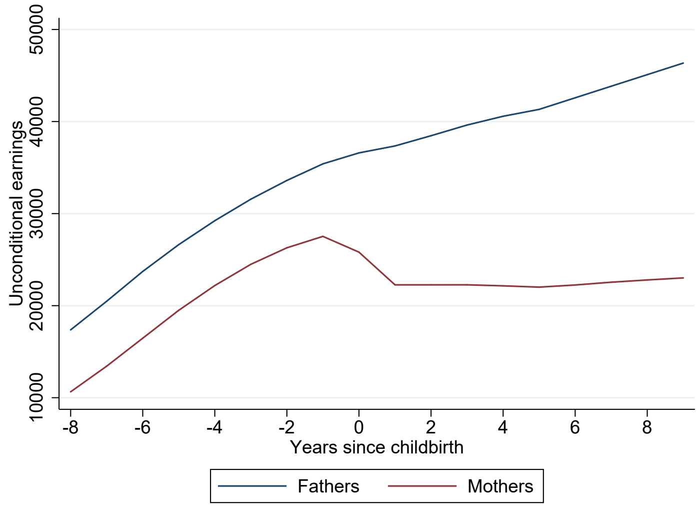
Figure A.1: Evolution of labor market outcomes for mothers and fathers around childbirth




(a) Fraction employed

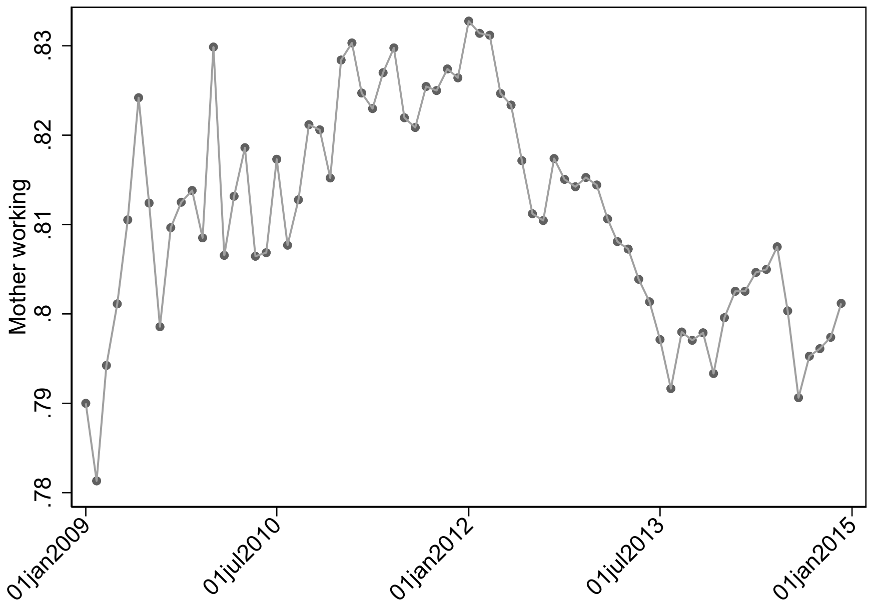



(b) Yearly hours of work




(c) Earnings

Figure A.2: Monthly labor supply for virtually constrained mothers 2009-2015.




*Notes*: Monthly employment status (working equals 1 if receiving income from wage-work and/or self-employment) of mothers with a younger child who were virtually constrained for an older sibling pre-reform.

Figure A.3: Maximum subsidizable price per hour over the years for daycare.

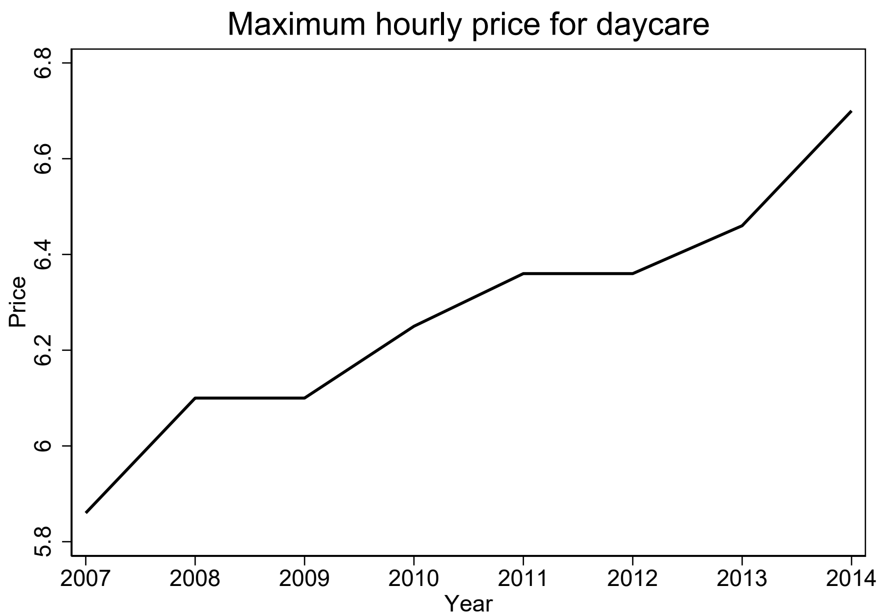
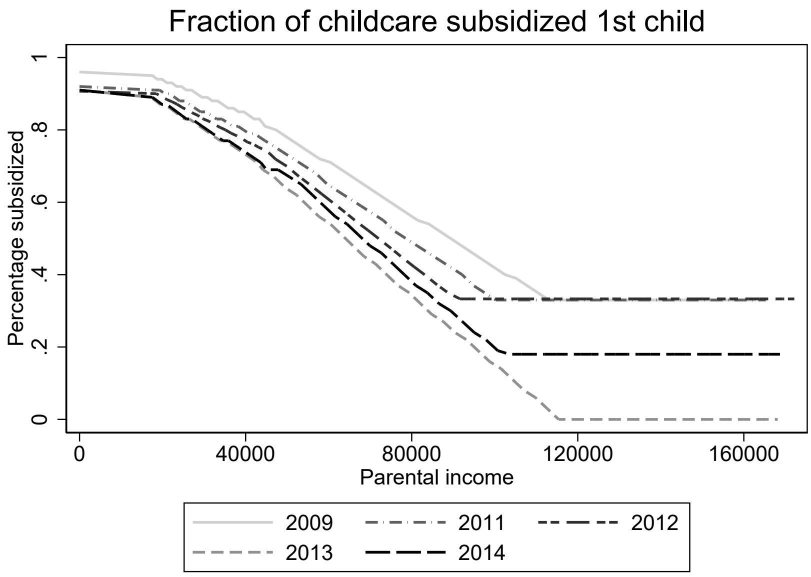
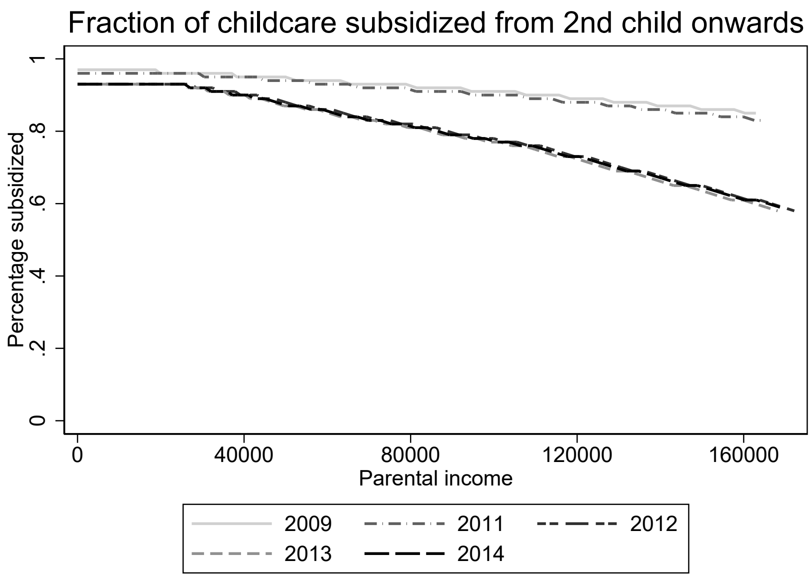


Figure A.4: Fraction of child care costs subsidized over the years by child rank

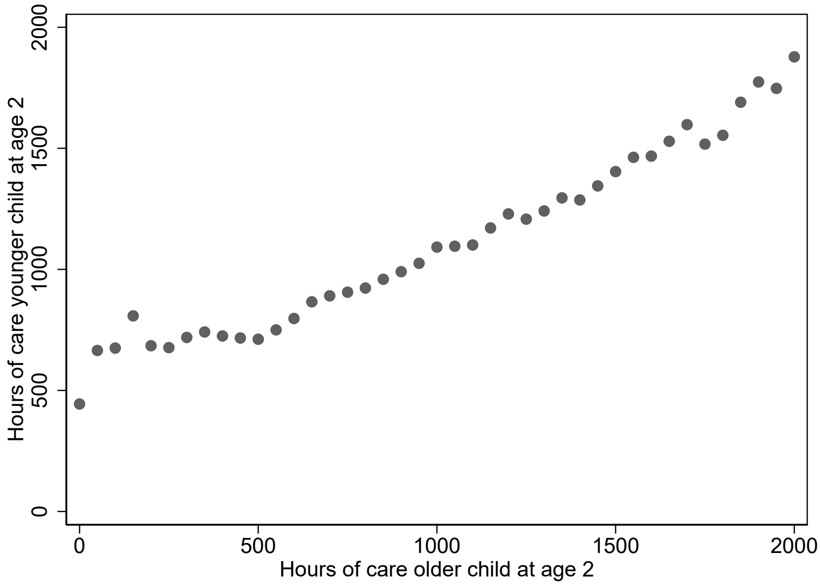



(a) 1st child



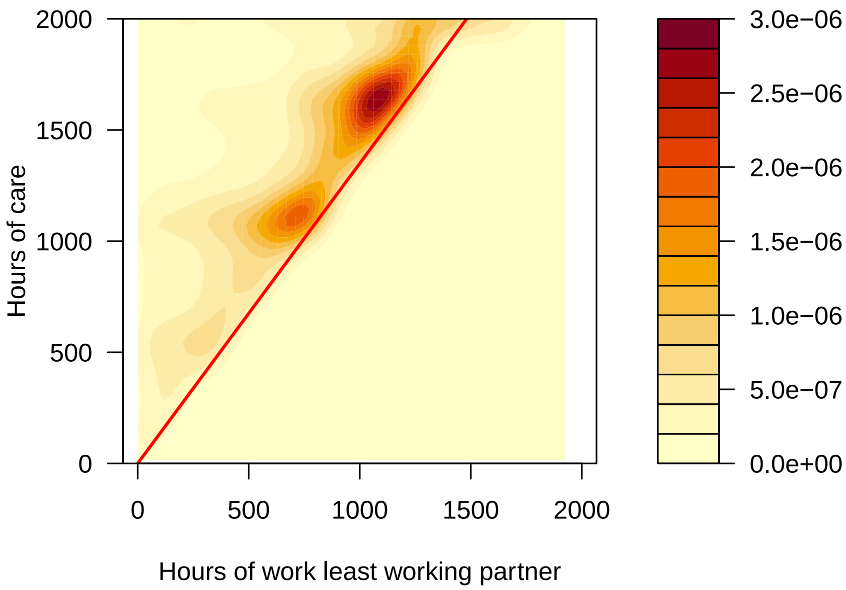

(b) 2nd or higher-parity child

Figure A.5: Average child care hours for the younger sibling versus that of the older sibling




*Notes*: Scatter plot of hours of subsidized childcare of the older sibling versus that of the younger sibling. Both are measured pre-reform and at age 2.

Figure A.6: Contour plot of childcare hours vs. workhours of the least working partner pre-2012



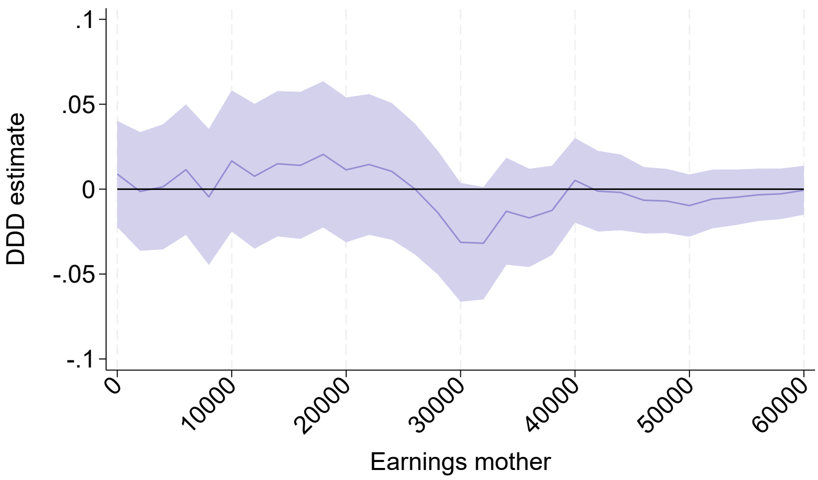

*Notes*: Heat map showing how the composition of our proxy treatment (virtually constrained) group is distributed over pre-reform hours of work and hours of care. The red line indicates the threshold (1.4× the hours of the least working partner), and only virtually constrained families are shown. The figure indicates mass throughout the distribution, with sligthly higher mass just over 1,000 yearly hours of work (or 20 weekly hours) of the least working partner.

Figure A.7: DDD estimates on the intensive margin


*Notes*: DDD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on maternal taxable income. The presented coefficients $\rho$ and associated confidence intervals stem from equation (4) where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size €1,000.

Figure A.8: Placebo DDD estimates for maternal hours of work and maternal earnings

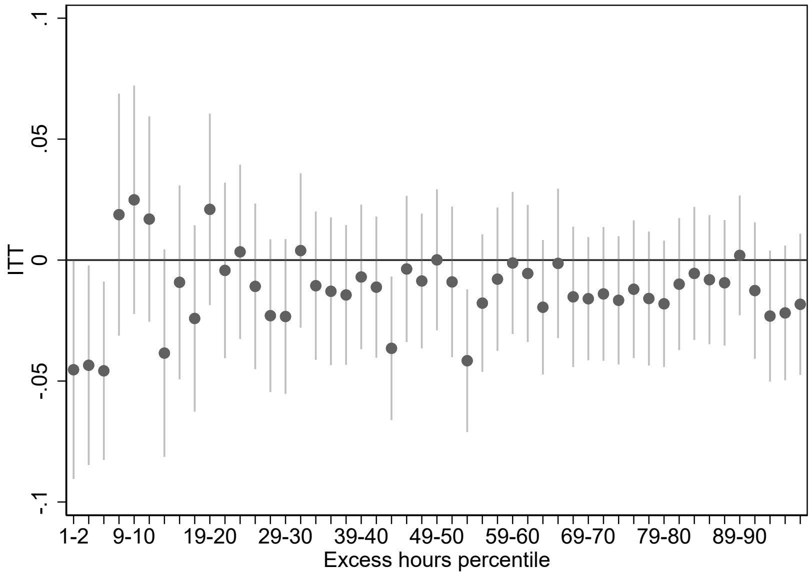

(a) Maternal hours of work




(b) Maternal earnings

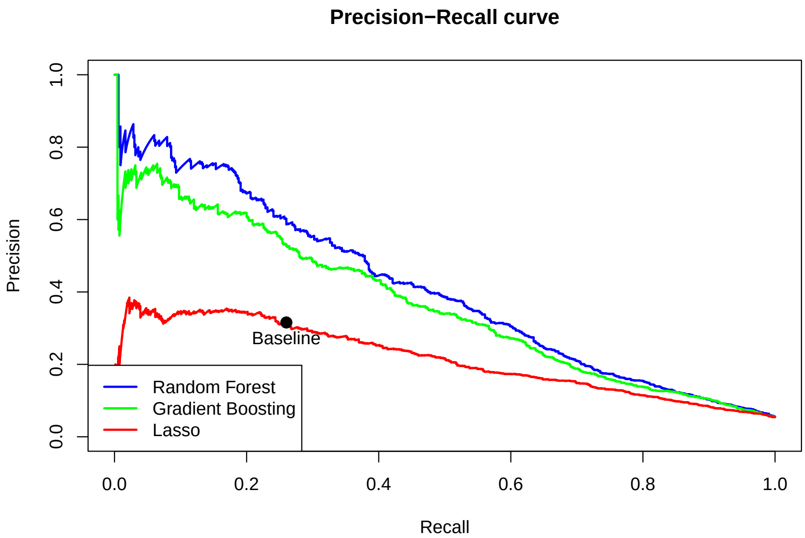
*Notes*: DDD estimates of placebo ITT effect of a placebo 2011 reform using distribution regressions on the cumulative distribution function (CDF) on maternal hours of work (panel A.8a) and maternal earnings (A.8b). Data based on pre-reform (i.e., before 2012) only. The presented coefficients $\rho$ and associated confidence intervals stem from equation (4) where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for maternal hours of work and bin size €1,000 for maternal earnings.

Figure A.9: *DDD estimates on mother working by excess hours percentile*



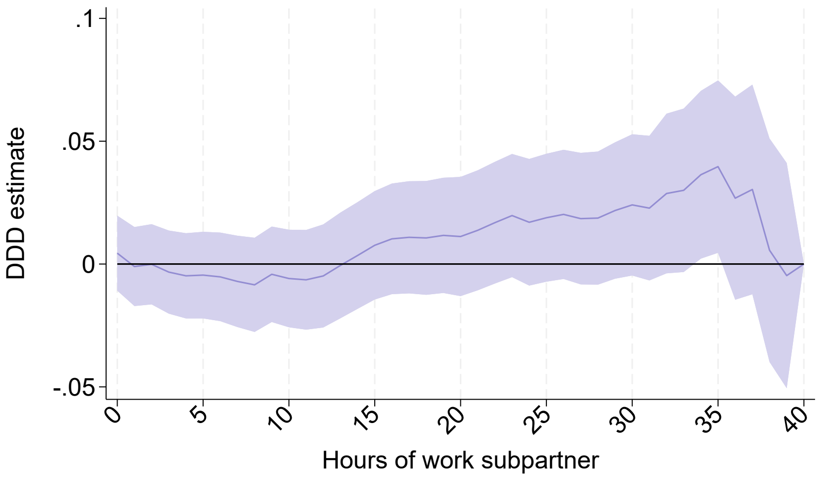

*Notes*: DDD estimates of the ITT effect of the 2012 reform, separately for 50 bins of the distribution of excess hours (1.4× hours of work − hours of care). The bottom 6 percentiles (i.e., the bottom 3 bins) constitute our virtually constrained group. The presented coefficients $\rho$ and associated confidence intervals stem from equation (4), where we replaced the three-way interaction $\big(G_i \times Z_i \times \mathbf{1}(T_i \geq 2012)\big)$ with the three-way interaction $\big(G_i \times B_i \times \mathbf{1}(T_i \geq 2012)\big)$, where $B_i$ denotes the set of 49 bins, leaving out the top bin.

Figure A.10: Precision versus Recall for the three ML algorithms

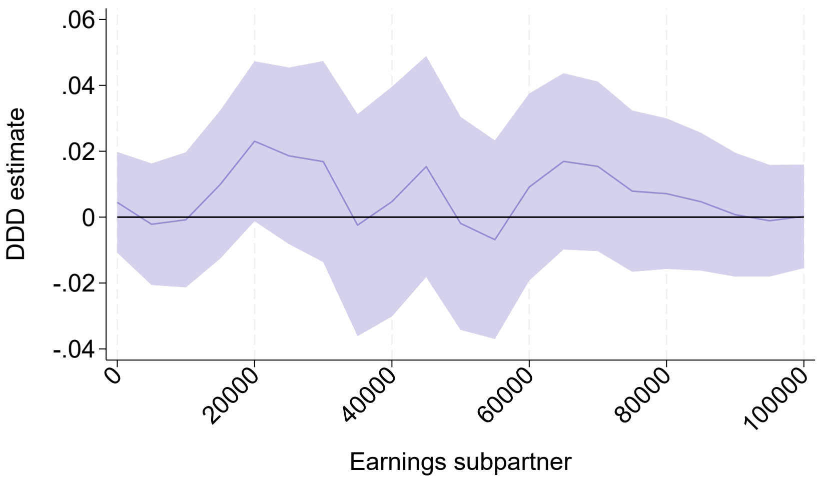



*Notes*: Precision–recall curves for the three machine learning algorithms considered in the analysis. Precision measures accuracy of positive predictions and is defined as the true positives over the total positives (true + false positive), while recall measures ability to find all actual positives (true positives over true positives + false negatives). Each curve traces model performance across all possible classification thresholds. Higher curves indicate better performance, reflecting a higher precision for a given level of recall. Baseline precision and recall refer to the binary classification of virtually constrained families we use in our main analysis.

Figure A.11: DDD estimates on hours of work and earnings of the father




(a) Hours work of the father




(b) Earnings of the father

*Notes*: DDD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on subsidy partner's hours of work (panel A.11a) and subsidy partner's earnings (panel A.11b). The presented coefficients $\rho$ and associated confidence intervals stem from equation (4) where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for weekly work hours, and bin size €1,000 for earnings.

# B Institutional background

## B.1 Changes in the childcare subsidy scheme

The primary reform we study is the 2012 constraint on subsidized child care hours based on work hours of the lesser-working parent. As mentioned in the main text, this reform was announced in June 2011 and took effect on January 1, 2012. Here, we discuss potential anticipation effects of the reform, as well as other concurrent changes in the subsidy scheme.

**Anticipation:** Figure A.2 shows monthly averages of the proportion of mothers working in virtually constrained families with a younger child (i.e., the group predicted to be most affected by the 2012 reform). The figure shows a sudden trend change in January 2012, coinciding with the reform's implementation. This suggests that there was little anticipation of the reform and that families understood its implications.

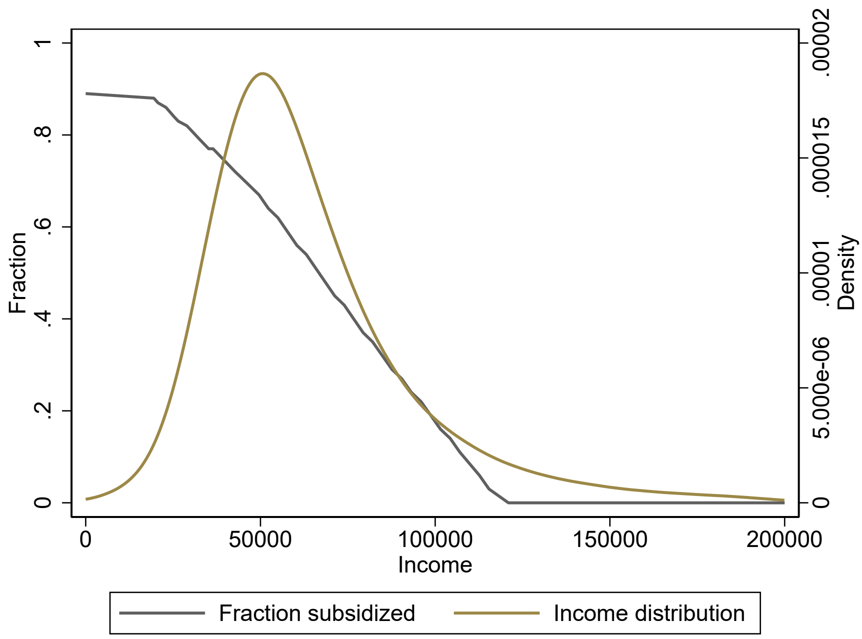
**The maximum price per hour $P_t^{max}$** Figure A.3 shows the maximum price per hour over the years for daycare. It can be seen that the maximum hourly price of daycare was usually adjusted for inflation. The actual prices followed a slightly different pattern: the average price for child care was lower than, but close to, the maximum price in 2007; while in 2013 the average price was 1.7% higher than the maximum price for day care (Van der Wiel and Van 't Riet, 2011).

**Subsidy rates** Child care costs are financed approximately one-third each by the government, employers (via a payroll premium), and parents. Figure A.4 shows subsidy rates over time for the first child (Panel a) and higher-parity children (Panel b). In addition to the 2012 hours constraint we study, there were contemporaneous changes in subsidy rates. The largest cuts for the first child occurred in 2013 and were concentrated among high-income households.

To limit exposure to these changes, our baseline sample excludes families with joint income above €90,000. As shown in Figure B.1, this restriction removes nearly all households in the income range where the reimbursement schedule flattens or drops to zero (above €118,000). Results are unchanged when further restricting to joint income below €60,000 (Appendix E.4). Since the reimbursement cuts were concentrated at the top of the income distribution, these restrictions substantially reduce potential confounding.

Another concern is the 2012 reduction in reimbursement rates for higher-parity children (Figure A.4, Panel b). Because child care use is serially correlated, virtually constrained families may use more care for older siblings and thus be more exposed to this cut. Our

Figure B.1: *Fraction subsidized in 2013 for 1st child and the income distribution.*

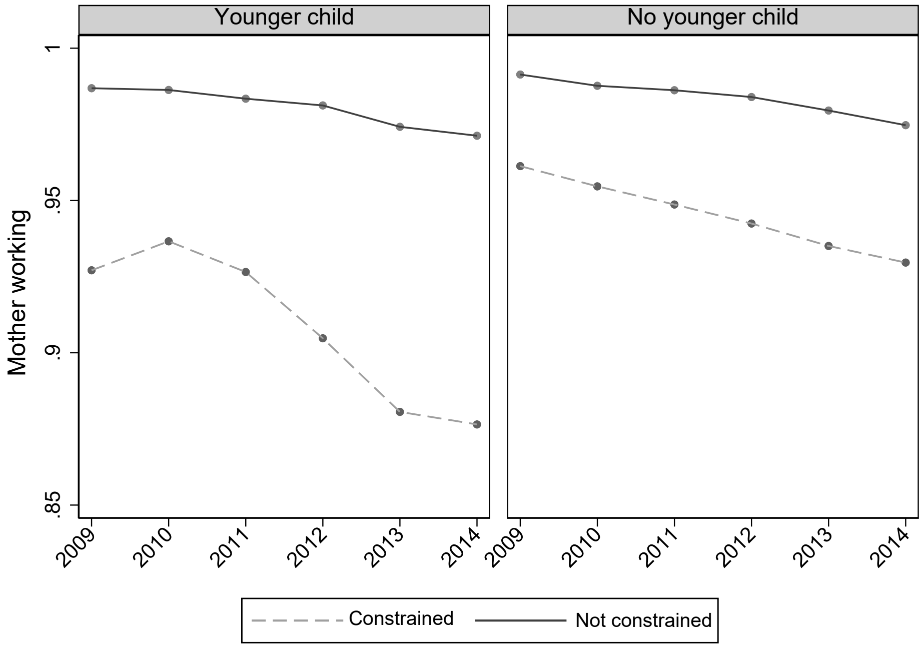



DDD specification absorbs the aggregate effect of the 2012 reimbursement reform through interactions between the proxy treatment indicator ($Z$) and year fixed effects. However, younger-child (YC) and no-younger-child (NYC) families are mechanically assigned to different reimbursement schedules (higher-parity vs. first-child). In principle, this could generate differential effects unrelated to the hours cap.

We address this concern in four ways. First, restricting the sample to joint income below €90,000 — and further to €60,000 in a robustness analysis — removes the households most affected by subsidy rate cuts. Estimated effects are stable across these samples. Second, explicitly controlling for (changes in) subsidy rates yields nearly identical estimates (Section E.4). Third, the response is sharply localized at the hours constraint threshold. Figure A.9 plots DDD estimates by percentiles of exposure, defined as 1.4× lesser-working partner's hours of work *minus* hours of care, with negative values indicating a family was virtually constrained. The decline in maternal employment is concentrated precisely in the bottom 6 percentiles where the constraint binds. No comparable effect appears beyond the 6 percentile threshold where the constraint was virtually binding. Such a discrete pattern cannot be attributed to reimbursement changes that vary smoothly with income. Fourth, among very intensive users of care ($>$ 900 hours annually), Figure B.2 shows responses are confined to only virtually constrained YC families. A subsidy rate-driven mechanism would predict broader effects across exposed groups.

In sum, the estimates hold when imposing income restrictions that exclude households most exposed to reimbursement cuts and when controlling directly for reimbursement schedules, and always display a sharp nonlinearity at the hours cap threshold. Taken

Figure B.2: Labor supply for mothers with high baseline child care use (i.e., > 900 hours).




Notes: All families have a child born between 2005-2009. Left panel refers to families with a younger child, while right panel refers to families without a younger child. The dashed line refers to families who were virtually constrained by the 2012 rules when the older sibling was aged 2. The solid line refers to families who were not virtually constrained (i.e., using more child care hours than allowed under the 2012 rules.)

together, these results make it unlikely that contemporaneous subsidy rate adjustments explain our findings. The empirical pattern aligns with a binding hours constraint, not with smooth price variation from reimbursement reforms.

**Treatment of host parent care** Apart from the decrease in the maximum hourly price for host parent care in 2010, the government additionally imposed tighter regulation of host parent care in the same year. In particular, host parents now require a vocational degree in social care or nursing and there is a tighter control on finance and quality. Host parent care represented 20% of the formal child care market in 2007; this increased to 24-25% in 2008/2009, before reducing in 2010 to a stable 15-17% for the period 2010-2014. Within host parent care, Intomart/GfK (2011) shows that the fraction of grandparents providing host parent care reduced from 47% in 2010 to 25% in 2011. While the substitution between different types of formal child care is interesting on itself, in this paper we are mainly interested in the effect of subsidies on aggregate formal child care use, and so we do not treat host parent care different from other types of formal child care.

## B.2 Parental leave policies

There are three types of parental leave in the Netherlands. First, maternity leave consists of 16 weeks, of which at least 4 weeks have to be taken before birth, and at least 6 after birth. It is paid at 100% up to a ceiling equivalent to the maximum daily payment for sickness benefit. Since it is mandatory, take up of maternity leave is near 100%. Second, during our study period there were also two days of paternity leave after birth, paid at 100% without a maximum. In 2013, 83% of fathers took leave, and 60% of these prolonged their leave by taking annual leave (den Dulk, 2017). Take up of paternity stayed constant between 2011 and 2017 (Korvorst, 2019; de la Croix et al., 2014). Third, both parents are entitled to parental leave for 26 working weeks per child since 2009. It is not paid, but some collective agreements offer partial payment of the leave. This leave can be taken until a child is 8 years old, and parents are flexible in how they want to use the leave. More than 50% of parents take parental leave before their kid turns two (CBS, 2011). Take-up of parental leave is higher for higher educated parents[1] and varies by sector. Public employees and health care workers more often have partially paid leave, and are also more likely to take up parental leave.

# C Measurement and descriptive statistics

## C.1 Child care use

While employment outcomes are our primary interest, the nature of the reform makes it informative to study responses on child care use as well. Table C.1 shows the effect of the reform on child care outcomes. These results should be interpreted with caution for two reasons. First, we observe only *subsidized* child care hours, which are not necessarily *actual* child care hours. After 2012, child care hours are censored at the maximum eligible hours, so there could be a mechanical drop in subsidized hours even when actual child care hours remain unchanged. We relieve some of this concern in Section C.2, where we compare actual versus subsidized child care use for a sub-sample with both measures and show that the two measures are closely aligned. Second, since families without a younger child by definition do not use child care, we can only rely on a DD specification instead of DDD. Even though the rich controls plausibly account for non-parallel trends across the virtually constrained and non-constrained YC families, the lack of a control group without any younger child renders the estimates somewhat less robust to differential dynamics between the virtually constrained and non-constrained YC families.

With these caveats in mind, Table C.1 shows that child care use decreased by 6.6 percent-

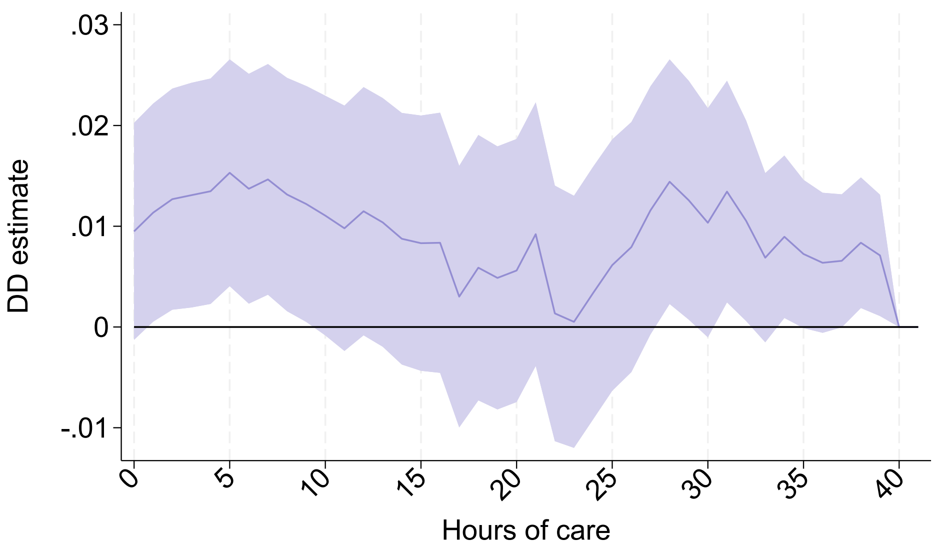
[1]Lower educated individuals are more likely to work in a sector where parental leave is unpaid, and at the same time they are less likely to be able to afford taking unpaid leave.

Table C.1: Main results for child care use

| | Child care use | | | Hours of child care | | |
|---|---|---|---|---|---|---|
| | Basic | Saturated | PDS | Basic | Saturated | PDS |
| YC - $N$=130,711 | | | | | | |
| DD | $-0.105^{***}$ | $-0.066^{***}$ | $-0.066^{***}$ | $-200.265^{***}$ | $-137.158^{***}$ | $-136.210^{***}$ |
| | (0.012) | (0.014) | (0.014) | (21.211) | (21.368) | (21.335) |
| $\mu_{s2}$ | 0.633 | 0.633 | 0.633 | 637.195 | 637.195 | 637.195 |

*Notes*: The row DD is a difference-in-difference specification among families with a younger sibling. The columns 'Basic' include a basic set of fixed effects of the DD specification; 'Saturated' includes all two-way interactions between the control variables and the DD fixed effects; 'PDS' refers to a model where the two-way interactions are selected on basis of post-double selection. Robust standard errors, clustered at the household level, are reported in parentheses. $\mu_{s2}$ is the mean of the relevant variable when the older sibling was aged 2. * p<0.1, ** p<0.05, *** p<0.01.

Figure C.1: DD estimates on hours of child care




*Notes*: DD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) of subsidized hours of childcare. A bin size is used 1 for weekly childcare hours.

age points at the extensive margin for the virtually constrained YC families compared to the non-constrained YC families — approximately a 10% reduction relative to the average child care use for the older sibling. This extensive-margin effect is mirrored in a reduction of 135 hours, or around 20%, in hours of child care. Figure C.1 shows the corresponding DD effect on the distribution of child care hours, revealing a reduction in hours of child care across virtually the entire distribution. While these results could be somewhat biased by the mechanical censoring or the lack of a control group to account for differential dynamics by the older sibling's virtual constraint status, plausibly part of the effects are genuine. Therefore, it seems safe to conclude that the 2012 reform led to reductions in child care use at both the extensive and intensive margins.

## C.2 Subsidized versus actual child care use

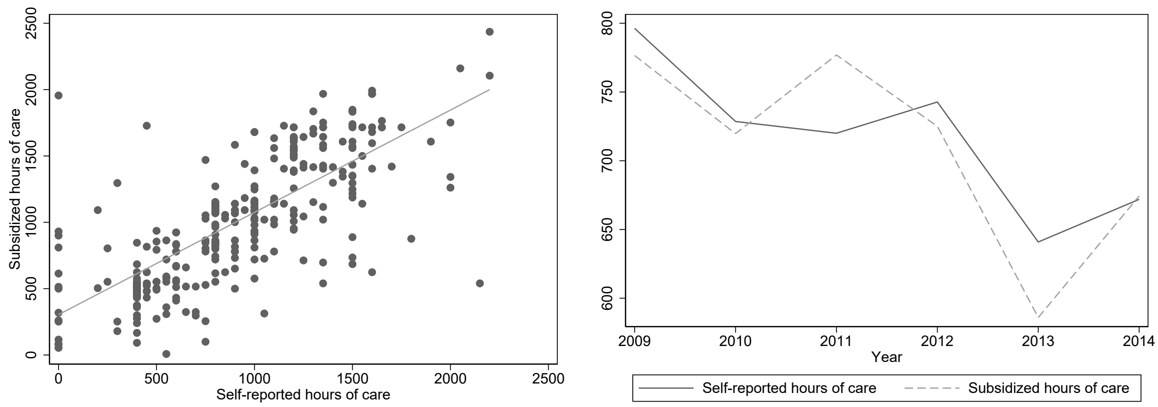
**Identification of the partner** The starting point of our analysis is a mother-child pair living in the same household. To determine the actual partner of the mother, which is the relevant 'subsidy partner' for the tax authorities, we use three separate data sources: the parent-child linkages, the marital status register and the municipality register with information on the partner living together in the same household. In more than 75% of the cases, the other parent of the joint child is living together with the mother, in which case the subsidy partner is identified. In cases where these deviate, we use the marital status and municipality register on partners, but these only include some basic demographic characteristics of the spouse, but not the identity. The bulk of the remaining cases represents the situation in which an individual is married to someone other than the parent of the individual's child, in which case we again can identify the subsidy partner. Given that the marital status register lacks the identity, we match identities by comparing the month of birth, gender, and country of birth of the spouse and partner. Only in very few cases where (i) an individual is married to someone other than the parent of the child, but they are not living together, or (ii) an individual is living together with someone other than the parent of the child, but they are not married, we cannot assign the subsidy partner with certainty. We treat these individuals as being single and exclude them from the analysis.

**Measurement of subsidies and subsidized child care** To test the accuracy of our data, we compared the actual subsidy received with the predicted subsidy based on Equation (1), using the hours of child care used, the average price per hour, the child's rank, and the taxable income from the individual and his or her subsidy partner. The correlation between the two is over 99% indicating that our formula for determining child care subsidies is highly accurate, and we can identify the subsidy partner with great accuracy.

Given the administrative records we use, there is also very little measurement error in our measure for subsidized child care use. A remaining issue is however that we observe *subsidized* but not *actual* child care use. If parents are not eligible for subsidies, we would not observe their actual child care use even if they were still using child care. Before 2012, this applied only to families where at least one of the parents was not working. After 2012, subsidized child care hours is additionally censored based on the hours worked of the lesser-working parent. Therefore, when studying *subsidized* child care use, a mechanical reduction in subsidized child care use does not necessarily reflect a change in actual child care use after 2012.

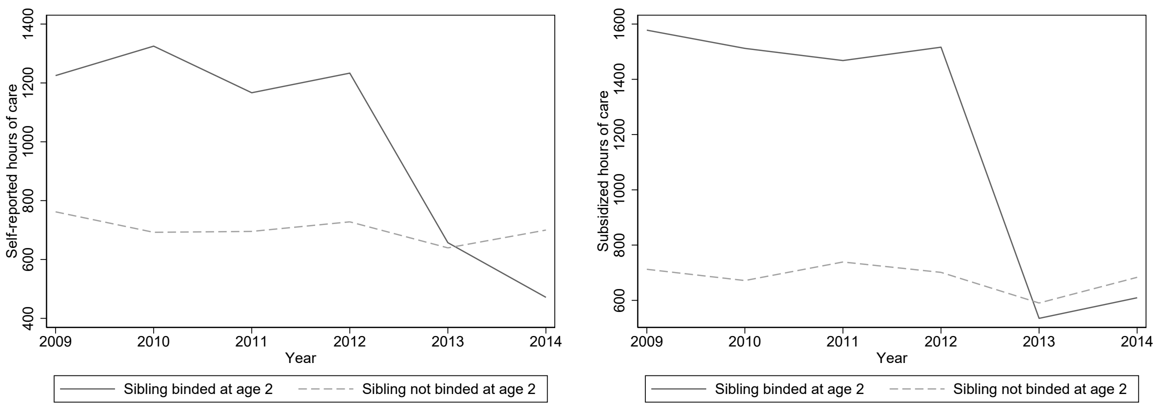
To gauge the severity of this limitation, we rely on the EU-SILC survey that can be linked to our administrative records for a small subsample of interviewed households. In Figure C.2 we first explore the correspondence of the self-reported actual hours of formal child care versus the subsidized hours of child care over all years (left-panel), and then the mean levels of each measures over the years (right-panel). There is a close connection between self-reported actual and subsidized child care hours. We also do not see any major deviations between the means after 2012. This provides some reassurance that subsidized hours of care provides a reliable proxy for actual child care hours.

Figure C.2: Subsidized hours of child care on basis of administrative data versus actual hours of childcare on basis of self-reports in survey data (left) and the mean values for each over the years (right).




Although this is based on a very small sample, Figure C.3 shows that, among the subsample for whom we observe both subsidized and self-reported actual child care use, child care hours drop similarly after 2012 for families whose focal child was virtually constrained by the 2012 reform. We do not see a comparable decrease for families whose focal child was not virtually constrained. This provides further reassurance that the decline in subsidized child care hours is not merely mechanical but reflects a decline in actual child care hours, although the exact magnitude of the drop may be subject to some uncertainty.

Figure C.3: Evolution of self-reported actual child care hours for families in which the focal child was virtually constrained or not (left) and the same for subsidized hours (right).




## C.3 Comparison to full sample

In this subsection we compare our estimation sample to the full sample of all families with a child born between 2005 and 2009. In our estimation sample, we restrict to families in which both parents were working as employees when the focal child was aged 2, because only then can we meaningfully determine whether the focal child was virtually constrained by the hours requirement imposed in the 2012 reform. Further, among families with another younger child, we restrict the sample to those with a younger child born 3 years after the focal child. Finally, we restrict the sample to families with a joint income below €90,000. These restrictions may lead to some sample selection. Table C.2 compares the mean of various variables across the full and estimation samples. Given the large sample size, all differences are statistically significant and are therefore omitted from the table for brevity.

Given our sample selection, it is not surprising that the fraction working is lower in the full sample, especially among mothers. At the same time, our estimation sample excludes a sizable number of self-employed mothers and fathers. This restriction may also explain differences in education and earnings between our estimation sample and the full sample. Families in our estimation sample are slightly more likely to use child care, but the number of hours is quite comparable. In terms of demographics, the differences are minimal. Overall, while some predictable differences exist between our estimation sample and the full sample, we believe our estimation sample is still representative of families in the Netherlands with young children during the period studied, at least among those with both parents working as employees.

Table C.2: Descriptive statistics for full versus estimation sample

| Variable | Full | Estimation |
|---|---|---|
| Gender focal child | 0.51 | 0.51 |
| Rank focal child | 1.86 | 1.88 |
| Born in the Netherlands | 0.85 | 0.89 |
| Education of the mother | | |
| Missing | 0.29 | 0.30 |
| Lower | 0.10 | 0.08 |
| Lower vocational | 0.24 | 0.28 |
| Higher vocational | 0.22 | 0.23 |
| University | 0.15 | 0.11 |
| Education of the father | | |
| Missing | 0.34 | 0.34 |
| Lower | 0.09 | 0.08 |
| Lower vocational | 0.21 | 0.26 |
| Higher vocational | 0.20 | 0.21 |
| University | 0.16 | 0.11 |
| Working mother $s_2$ | 0.82 | 1.00 |
| Working father $s_2$ | 0.96 | 1.00 |
| Self-employment mother $s_2$ | 0.11 | 0.00 |
| Self-employment father $s_2$ | 0.22 | 0.00 |
| Hours of work mother $s_2$ | 1145.67 | 1124.13 |
| Hours of work father $s_2$ | 1912.82 | 1945.35 |
| Earnings mother $s_2$ | $22,836.03$ | $19,942.03$ |
| Earnings father $s_2$ | $46,292.00$ | $40,362.20$ |
| Childcare use $s_2$ | 0.53 | 0.63 |
| Hours of care $s_2$ | 1044.39 | 1008.43 |
| N | 1,796,344 | 602,035 |
| Unique families | 485,912 | 208,872 |

*Notes*: Descriptive statistics (means) for the full sample of all parents with a child born between 2005 and 2009 (column 1), compared with our estimation sample (column 2). $s_2$ denotes that the variable is measured when the focal child (i.e., the older sibling) was aged 2.

# D Machine Learning Prediction

This section outlines our machine learning prediction approach, detailing the predictors used, prediction procedure, algorithms employed, parameter tuning methods, evaluation metrics, and the classification of high- and low-probability groups.

## D.1 Selection of Predictors

In our baseline analysis, we use the virtual binding status of an older sibling at age 2 as the *sole* predictor for the younger sibling's binding status. This approach provides a clear dichotomous distinction between proxy treatment and control groups. For robustness, our machine learning prediction approach expands the predictor set with variables from four broad categories.

First, we include demographic information such as parents' age and educational attainment. Second, we incorporate work and income data measured when the older sibling was age 2, including each parent's employment status, working hours, earnings, taxable income, and the family's joint taxable income. Third, we add information on the older sibling's childcare use and hours at age 2. Fourth, we construct additional measures to assess the relevance of the hours constraint, including both the ratio and difference between actual childcare hours and entitled subsidizable hours (calculated as 1.4 $\times$ the hours worked by the lesser-working parent).

## D.2 Overview of the Prediction Procedure

As described in Section 3, our full dataset comprises 602,035 observations from 2007 to 2014, which are categorized into the YC sample (families with a younger child) and the NYC sample (families without a younger child). The YC sample is further divided into pre-reform and post-reform subsamples. For estimating the machine learning algorithms, we rely exclusively on the pre-reform YC subsample. However, with the hours constraint binding for the younger child in only 3,329 out of 61,279 cases in this subsample, our target variable to predict is highly imbalanced. This imbalance poses a challenge, as standard machine learning algorithms often struggle to predict such rare events effectively.

To address this challenge, we create a "balanced sample" that includes all observations with binding hours constraints for the younger child alongside an equal number of randomly selected observations without such constraints from the pre-reform YC subsample. This "balanced sample" is then split into two mutually exclusive subsets: a training sample (75% of the data) and a test sample (25%). Using the training sample, we train three machine learning algorithms: random forest, gradient boosting, and LASSO. Their performance is then evaluated on the test sample. After identifying the best-performing algorithm, we apply it to our complete dataset of 602,035 observations to predict the probability of each family being bound by the hours constraint.[2]

Although probabilities predicted from algorithms trained on the "balanced sample" are biased, they can be corrected using Bayes' rule to align with the original sample distribution (Einav et al., 2018). Let $D$ represent the younger child's binding status, $B$ indicate inclusion in the "balanced sample," and $\hat{p}_b(x) = \hat{P}(D = 1|B = 1, X = x)$ denote the predicted probability for $D = 1$ in the "balanced sample" for given predictors $X = x$. Let $s = P(B = 1|D = 0)$ represent the sampling probability for unconstrained families included in the "balanced sample". Using $\hat{p}_b(x)$ and $s$, we can calculate the corrected probability for families with predictors $x$ as:

$$\hat{p}_o(x) = \frac{s\hat{p}_b(x)}{1 - (1 - s)\hat{p}_b(x)}$$

[2]For families in the NYC sample, this probability reflects their likelihood of facing a binding hours constraint if they had a younger child three years younger than the "focal" child.

Importantly, since $\hat{p}_o(x)$ is a monotone function of $\hat{p}_b(x)$, the Bayes' transformation is rank preserving. This means when comparing model performance and setting thresholds for group selection, we can simply use the unadjusted predicted probability $\hat{p}_b(x)$.

## D.3 Tuning the Prediction Parameters

To optimize each machine learning algorithm's performance, we tune key parameters using cross-validation. We divide the training sample into five equal-sized folds. For each parameter configuration, we train the algorithm five times, each time using four folds for training and evaluating performance on the remaining fold. We then select the parameter configuration that maximizes our performance metric and retrain the model on the entire training sample.

We use the area under the curve (AUC) as our performance metric, which is standard in machine learning for classification tasks. AUC represents the area under the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate against the false positive rate. The AUC value indicates the probability that the model will assign a higher predicted probability to a randomly chosen positive case than a randomly chosen negative case. In our context, AUC helps assess how well an algorithm distinguishes between binding and non-binding cases, with higher values indicating better discriminative ability.

For random forest, we tune three key parameters: the number of trees, the number of variables considered at each split, and the minimum observations at each node. AUC is highest for 2,000 trees, 4 covariates randomly considered at each split (*mtry*), and the minimum number of observations per terminal node (*minimum leaf size*) equal to 1. For gradient boosting, we also tune three parameters: the number of trees, interaction depth, and learning rate. The highest AUC is achieved with 500 trees, an interaction depth of 7, and a learning rate of 0.1 For LASSO, we tune only the regularization parameter $\lambda$, with optimal performance at $\lambda = 0.00247$.

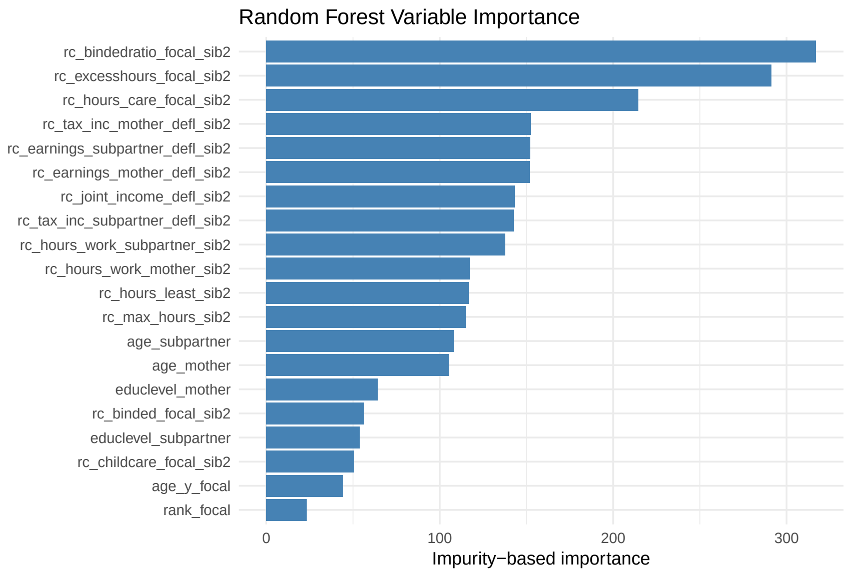
## D.4 Model Performance

The random forest achieves the highest classification performance, with an AUC of 0.829, followed by gradient boosting (0.807) and LASSO (0.776). As a complementary measure of model performance, we also examine Precision–Recall curves. Precision captures the share of predicted treated observations that are truly treated, while recall measures the share of truly treated observations that are correctly identified. Because precision–recall curves are sensitive to class prevalence (i.e., precision is grossly overestimated when applied on a balanced sample), we constructed a small, unbalanced hold-out sample solely for visualization of the precision-recall curve. All final models and predictions reported in the main analysis are estimated using the full sample, including all treated units.

The precision-recall curve is provided in Figure A.10. For comparison, we also included the precision and recall of our baseline binary classification of treatment and control groups based on whether the family was virtually constrained for an older sibling. It turns out that our binary baseline reaches a level of precision of 32% and recall of 26%, reaching the frontier of the precision-recall curve of the LASSO algorithm. Both the random forest and gradient boosting are able to reach higher levels of precision and recall, with random forest predictions again outperforming the gradient boosting algorithm. Based on both metrics, the random forest algorithm performs best, and we use this algorithm to predict the probability of being affected by the hours constraint for the full sample.

Figure D.1 shows the variable importance of our preferred random forest algorithm. It is reassuring that the combination of the continuous variables that determine the intensity of hours constraint show up as most important in driving the random forest prediction.

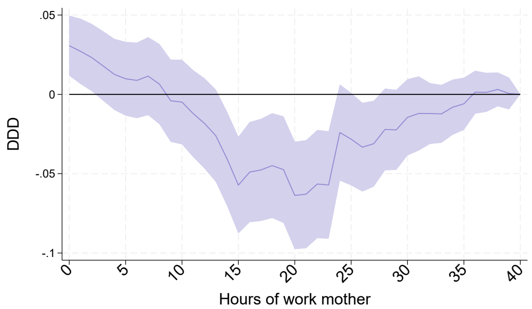
Figure D.1: Variable importance for the random forest algorithm.



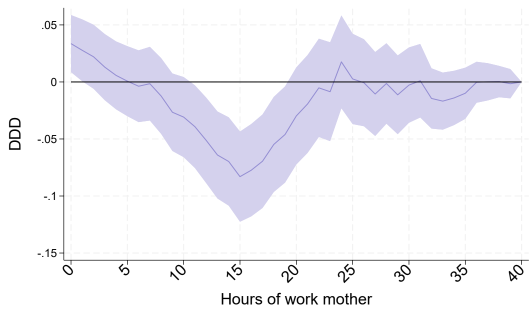

## D.5 Classification of High- and Low-Probability Groups

Because the random forest is trained on a balanced sample, its predicted probabilities are not interpretable in absolute terms and are used solely for ranking individuals by predicted exposure to the reform. To define empirically meaningful groups, we anchor classification thresholds to information on historical incidence: in prior years, approximately 6 percent of individuals were virtually affected by the 2012 reform. We therefore classify individuals in the top 6 percent of the predicted risk distribution as those the model ranks as most likely to be affected, at a rate consistent with historical incidence. Individuals in the bottom 60 percent of the distribution are classified as unlikely to be affected, while observations in the intermediate range are excluded. This choice reflects a precision–contamination tradeoff: restricting the control group to the lower tail of the risk distribution reduces misclassification from latent exposure, while retaining a sufficiently large sample for inference. Reassuringly, including all remaining (i.e., those below the 94th percentile) as controls, or using those below the 75th or 50th percentile as controls does not change the estimates (see Figures D.2 and D.3).

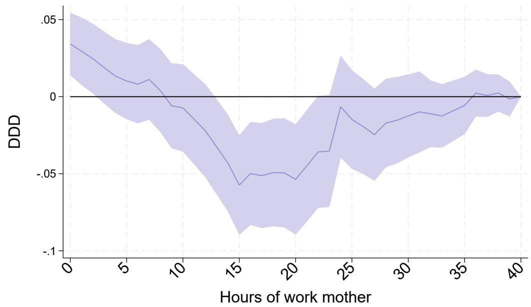
Figure D.2: Machine learning results for maternal hours of work based on alternative definitions of the control group.




(a)



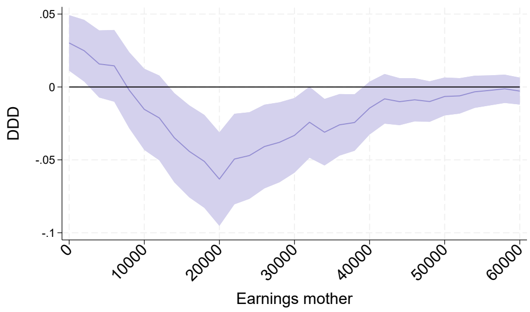

(b)



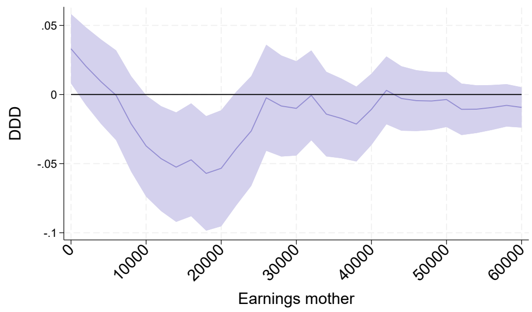

(c)

Notes: Triple difference distribution regressions on hours of work for the mother. Our baseline estimation uses the top 6 percentiles as the treatment group and the bottom 60 percentiles as control. Here we test robustness against using all individuals in the control (i.e., bottom 94 percentiles, panel a), using the bottom 50 percentiles (panel b) and using the bottom 75 percentiles (panel c).

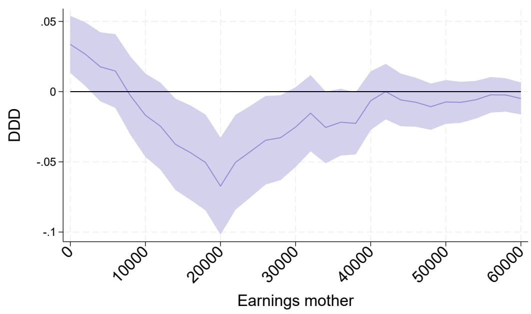
Figure D.3: Machine learning results for maternal earnings based on alternative definitions of the control group.




(a)




(b)




(c)

Notes: Triple difference distribution regressions on earnings for the mother. Our baseline estimation uses the top 6 percentiles as the treatment group and the bottom 60 percentiles as control. Here we test robustness against using all individuals in the control (i.e., bottom 94 percentiles, panel a), using the bottom 50 percentiles (panel b) and using the bottom 75 percentiles (panel c).

# E Additional Robustness Checks

## E.1 Endogenous selection

**Fertility decisions** Our proxy treatment group is defined as families with an older sibling virtually constrained by the hours constraint pre-reform and having a younger sibling aged 1-3 in the years post-reform. Moreover, in the DDD estimation we rely on families *without* a younger child as a control group. This strategy may lead to biased estimates if there is endogenous selection into proxy treatment and control groups, which could be the case if the reform affected fertility decisions.

To test potential fertility responses to the reform, we define the dependent variable as having another child in a given year, and estimate a DD regression where the proxy treatment group constitutes families with a sibling virtually affected by the reform. Table E.1 shows the results of the DD regression without and with covariates. Both estimates suggest a very precisely estimated zero effect on fertility. This is in line with recent evidence from the U.S., a context with much higher child care costs, which shows that child care subsidies (specifically, the Child and Dependent Care Tax Credit or CDCTC) do not affect fertility decisions (Averett and Wang, 2023).

Table E.1: Fertility responses

| | Without covariates | With covariates |
|---|---|---|
| DD | $-0.002$ | $-0.001$ |
| | (0.002) | (0.002) |
| $N$ | 1,295,612 | 1,295,612 |

*Notes*: DD estimates based on the combined YC and NYC samples. Robust standard errors, clustered at the household level, are reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

**Self-employment** Another source of endogenous selection could arise if mothers decide to switch to self-employment in response to the reform. This would imply that we no longer observe their working hours and earnings, and this could bias our intensive margin estimates. Table E.2 however shows no sign of any response on maternal self-employment in response to the reform. Taken together, endogenous selection into our proxy treatment and control groups based on fertility or self-employment changes does not bias our estimates.

## E.2 Subsidy rate and other tax reforms

As discussed in Section B.1, apart from imposing the hours requirement in 2012, there were other concurrent reforms in child care subsidies. In addition, there were also concurrent reforms in the tax system affecting parents with young children. In particular, the Dutch

Table E.2: Effects on self-employment

| | Basic | Saturated | PDS |
|---|---|---|---|
| Self-employment mother | −0.013 | −0.015 | −0.015 |
| | (0.009) | (0.011) | (0.011) |
| $N$ | 602,035 | 602,035 | 602,035 |

*Notes*: As in Table 2. * p<0.1, ** p<0.05, *** p<0.01.

variant of the Earned Income Tax Credit (EITC, Dutch acronym IACK) as well as the labor tax credit (Dutch "arbeidskorting") were subject to changes over the years.[3] To check whether these concurrent changes in subsidy rates and tax credits drive our results, we adopt an approach similar to Gruber and Saez (2002) and Dahl and Lochner (2012) in simulating the predicted (changes in) subsidy rates and tax credits while holding the joint income of parents constant at the level when the older sibling was aged 2 whilst only adjusting for inflation. We then add the levels and changes of predicted subsidy rates and subsidy amounts, as well as the levels and changes of tax credits, to our DDD specifications to see whether the results change. Table E.3 shows the results, which are very similar to our baseline results. Figure E.1 also shows very similar patterns to our baseline results, although the precision is slightly lower. Overall, these results suggest that concurrent reforms in subsidy rates and tax credits are not contaminating our effects of the 2012 reform constraining the hours of child care.

Table E.3: Effects on maternal working with controls for subsidy rate changes

| | Basic | Saturated | PDS |
|---|---|---|---|
| Mother working | −0.029*** | −0.032*** | −0.031*** |
| | (0.010) | (0.013) | (0.013) |
| $N$ | 602,035 | 602,035 | 602,035 |

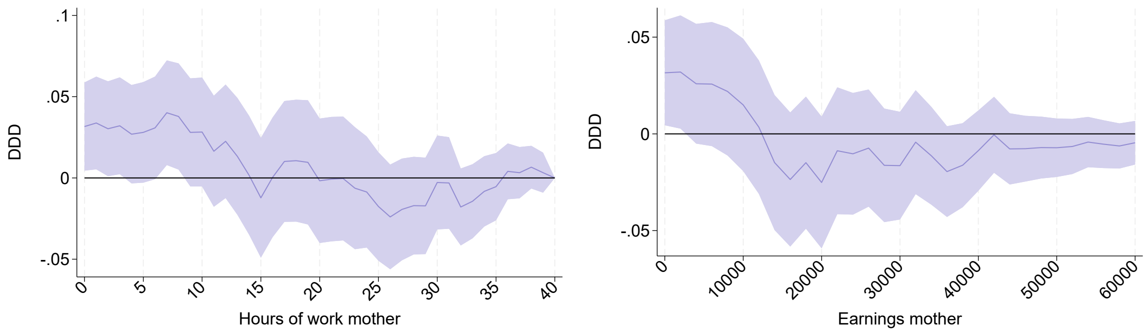
*Notes*: DDD regressions as in Table 2, but with additional control variables that capture the level of the predicted subsidy rate and predicted tax credits, as well as the change in the subsidy rates and tax credits between the younger sibling in the respective calendar year and the older sibling when aged 2. * p<0.1, ** p<0.05, *** p<0.01.

## E.3 Doubly robust estimator

Ortiz-Villavicencio and Sant'Anna (2025) recently showed that the DDD estimator including covariates can no longer be interpreted as the difference between two DD estimators. They suggest a doubly-robust weighted DDD estimator that flexibly allows for covariates and still recovers the correct treatment effect. Their framework is general and allows for staggered adoption, yet in our case treated families are all first treated in 2012 (i.e., the

[3] De Boer et al. (2018) provide an overview of the changes in the tax and family subsidy system in the Netherlands. Between 2005 and 2017, there were changes in the tax system mostly benefiting dual earner couples and thus incentivising both partners to work.

Figure E.1: DDD estimates on maternal work hours (left-panel) and maternal earnings (right-panel) - including controls for concurrent reforms




*Notes*: DDD estimates of the ITT effect of the 2012 reform using distribution regressions on the cumulative distribution function (CDF) on maternal hours of work (left panel) and maternal earnings (right panel). The presented coefficients $\rho$ and associated confidence intervals stem from equation (4) where the outcome $Y$ is replaced by $p(Y \leq y)$ with bin size 1 for maternal weekly work hours, and bin size €1,000 for maternal earnings. Additional control variables are included that capture the level of the predicted subsidy rate and predicted tax credits, as well as the change in the subsidy rates and tax credits between the younger sibling in the respective calendar year and the older sibling when aged 2.

year of the reform). The parameter of interest is

$$ITT(2012,t) = E[Y_{i,t}(2012) - Y_{i,t}(\infty)|G_i = 1, Z_i = 1] \tag{E.1}$$

which is the intention-to-treat effect for the group with a younger child ($G_i = 1$) who are virtually constrained ($Z_i = 1$), comparing their actual outcomes $Y_{i,t}(2012)$ to the counterfactual of what would have happened had they not been treated $Y_{i,t}(\infty)$.

Under the parallel trends in relative changes assumption, and relative to the pre-reform year 2011, the estimator can be obtained as

$$\begin{aligned}\widehat{ITT}(2012,t) = &\Big(E_n[Y_{i,t} - Y_{i,2011}|G_i = 1, Z_i = 1] - E_n[Y_{i,t} - Y_{i,2011}|G_i = 1, Z_i = 0]\Big) \\ &- \Big(E_n[Y_{i,t} - Y_{i,2011}|G_i = 0, Z_i = 1] - E_n[Y_{i,t} - Y_{i,2011}|G_i = 0, Z_i = 0]\Big)\end{aligned} \tag{E.2}$$

where $E_n[A] = n^{-1}\sum_{i=1}^{n} A_i$ denotes the sample mean.

When incorporating covariates, the doubly robust (DR) DDD estimator is derived as

$$\begin{aligned}\widehat{ITT}(2012,t)_{dr} = &E_n\Big[\Big(\hat{w}_{trt}^{G_i=1,Z_i=1} - \hat{w}_{comp}^{G_i=1,Z_i=0}\Big)\Big(\bar{Y}_t - \bar{Y}_{2011} - \hat{m}_t^{G_i=1,Z_i=0}(X)\Big)\Big] \\ &+ E_n\Big[\Big(\hat{w}_{trt}^{G_i=1,Z_i=1} - \hat{w}_{comp}^{G_i=0,Z_i=1}\Big)\Big(\bar{Y}_t - \bar{Y}_{2011} - \hat{m}_t^{G_i=0,Z_i=1}(X)\Big)\Big] \\ &+ E_n\Big[\Big(\hat{w}_{trt}^{G_i=1,Z_i=1} - \hat{w}_{comp}^{G_i=0,Z_i=0}\Big)\Big(\bar{Y}_t - \bar{Y}_{2011} - \hat{m}_t^{G_i=0,Z_i=0}(X)\Big)\Big]\end{aligned} \tag{E.3}$$

where

$$\hat{w}_{trt}^{G_i=1,Z_i=1} = \frac{1\left(G_i=1,Z_i=1\right)}{E_n\Big[1(G_i=1,Z_i=1\Big]} \quad \text{and} \quad \hat{w}_{comp}^{G_i=g,Z_i=z} = \frac{\frac{1(G_i=g,Z_i=z)\cdot\hat{p}^{G_i=1,Z_i=1}(X)}{\hat{p}^{G_i=g,Z_i=z}(X)}}{E_n\Big[\frac{1(G_i=g,Z_i=z)\cdot\hat{p}^{G_i=1,Z_i=1}(X)}{\hat{p}^{G_i=g,Z_i=z}(X)}\Big]}, \tag{E.4}$$

in which $\hat{p}^{G_i=g,Z_i=z}(X) = P\Big[G_i = g, Z_i = z|X\Big]$ is the generalized propensity score, and $\hat{m}_t^{G_i=g,Z_i=z}$ is the predicted value from an regression of outcome $Y_{i,t}$ on the covariates and a post-2012 dummy for the subgroup $G_i = g$ and $Z_i = z$.

In our preferred specification we control for hours of work and hours of child care when the focal child was aged 2. These are potentially important controls as the labor market dynamics may well differ across families with different baseline levels of work hours and child care hours. However, these variables are interfering with the "strong overlap" assumption in Ortiz-Villavicencio and Sant'Anna (2025). This assumption basically states that one cannot perfectly predict group membership (i.e., groups stratified by whether they are virtually constrained and having a younger child) based on the covariates. Since virtually constrained families *are* defined by the combination of work hours and child care hours, the generalized propensity scores reach near-perfect prediction for certain families. This leads to some convergence problems and inflates standard errors for the resulting doubly-robust estimator.

Hence, our preferred estimator is the regression adjustment (RA) DDD estimand in Ortiz-Villavicencio and Sant'Anna (2025), which is given by

$$\widehat{ITT}(2012,t)_{ra} = E_n\Big[\hat{w}_{trt}^{G_i=1,Z_i=1}\Big(Y_{i,t} - \hat{m}_t^{G_i=1,Z_i=0}(X) - \hat{m}_t^{G_i=0,Z_i=1}(X) + \hat{m}_t^{G_i=0,Z_i=0}(X)\Big)\Big]. \tag{E.5}$$

We pool years post 2012 to obtain the average DDD estimates post-treatment, and compute standard errors based on 500 bootstrap replications. The resulting estimates are provided in Table E.4.

Table E.4: DDD estimates including covariates

| | Mother working | | |
|---|---|---|---|
| | $ITT(2012,t)$ | $ITT(2012,t)_{dr}$ | $ITT(2012,t)_{ra}$ |
| DDD | $-0.031^{***}$ | $-0.024$ | $-0.025^{***}$ |
| | (0.010) | (0.065) | (0.006) |
| $N$ | 463,268 | 463,268 | 463,268 |

*Notes*: DDD estimates based on estimators in Ortiz-Villavicencio and Sant'Anna (2025). "dr" stands for doubly robust, "ra" for regression adjustment. Standard errors obtained using 500 bootstrap replications. * p<0.1, ** p<0.05, *** p<0.01.

The standard DDD estimator without covariates is very similar to our baseline estimates in Section 5. When implementing the doubly robust estimator with hours of work and childcare as controls (column 2), the strong overlap assumption is mildly violated, leading to inflated standard errors even though the point estimates remain close to the baseline. Our preferred regression-adjustment estimator (column 3) yields a point estimate of $-0.025$, which is statistically significant and comparable in magnitude. Overall, in the RA–DDD specification contrasting 2011 with the post-reform years, including covariates slightly attenuates the estimate but leaves it similar in magnitude and statistically significant.

## E.4 Other sensitivity analyses

Table E.5 subjects our analysis of maternal employment at the extensive margin to a host of specification checks. Column 1 repeats our preferred estimate from the DDD specification with controls selected by post-double selection. In our main specification, we made a number of choices in specifying the sample or model specification. Here we test sensitivity to these assumptions. Figures E.2 and E.3 show the corresponding effects on hours of work and earnings at the intensive margin, respectively.

Table E.5: Robustness checks

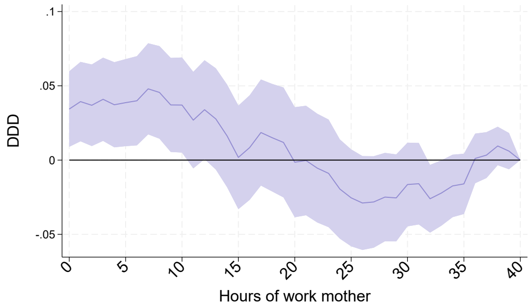
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Mother working | $-0.034^{**}$ | $-0.032^{***}$ | $-0.039^{**}$ | $-0.033^{**}$ | $-0.033^{**}$ | $-0.024^{**}$ | $-0.034^{**}$ |
| | (0.013) | (0.012) | (0.016) | (0.013) | (0.013) | (0.012) | (0.016) |
| $N$ | 602,035 | 674,632 | 421,394 | 602,035 | 602,035 | 651,186 | 473,515 |

*Notes*: DDD Estimates with control variables selected based on post-double selection (PDS). (1) refers to our baseline estimates; (2) including families with joint income above €90k; (3) excluding families with joint income below €60k; (4) includes age in quarters of the focal child; (5) allows for other two-way interactions between the various control variables selected by post-double selection; (6) includes all eligible families, not just those with an exact 3 year age difference between the two siblings; (7) excludes the year 2012. Robust standard errors, clustered at the household level, are reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

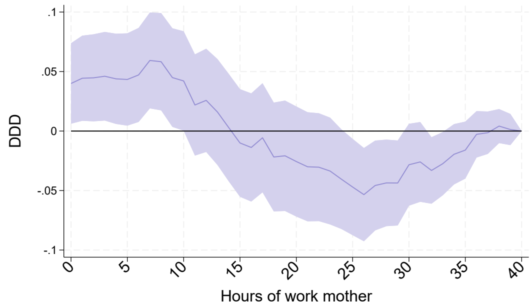
**Income threshold** Since higher-income families were most affected by the subsidy rate reforms, our main sample dropped all families with a joint income above €90,000 to avoid contamination. In column 2 of Table E.5, we show that including the higher income families leads to very similar estimates. Using a tighter restriction of dropping everyone with an income above €60,000 – making it very unlikely that these families are affected by the subsidy rate reform – again yields very similar results (column 3). Figures E.2 and E.3 show the corresponding distribution regression results for hours of work and earnings, respectively, again with very similar findings.

**Specification of control variables** In our baseline specification we choose a set of control variables, including age in years of the focal child (older sibling). To allow for more subtle differences in child care use across ages, in column 4 we include age in quarters with very similar results. In column 5 we extend our set of controls by allowing for an

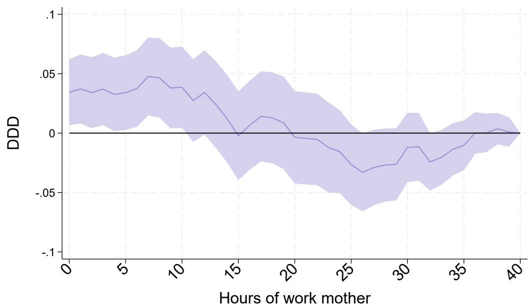
Figure E.2: *Robustness checks for mother's working hours.*



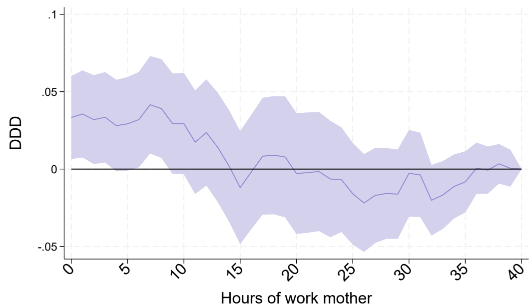

(1)



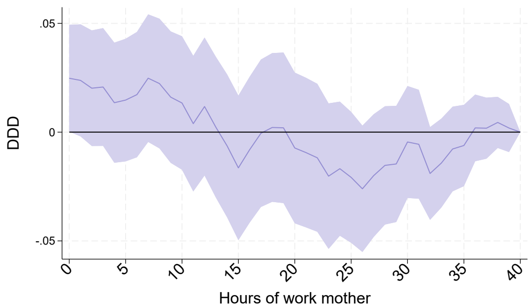

(2)



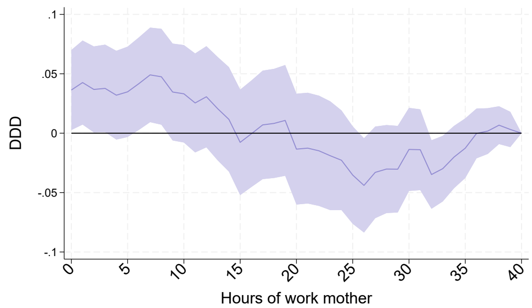

(3)




(4)



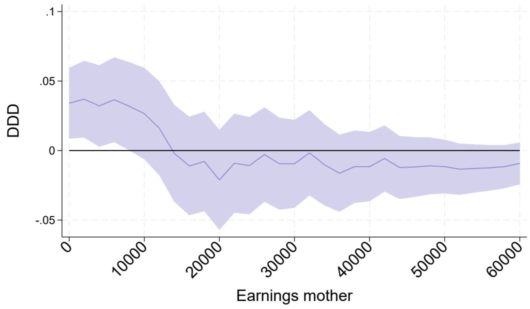

(5)



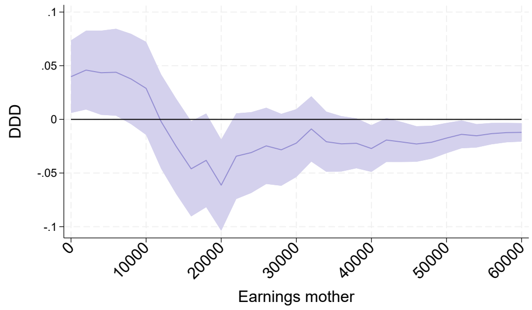

(6)

Notes: Labels as in Table E.5: (1) including families with joint income above €90k; (2) excluding families with joint income below €60k; (3) includes age in quarters of the focal child; (4) allows for other two-way interactions between the various control variables selected by post-double selection; (5) includes all eligible families, not just those with an exact 3 year age difference between the two siblings; (6) excludes the year 2012. Robust standard errors, clustered at the household level, are reported in parentheses.

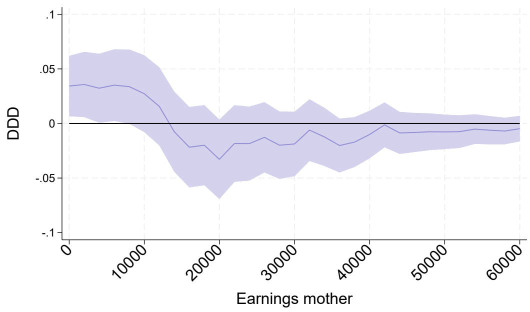
Figure E.3: *Robustness checks for mother's earnings.*



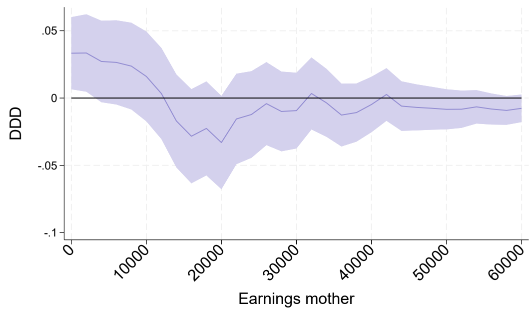

(1)



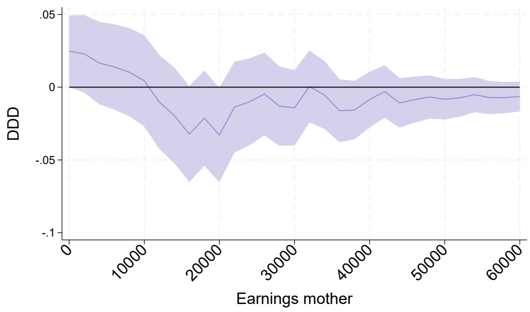

(2)



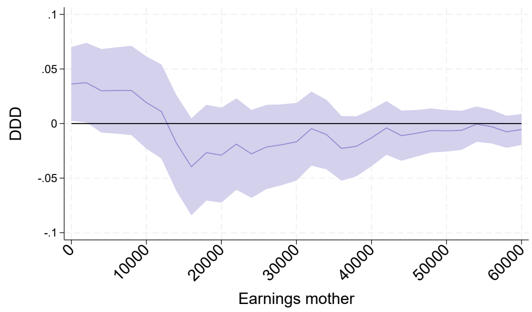

(3)




(4)




(5)




(6)

Notes: Labels as in Table E.5: (1) including families with joint income above €90k; (2) excluding families with joint income below €60k; (3) includes age in quarters of the focal child; (4) allows for other two-way interactions between the various control variables selected by post-double selection; (5) includes all eligible families, not just those with an exact 3 year age difference between the two siblings; (6) excludes the year 2012. Robust standard errors, clustered at the household level, are reported in parentheses.

additional set of two-way interactions across all control variables selected by post-double selection (Belloni et al., 2014). The results hardly change.

**Sample restrictions** In our main sample, we restricted to families where the age gap between the older and younger sibling was exactly 3 years. When we lift this restriction in column 6, the point estimates for mother working are somewhat smaller compared to our main estimates, but none of our conclusions is affected. In column 7 we exclude the year 2012 since some individuals may have learned about the new rules only in or after 2012.[4] The results remain very similar.

# F Partial Identification of the CATT Parameter

In this section, we explore partial identification of the treatment effect without imposing Assumption 5 (*Homogeneity in the CATT between subgroups*). We show that even when the two CATT parameters $\Delta_1$ and $\Delta_0$ differ, the Wald-DDD estimator $\widehat{W}_{DDD}$ can still yield informative bounds on $\Delta_1$ provided that the two parameters share the same sign (i.e., $\Delta_1\Delta_0 \geq 0$). This result is formalized in Corollary 1:

**Corollary 1** *If Assumptions 1-4 hold and the two CATT parameters $\Delta_1$ and $\Delta_0$ have weakly the same sign (i.e., $\Delta_1\Delta_0 \geq 0$), then:*

(i) *If $\Delta_1, \Delta_0 \geq 0$, $\frac{1}{1+\widehat{\alpha}}\widehat{W}_{DDD}$ yields a lower bound on $\Delta_1$, i.e., $\Delta_1 \geq E\big[\frac{1}{1+\widehat{\alpha}}\widehat{W}_{DDD}\big]$.*

(ii) *If $\Delta_1, \Delta_0 \leq 0$, $\frac{1}{1+\widehat{\alpha}}\widehat{W}_{DDD}$ yields an upper bound on $\Delta_1$, i.e, $\Delta_1 \leq E\big[\frac{1}{1+\widehat{\alpha}}\widehat{W}_{DDD}\big]$, which implies $|\Delta_1| \geq |\widehat{W}_{DDD}|$. That is, the magnitude of $\frac{1}{1+\widehat{\alpha}}\widehat{W}_{DDD}$ yields a lower bound for the absolute size of the adverse CATT parameter $\Delta_1$.*

Given Corollary 1, whenever the two CATT parameters share the same sign — whether both positive or both negative — $\frac{1}{1+\widehat{\alpha}}\widehat{W}_{DDD}$ always yields an informative lower bound on the magnitude of $\Delta_1$.

# G Additional Details for the GMM Estimator

Combining Assumption 3 (*Stationarity*) and the first-stage relationship in Equation (7) implies that $E[D_i|Z_i, G_i = 1, T_i \geq 2012] = E[D_i|Z_i, G_i = 1, T_i < 2012] = \alpha + \beta Z_i$.

[4] Implementation of the reform was done on the 1st of January, 2012 and the reform was announced roughly 6 months before the introduction in June 2011 (Parliament, 2011). However, since the rules were new and child care subsidies are deposited based on parental expectations of working hours, it could be that some families only learned about the new rules during or at the end of 2012. Figure A.2 shows that we already see some reductions in labour supply starting in the early months of 2012 among affected families, suggesting that most families were aware of the new rules.

Substituting this into Equation (8), we can express the conditional expectation of $Y_i$ as a linear projection on $Z_i$, $T_i$, $G_i$ dummies and their interactions as follows:

$$\begin{aligned} E[Y_i|Z_i, T_i, G_i] =& \delta_t \mathbf{1}(T_i = t) + \kappa_t G_i \mathbf{1}(T_i = t) + \theta_t Z_i \mathbf{1}(T_i = t) + \tau Z_i G_i + \\ & \lambda\alpha G_i \mathbf{1}(T_i \geq 2012) + \lambda\beta Z_i G_i \mathbf{1}(T_i \geq 2012). \end{aligned} \tag{G.1}$$

The moment conditions corresponding to the first-stage and reduced-form Equations (7) and (G.1) are:

$$E\Big[\big(1 \quad Z_i\big)'\big(G_i \mathbf{1}(T_i < 2012)(D_i - \alpha - \beta Z_i)\big)\Big] = 0 \tag{G.2a}$$

$$E\left[\begin{array}{c} \Big(\big\{\mathbf{1}(T_i = t) \quad G_i\mathbf{1}(T_i = t) \quad Z_i\mathbf{1}(T_i = t)\big\}_{t\in\{2009,2014\}} \quad Z_iG_i \quad Z_iG_i\mathbf{1}(T_i \geq 2012)\Big)' \\ \Big(Y_i - \delta_t - \kappa_t G_i - \theta_t Z_i - \tau Z_i G_i - \lambda\alpha Z_i G_i \mathbf{1}(T_i \geq 2012) - \lambda\beta Z_i G_i \mathbf{1}(T_i \geq 2012)\Big) \end{array}\right] = 0. \tag{G.2b}$$

Let $\Lambda \equiv \big(\lambda, \alpha, \beta, \tau, \{\delta_t, \kappa_t, \theta_t\}_t \in \{2009, ..., 2014\}\big)$ denote the vector of population parameters.[5] We construct the optimal GMM estimator by weighting these moment conditions by the inverse of their variance:

$$\widehat{\Lambda} \equiv \arg\min_{\Lambda} \Big[\frac{1}{N}\sum_{i=1}^{N} g_i(\Lambda)\Big]' \hat{V}^{-1} \Big[\frac{1}{N}\sum_{i=1}^{N} g_i(\Lambda)\Big],$$

where

$$g_i(\lambda) \equiv \left[\begin{array}{c} \big(1 \quad Z_i\big)'\big(G_i \mathbf{1}(T_i < 2012)(D_i - \alpha - \beta Z_i)\big) \\ \Big(\big\{\mathbf{1}(T_i = t) \quad G_i\mathbf{1}(T_i = t) \quad Z_i\mathbf{1}(T_i = t)\big\}_{t\in\{2009,2014\}} \quad Z_iG_i \quad Z_iG_i\mathbf{1}(T_i \geq 2012)\Big)' \\ \Big(Y_i - \delta_t - \kappa_t G_i - \theta_t Z_i - \tau Z_i G_i - \lambda\alpha Z_i G_i \mathbf{1}(T_i \geq 2012) - \lambda\beta Z_i G_i \mathbf{1}(T_i \geq 2012)\Big) \end{array}\right]$$

and $\hat{V}$ is an estimate of the variance of $g_i(\Lambda)$.

**Incorporating covariates** To simplify algebra, so far we have focused on the basic empirical specification without covariates. The GMM framework in Equation (8) can be easily extended to add covariates analogous to Equation (4) as follows:

[5] With two groups and six periods, $\Lambda$ contains 22 parameters, and Equations (G.2a)–(G.2b) jointly provide 22 moment conditions.

$$
\begin{aligned}
Y_i = \sum_{t=2009}^{2014} \delta_t \times \mathbf{1}(T_i = t) + \sum_{t=2009}^{2014} \kappa_t \big(G_i \times \mathbf{1}(T_i = t)\big) + \sum_{t=2009}^{2014} \theta_t \big(Z_i \times \mathbf{1}(T_i = t)\big) + \tau(Z_i \times G_i) \\
+ \lambda \big(G_i \times D_i \times \mathbf{1}(T_i \geq 2012)\big) + f\big(X_i; G_i, Z_i, T_i\big) + \varepsilon_i,
\end{aligned}
\tag{G.3}
$$

where $f\big(X_i; Z_i, G_i, t\big)$ denotes the main effects of covariates $X_i$, as well as their two-way interactions with calendar year, group, and proxy treatment assignment dummies, chosen by the post-double selection procedure. Incorporating these covariates allows us to relax Assumption 4 (*Parallel Trends in Relative Changes*) to its conditional form, while also resulting in more precise estimates.