What Causes the Child Penalty? Evidence from Same Sex Couples and Policy Reforms*

Martin Eckhoff Andresen† Emily Nix‡

Abstract

Women experience significant reductions in labor market income following the birth of children, while their male partners experience no such income drops. This "relative child penalty" has been well documented and accounts for a significant amount of the gender income gap. In this paper we do two things. First, we use a simple household model to better understand the potential mechanisms driving the child penalty, which include gender norms around child care, female preferences for child care, efficient specialization within households, and the biological cost of giving birth. The model, combined with the estimated child penalties for heterosexual and same sex couples, suggests that the child penalty experienced by women in heterosexual couples is primarily explained by female preferences for child care and gender norms, with a smaller contribution due to the biological costs of giving birth. Second, we provide causal estimates on the impact of two family policies aimed at reducing the relative child penalty: paternity leave and subsidized early child care. Our precise and robust regression discontinuity results show no significant impact of paternity leave use on the relative child penalty. Early subsidized care seems to have more promise as a policy tool for affecting child penalties, as we find a 25% reduction in child penalties per year of child care use from a large Norwegian reform that expanded access to child care.

JEL-codes: I21, J13, J22, J71

Keywords: Gender wage gap, labor supply, child penalty, paternity leave, child care, same sex couples, event study, regression discontinuity, instrumental variables

*We thank seminar participants at the University of Rochester, Claremont McKenna University, Statistics Norway, RAND, Arizona State University, LSU, the University of Oslo, VATT Helsinki, Warwick University and Erasmus University. We also thank Kenneth Aarskaug Wiik, Edwin Leuven, Matias D. Cattaneo, Sebastian Calonico, Heather Antecol, Adam Sheridan, Trude Gunnes, Petra Persson, Antonio Dalla Zuanna, Thor Olav Thoresen and Nina Drange for helpful comments and suggestions. All errors remain our own. Andresen gratefully acknowledges financial support from the Norwegian Research Council (grant no. 236947). This version: March 25, 2019. Latest version here.

†Statistics Norway, mrt@ssb.no

‡Corresponding Author: University of Southern California, enix@usc.edu, 701 Exposition Blvd Ste. 231, Los Angeles, CA 90089-1422

1 Introduction

The gender income gap has narrowed significantly over the past 50 years.[1] However, one component of the gender income gap has proven to be relatively persistent: the income penalty women in heterosexual couples experience after the birth of children. In contrast, men in heterosexual couples experience no such income penalty upon the birth of children. This income penalty experienced by women is often termed the "child penalty"[2] and its importance has recently been documented in a variety of countries such as the United States, Denmark, Norway, the United Kingdom and Sweden (see Chung *et al.* (2017), Kleven *et al.* (2018), Bergsvik *et al.* (2019), Kuziemko *et al.* (2018) and Angelov *et al.* (2016)). As other determinants of the gender income gap have declined in importance, the proportion of the gap that can be explained by the "relative child penalty", the difference in the child penalty experienced by fathers compared to mothers, has increased. Kleven *et al.* (2018) show that in Denmark the relative child penalty accounted for 80% of the gender gap in 2013, compared to only 40% in 1980.[3]

The stubborn persistence of the relative child penalty among heterosexual couples is a puzzle, particularly given the overall decline in gender wage gaps. In this paper we attempt to understand the relative child penalty and how it might be reduced. First, we try to understand why the relative child penalty exists by using a household model and comparing the child penalties of heterosexual and same sex couples, estimated in an event study framework, which the model predicts should behave differently depending on what causes the child penalty. Second, we present causal evidence on the impact of two common policies proposed to reduce the relative child penalty: paternity leave and use of subsidized formal childcare.

In the first half of the paper, we consider commonly suggested mechanisms behind the child

[1] See Blau and Kahn (2000). Additionally, economists have provided evidence on a number of explanations for this decline, such as the narrowing of the gender education gap, the decrease in labor force discrimination, and family oriented policies. For an overview, see Olivetti and Petrongolo (2016).

[2] We acknowledge that some readers may object to using the term "penalty" if this phenomenon is not driven by discrimination. This paper aims at disentangling the mechanisms behind these disparate income penalties, but we will use the term "child penalty" for the income loss following child birth independently of the mechanism, in line with the literature.

[3] Of course, other determinants of the remaining gender gap are also important, and may interact with the impact of children. For example, Goldin (2014) focuses on the structure of the labor market as an explanation for the remaining gender gap.

penalty: gender norms, female preferences for child care, efficient within household specialization, and biology. To understand which of these mechanisms drives the relative child penalty, we estimate and compare the child penalties among same sex male and same sex female partners to the child penalties experienced by heterosexual couples using administrative data from Norway. Our approach is motivated by suggestive evidence that same sex couples split household chores more evenly (Goldberg *et al.*, 2012). If the absence of pre-set gender roles lead same sex couples to also split the burden of child care more evenly, child penalties may look very different among same sex couples. To identify the child penalties within each couple type, we use an event study approach as in Kleven *et al.* (2018).

To more formally understand how our results can disentangle the roles of preferences, the biological costs of giving birth, household specialization and gender norms around child care in the heterosexual relative child penalty, we build a simple model of household labor supply before and after the arrival of children. In the model, partners may differ in their relative productivity in the labor market versus home production, men and women may have different preferences for child care, and pregnancy imposes a fixed cost to the woman physically bearing the child. We model gender norms as a disutility for men in heterosexual couples from women working outside the home after the child is born, as in Fernández *et al.* (2004). The model yields the following intuitive predictions. As expected, and by construction, each of these mechanisms yield a relative child penalty for heterosexual couples. If household specialization drives the relative child penalty within heterosexual couples, the model predicts similar child penalty patterns in otherwise similar same sex couples. If part of the relative child penalty is driven by the costs of giving birth, the model predicts a relative child penalty for the pregnant mother versus the non-pregnant mother among same sex female couples, but no such difference among same sex male partners. If intra-couple gender norms cause the relative child penalty in heterosexual couples, the model predicts that we will not find relative child penalties among same sex couples. If women have greater preferences for child care than men, the model predicts child penalties for both partners in same sex female couples and smaller or no penalties for partners in same sex male couples. However, if child penalties are driven by preferences for child care, the model also

predicts that child penalties for lesbian mothers will be smaller than for heterosexual women. This result is driven by the fact that heterosexual women can lean on their male partners, who derive less utility from time with children, to make up for the time they spend in home versus market production.

Similar to previous papers, we find that women in heterosexual couples experience an average drop in income of approximately 22% following the birth of the first child, and this drop persists over time. Their male partners experience no child penalty in income. We also show that this large drop in female income translates to an overall household income drop of 6-8% for heterosexual households, and this household income penalty also persists over time. For female same sex couples we find an initial 13% drop in the income of the partner who gives birth. Her partner experiences an initial income drop of 5%. Despite a larger immediate drop in income, the mother who gives birth catches up with her partner around two years after birth, and from that point on both mothers experience similarly sized child penalties which decrease over time; by four years after birth there is no longer a child penalty. While the initial household income penalty experienced by lesbian couples on the birth of the first child is statistically indistinguishable from the same income penalty experienced by heterosexual couples (although shared more evenly between partners), by five years after birth lesbian couples no longer experience a household income penalty. Since the model predictions regarding specialization require comparisons of child penalties across couples with similar comparative advantage differentials, we expand on the traditional child penalty event study by introducing two approaches motivated by the household model to control for comparative advantage differences across couple types. The differences between heterosexual and lesbian couples remain in these specifications.

These patterns suggest that while biology may play a small role, the majority of the relative child penalty experienced by heterosexual couples is due to preferences and gender norms. While the population of same sex male couples with children is very small, we find no income penalty for either spouse. This is also consistent with a dominant gender norms and female preferences mechanism, and a smaller role played by biology. Last, we investigate the possibility that all of the differences are in fact driven by same sex couples caring less about their children's out-

comes. While this assumption would be consistent with our results, it is not consistent with one additional result: children of same sex couples outperform children of heterosexual couples on English, reading and math tests at age 10, even after conditioning on a large range of observable differences between the couple types.

To further understand the anatomy of the child penalties and how they differ between couple types, we next decompose the overall income penalty into a series of potential decisions made by couples after birth which all may impact income: total contracted hours, binary indicators of employment at various levels, family friendliness of the employer, and sickness absence. Results indicate that the differences between lesbian and heterosexual couples are primarily driven by different responses at the intensive margin of labor supply, not at the extensive margin, nor through differences in occupational sorting.

We next turn to investigate the impact policy might have on the relative child penalty. Policy makers might wish to know how to decrease the relative child penalty in order to reduce the overall gender income gap, particularly given the results from the first half of the paper. In the second half of this paper, we estimate the impact of two commonly proposed family policies aimed at reducing the relative child penalty: paternity leave and subsidized early child care. Paternity leave may reduce the relative child penalty by targeting fathers while subsidized access to high quality child care may reduce the relative child penalty by providing households with a viable substitute for the mother's time at home.

For paternity leave, we use a regression discontinuity design to estimate the impact of six reforms to the paid paternity leave quota in Norway from 2005-2014. Using robust semi-parametric RD methods we estimate a strong first stage: the reforms significantly increased paternity leave takeup. However, despite fathers taking additional leave, we find no significant impact on either spouse's labor income. Consistent with the lack of impact on individual incomes, there is no impact of paternity leave on the relative child penalty. Pooling all reforms, we can rule out reductions in mothers' earnings from an extra week of paternity leave larger than around 5 to 7 per cent of the child penalty.

Paternity leave use may, however, impact the relative distribution of home and market work

between the two spouses in a way that does not necessarily show up in earnings. To see if this is the case, we use the same paternity leave reforms to estimate the impact of leave use for the first child on leave use for subsequent children. If paternity leave use affects norms and preferences related to child care, we might expect to see fathers who are induced by the policy to take additional leave for the first child to also increase their use of paternity leave for subsequent children. Again, however, this is not what we find. Instead, our precise and robust RD estimates show no impact of leave use on future take up of leave for any of the reforms, with non-significant estimated effects of less than 0.1 additional week of leave taken by fathers for subsequent children per week of leave use for the first child.

In the final section of the paper we use a large-scale Norwegian reform from 2002 that expanded child care availability for 1-2 year olds to investigate the effect of access to high quality child care on parent's child penalties over time. The market for care for toddlers was severely rationed before this reform. The reform increased subsidies to child care institutions, leading to a rapid expansion of care slots. To identify the impact of increased access to high quality child care, we exploit the variation across municipalities and over time in construction of new slots and centers, instrumenting individual child care use with the rationed, municipality-level availability of slots in a variation of the setup in Andresen and Havnes (2019). Results indicates positive effects on mothers' labor income at ages 2 and 3 that scales to reduce the child penalty experienced by mothers by around 25% for each additional full year of early child care use, although the impacts are not persistent in the long run.

Our paper is most closely related to the literature on child penalties. We use the simple event study approach from Chung *et al.* (2017), Kleven *et al.* (2018), Bergsvik *et al.* (2019), and Angelov *et al.* (2016) to identify child penalties.[4] Together, our results and the results from these papers suggest that there does not currently exist a sample of heterosexual couples, whether in different countries, educational groups, or socioeconomic class, that does not experience large relative child penalties. However, as we show in our household model, it is impossible to understand why these relative child penalties occur by estimating child penalties for heterosexual couples

[4]Lundborg *et al.* (2017) also show the child penalty occurs among heterosexual couples who use IVF to get pregnant, which may be even closer to the process that same sex couples experience when conceiving children.

alone. In this paper, we find very different patterns when estimating the same event study for same sex couples, and use these results combined with predictions from the household model to shed some light on why heterosexual couples experience such large relative child penalties. Related to our results, Kuziemko *et al.* (2018) also find evidence that preferences of heterosexual women may play an important role in the child penalty. Specifically, they show that women in heterosexual couples exhibit time inconsistency in these preferences, finding that women report more negative opinions toward female employment after giving birth relative to before birth. Another closely related paper is Kleven *et al.* (2019) which shows the same general pattern in child penalties across a number of different countries and finds that the magnitude of the child penalties experienced by women are correlated with elicited gender norms.

Our paper also contributes to a smaller literature focused on same sex couples and their children. Baumle (2009) finds that in the United States, partnered gay men on average earn less than partnered heterosexual women, while the opposite is true for partnered lesbian women. Schneebaum (2013) also finds that lesbian women earn more than heterosexual women, but focuses on the differences between primary and secondary earners, as well as those with and without children. Black *et al.* (2007) review existing data, provide additional summary statistics for the United States, and suggests a role for economics in understanding household choices of gay and lesbian couples. Looking more specifically at parenting, Goldberg *et al.* (2012) look at a sample of 55 lesbian couples and find they report sharing household chores and child care more evenly than a comparison group of 65 heterosexual parents. Others have investigated labor supply (Antecol and Steinberger, 2013), parental leave use (Evertsson and Boye, 2018; Rudlende and Lima, 2018) and time use (Martell and Roncolato, 2016) for same sex couples, as well as the impact of legal recognition (Alden *et al.*, 2015). Finally, Moberg (2016) and Rosenbaum (2019) investigates the differential response to child birth across heterosexual and same sex couples in Sweden and Denmark.

While comparing the outcomes of children born to same sex and heterosexual couples is not the focus of this paper, we also present evidence that children of same sex (mostly lesbian) couples have higher math, English, and reading scores at age 10, and the effect is significant at the 99th

percentile for English and reading scores. These results remain significant when controlling for a large range of observable differences between heterosexual and same sex couples. This result contributes to a charged debate in the United States, as demonstrated in oral arguments for the landmark 2015 Supreme Court case *Obergefell v. Hodges,* which legalized same sex marriage. Previous studies of children born to same sex couples have been criticized by both sides of the debate on the basis of three methodological concerns: non-representative samples,[5] mislabeling children from heterosexual couples as children of homosexual couples or vice versa,[6] and small sample size. In this paper, our use of administrative data containing the population of children of same sex couples in Norway and the ability to identify such children accurately largely overcomes these concerns.

The second half of our paper contributes to the literatures on paternity leave and child care, by looking specifically at the impact of these policies on the individual and relative child penalties. A number of papers have estimated the impact of paternity leave policies on different outcomes.[7] A few particularly relevant studies include Cools *et al.* (2015) and Kotsadam and Finseraas (2011; 2013) who find positive impacts of the Norwegian paternity leave policies on child outcomes and their later equality in division of household work looking at the 1993 reform.[8] Dahl *et al.* (2014) find substantial peer effects of the Norwegian policy in 1993 using a regression discontinuity approach, and we use a similar approach to identify the causal effect of exposure to paternity leave on the amount of leave taken for later children. Rege and Solli (2013) find a decrease in father earnings long term in Norway from the 1993 reform using a difference in difference approach and Johansson (2010) finds that a Swedish policy increased mother's earnings but had no impact on fathers. Ekberg *et al.* (2013) find that fathers are no more likely to take sick leave to care for a sick child long term using a Swedish reform, and Patnaik (2019) finds a large and persistent change

[5]Studies often used "opportunity samples" where couples volunteer to participate.

[6]In particular, a number of studies label children born to a heterosexual couple, which later divorces and one spouse enters a same sex relationship, as children of homosexual couples. Under this approach, if these children do worse than children in stable heterosexual couples, it is impossible to disentangle the impact of divorce versus having one set of same sex parents.

[7]A larger literature looks at the impact of maternity leave on maternal earnings and child outcomes. See, for example, Lalive and Zweimüller (2009); Lalive *et al.* (2014); Carneiro *et al.* (2015); Baker and Milligan (2015).

[8]Halrynjo and Kitterød (2016) find small and contradictory effects from quasi-experimental evaluations in a survey of studies on Nordic daddy quotas.

in the division of household labor from a Canadian daddy quota. This selection of papers from a broader literature captures the fact that existing work on paternity leave finds either no impact or positive impacts on children. The literature finds either no impact or a decrease in fathers' income and an increase in mothers' income, pointing at least to the possibility that paternity leave may decrease the relative child penalty. We add to this literature by exploiting six consecutive paternity leave reforms, one of which decreased the quota, using robust semi-parametric regression discontinuity methods. We show that the impact is symmetric across reforms that increased and reduced the quota, and by stacking all six reforms we provide precise zero estimates for the effect on labor income. Furthermore, we show no effect of exposure to paternity leave for the first child on leave use for subsequent kids, suggesting that preferences for leave taking is not affected by exposure to paternity leave.

Our results on paternity leave are also related to Antecol *et al.* (2018) who find that gender neutral tenure clock stopping policies do not help women in academia, and may even hurt their careers. We examine a similar shift toward more gender neutral leave policies, and find that the results from Antecol *et al.* (2018) are not unique to academia. Paternity leave does not help women's careers, at least not in terms of income, across the population of professions in Norway.

Finally, we contribute to the large literature on the impact of child care use on female labor supply. Most closely related is Andresen and Havnes (2019) on which we build. Havnes and Mogstad (2011) find no effects of a similar expansion of care for older kids' outcomes in Norway in the mid 1970's. Other related papers in this field are summarized in e.g. Blau and Currie, 2006; Akgunduz and Plantenga, 2018; Morrissey, 2016. In this paper we focus specifically on the impact of access to child care on the individual and relative child penalties experienced by men and women within heterosexual couples. Although the literature finds mixed evidence, we find positive impacts of child care access on female labor market outcomes, and thus that child care may reduce the relative child penalty experienced by heterosexual women.

The remainder of the paper is organized as follows. In Section 2 we present a model for household labor supply in the presence of children and derive testable predictions. In Section 3 we describe our approach to identify child penalties across couple types. In Section 4 we outline

the institutional background and the data, and in Section 5 we present the main results. In Section 6 we present our empirical strategies and results on the impact of paternity leave and access to child care on the heterosexual child penalties. In Section 7 we conclude.

2 A model of household labor supply in the presence of children

In this section we develop and solve a simple household model. The model includes the most commonly suggested mechanisms for the child penalty: gender norms around child care, specialization within households, female preferences for child care, and the impact of giving birth. The solutions of the model provide testable predictions that we bring to the data. Our model is loosely adapted from similar household models in Fernández *et al.* (2004) and Olivetti (2006). Once solved, the model shows that while each of these mechanisms generate a relative child penalty for heterosexual couples, comparisons with same sex couples will allow us to distinguish between mechanisms when we estimate individual and relative child penalties for heterosexual and same sex couples in the data.

There are three periods. In the first period, households consist of two adults. In the second period, the child arrives in the household (either adopted or birthed by a female adult).[9] In the second and third period, the household consists of the two adults and the child. Each adult is endowed with 1 unit of time in every period. In each period households choose the amount of labor each adult allocates between home and labor market production. The two adults may be of any gender (man and women, two men, or two women). The quasi linear utility function of each spouse $i \in a, b$ is given by:

$$U_i\left(c, \theta, t_{-i}\right) = c + \left(1 + a_t\right) \beta \ln \theta + \eta \ln \left(1 - t_i\right) \bar{X}_i - \alpha t_{-i} \bar{Z}_i \tag{1}$$

where c is consumption and θ is household production. a_t represents the additional utility from household production once the child arrives (so $a_1 = 0$, and $a_2, a_3 > 0$). $\bar{Z}_i$ is an indicator

[9]We do not model the fertility decision or allow parents to make labor market decisions in anticipation of children. While these are important issues (see for example Bursztyn *et al.* (2017)), they are beyond the scope of this paper. We do allow for an income gap before children, which could capture some of these points.

equal to 1 if the individual is a male married to a female in periods 2 and 3, $\bar{X}_i$ is an indicator equal to 1 if the individual is female and 0 if the individual is male, and t_i and t_{-i} are own and spouse's labor supply. β represents the preferences for home production, which are shifted by a_t at the arrival of the child, and η is the additional utility women get from being at home with children, capturing potential differences in gender preferences over time with children. α is the disutility men get from each hour their wife works when they have children, capturing gender norms around child care.[10]

There is no saving or borrowing, and in each period household consumption is joint and equal to the sum of spouses' earnings. For simplicity, we do not model wage setting, and simply take as given the wages of each spouse w_i and w_{-i}, so that

$$c = w_i t_i + w_{-i} \left(1 - \delta_t \bar{S}\right) t_{-i}$$

where $\bar{S}$ is an indicator equal to 1 in period 2 and 3 if the spouse is a woman who gave birth. δ_t is the productivity shock, which we think of as capturing the health shock of giving birth, as well as other biological components such as breast feeding.

Household goods (including child quality) are produced by the following production function

$$\theta = k_i h\left(1 - t_i\right) + k_{-i} h\left(1 - t_{-i}\right) \tag{2}$$

where $k_i \geq 0$ are productivity parameters, $h^{'} > 0$, $h^{''} \leq 0$, and $h\left(0\right) = 0$.

The household maximizes utility by choosing each spouse's division of labor in each period, where household utility is given by

$$\sum_i \lambda_i U_i \left(c, \theta, t_{-i}\right)$$

[10]Survey evidence shows large differences in the norms towards working women with young children compared to working women without children. As an example, 80% of the respondents in the ISSP in 2002 think that married women without children should work full time in the United States, while only around 15% think the same about women with children below school age. Similar differences appear for other countries, including Sweden and Denmark, see International Social Survey Program (ISSP) from 2002. See also Kleven *et al.* (2019).

and λ_i is the weight of each spouse in household decisions. This assumes Pareto efficiency in household decisions and is consistent with a number of household bargaining problems.[11]

There are no dynamics to the problem. This means we can solve the problem sequentially, maximizing t_a and t_b in each period. For each period, the couples solve the following equation, taking the home production process in equation 2 as given:

$$\max_{t_a,t_b} \; (\lambda_a + \lambda_b)\left(w_a t_a + w_b t_b - \delta w_b t_b \bar{S} + \beta \ln \theta\right) + \lambda_a \eta \ln(1-t_a)\,\bar{X}_a + \lambda_b \eta \ln(1-t_b)\,\bar{X}_b - \lambda_a \alpha t_b \bar{Z}_a \tag{3}$$

The first order conditions are:

$$\frac{(1-\delta)\,\bar{S}_i w_i}{k_i} = \frac{(1+a_t)\,\beta h'(1-t_i)}{k_i h(1-t_i) + k_{-i} h(1-t_{-i})} + \frac{\lambda_i \eta \bar{X}_i}{k_i(\lambda_i + \lambda_{-i})(1-t_i)} + \frac{\lambda_{-i}\alpha \bar{Z}_{-i}}{k_i(\lambda_i + \lambda_{-i})}$$

These wage equations yield the following predictions:

1. **Female preferences for child care:** The income penalty is increasing for all women as η increases. The income penalty for heterosexual men is decreasing. However, for any given $\eta > 0$, the increase in the income penalty experienced by lesbian women due to an increase in η is smaller than the increase in the income penalty for heterosexual women. The relative child penalty for heterosexual couples is increasing in η at an increasing rate if $h'' < 0$ and at a constant rate otherwise. The child penalty for lesbian couples is zero if $\delta = \frac{w_a}{k_a} - \frac{w_b}{k_b} = 0$. Otherwise, there is no contribution to any existing relative child penalty for lesbian couples so long as h'' is constant. By construction, η has no impact on the incomes of gay men, and cannot account for a relative child penalty for gay men.

2. **Biology:** The income penalty is increasing for the woman who gives birth as δ increases. The relative child penalty for lesbian and heterosexual couples is increasing in δ at an increasing rate if $h'' < 0$ and at a constant rate otherwise. δ has no impact on the income or relative child penalty of gay men by construction.

[11] This is a very simple model by design. It assumes Pareto efficiency, but this has some important drawbacks. See Del Boca and Flinn (2012) for a discussion of alternative approaches. Notice that we assume that the bargaining weights do not vary by couple type. An alternative approach to capture gender norms could be to assume that in same sex couples $\lambda_a = \lambda_b$ and in heterosexual couples $\lambda_a > \lambda_b$, where λ_a represents the Pareto weight of the man.

3. **Gender norms:** The income penalty for heterosexual women is increasing as α increases and the income penalty for heterosexual men is decreasing. The relative child penalty for heterosexual couples is increasing in α at an increasing rate if $h'' < 0$ and at a constant rate otherwise. By construction, α has no impact on the income and relative child penalties of gay and lesbian women.

4. **Intra-household specialization:** Let spouse a have a comparative advantage in market work, so that $\frac{w_a}{k_a} \geq \frac{w_b}{k_b}$. The income penalty for spouse a is decreasing as $\frac{w_a}{k_a} - \frac{w_b}{k_b}$ increases, while the income penalty for spouse b is increasing as $\frac{w_a}{k_a} - \frac{w_b}{k_b}$ increases. The relative child penalty for heterosexual, lesbian, and gay couples is increasing as $\frac{w_a}{k_a} - \frac{w_b}{k_b}$ increases.

In Appendix Table A1 we summarize the main predictions of the model. Every mechanism leads to a child penalty that differs between mothers and fathers in heterosexual couples, which is why it is so hard to disentangle mechanisms when looking only at heterosexual couples. Adding same sex couples allows us to distinguish between mechanisms. Based on the model, we can rule out specialization if we compare similar couple types in terms of market and household productivity and we don't see a similar relative child penalty for lesbian and gay couples and heterosexual couples in periods 2 and 3. Biology plays a role if we see an income penalty for the woman giving birth and a relative child penalty for lesbian and heterosexual couples. We can rule out that women simply get greater direct utility from childcare if we don't see an income penalty for both women in same sex female couples.

Perhaps the most surprising result that comes out of the model is the fact that the child penalties for lesbian women due to female preferences for child care will be smaller than the child penalty for heterosexual women due to the same mechanism, which we also show via a simulation of the predicted wages as η, the female preference for child care, increases in Figure 1. The intuition is that in heterosexual couples, the husband will increase labor supply to the market in order to compensate for lost income from the mother, while in lesbian couples both spouses will have to balance their mutual desires to spend more time at home with the need to maintain consumption by providing labor to the market. This will be an important caveat for our results. We also report simulations demonstrating the impact of each of the other mechanisms in Figure

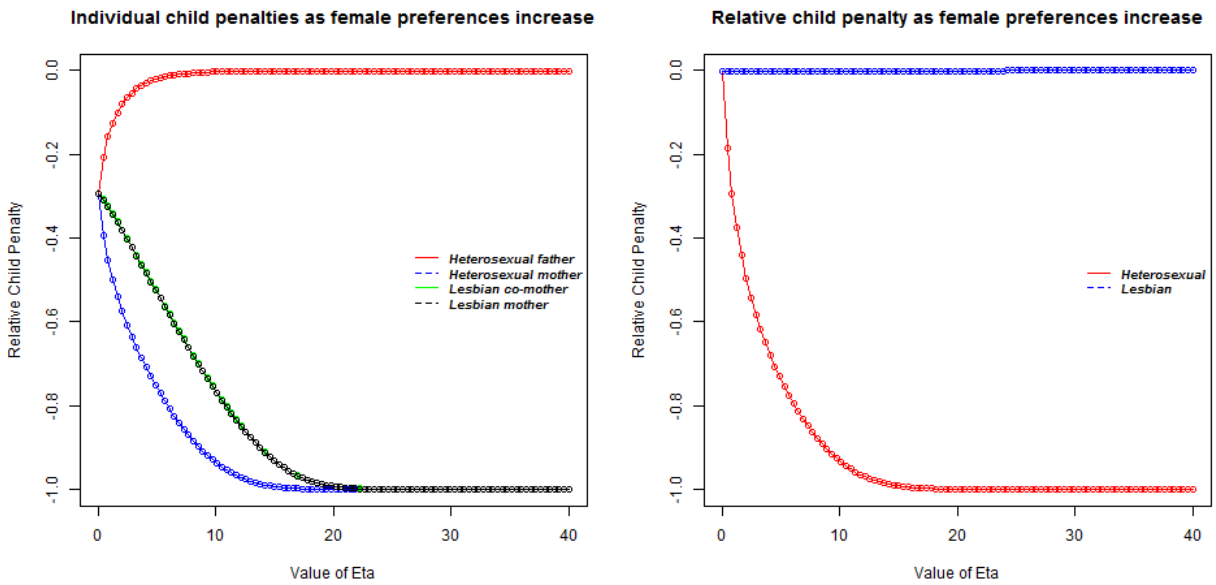


Figure 1: Model Predictions: Simulations for Preferences and Biology

Note: Left panel show individual income penalties relative to full time income in period 1, and right panels show child penalty by couple type. To produce the simulations we set $h(1-t_i) = 1 - t_i$. The baseline parameter values are: $k_a = k_b = 1$, $\lambda_a = \lambda_b = .5$, and $\beta = 5$. At baseline, wages of both partners are normally distributed with mean 10 and standard deviation 1. At baseline $\alpha = \eta = \delta = 0$. In panel 1, we solve for 100 equally spaced grid points of $\eta \in [0, 40]$, keeping all other values fixed.

A1 in the Appendix. These figures plot the child penalty, the percentage change in income relative to the first period of each couple on the left hand side and the relative child penalty, the difference between the child penalties of spouse a and b on the right hand side, based on the time allocations that maximize equation 3 as we vary each parameter (η, δ, α, and w_a) individually.

3 Empirical strategy

To bring the model predictions to the data, we must first identify child penalties across couple types. To identify the child penalty for each partner in each couple type we adopt an event study framework as in Kleven *et al.* (2018). The choice to have children is potentially endogenous to many other determinants of income. However, the precise timing of birth allows us to address this endogeneity. Specifically, if children impact a given labor market outcome of interest such as income, then the precise year in which the child arrives will correspond to a sharp discontinuity in income. Provided the other determinants of income do not also experience discontinuous changes when the child arrives for reasons other than the child's arrival, we can attribute the corresponding discontinuity in income to the arrival of children.

This suggests a simple regression of the outcome of interest on event time dummies to identify child penalties. For our main results we also include gender specific age and year dummies which control flexibly for gender specific life-cycle and time trends in income. The results with only event time dummies are included in Figure A2 in the Appendix and are very similar, but Kleven *et al.* (2018) show that including age and time dummies performs better in identifying child penalties. Event study frameworks such as this have been used to investigate, among other things, the economic impacts of inheritances (Druedahl and Martinello, 2016), hospital admissions (Dobkin *et al.*, 2018) and family health shocks (Fadlon and Nielsen, 2017).[12]

More formally, let t represent event year, with $t = 0$ corresponding to the year in which the couple's first child is born. Let y_{it} be the labor market outcome of interest for individual i at event

[12]Borusyak and Jaravel (2016) revisits the identification problem in event study designs, pointing to the challenge of aggregating post-event dummies and the impossibility of identifying cohort or individual fixed effects together with age and event time dummies. Fortunately, these are not problems in our setting.

time t. We estimate the following equation to identify the child penalties

$$y_{it} = \overbrace{\sum_{j \neq -1} \sum_{k} \alpha_{jk} \mathbb{1}[t = j, K_i = k]}^{\text{Parent-type event time dummies}} + \overbrace{\sum_{l} \sum_{m} \beta_{lm} \mathbb{1}[age_{it} = l, X_i = m]}^{\text{Gender- specific age profiles}} \tag{4}$$
$$+ \underbrace{\sum_{n} \sum_{o} \gamma_{no} \mathbb{1}[T_{it} = n, X_i = o]}_{\text{Gender-specific year shocks}} + \underbrace{\sum_{p} \eta_p \mathbb{1}[K_i = p]}_{\text{Type fixed effects}} + \epsilon_{it}$$

Where X_i is the gender (male, female) of parent i, age_{it} is the age of parent i at event time t, T_{it} is the calendar year for individual i at event time t, and K_i is the parent type: mother or father in heterosexual couple, mother or co-mother in a lesbian couple, and father or co-father in a gay couple. $\mathbb{1}[A]$ is the indicator function for event A. Standard errors are clustered by couple and robust to heteroskedasticity. The event time dummy the year before birth is omitted, which implies that all estimates of event dummies are relative to the year before birth for that specific parent type. Note that while we allow life-cycle and time trends to vary by gender, we do not allow them to differ within gender.[13] Equation (4) is equivalent to running the regressions separately for mothers and fathers if we only estimate the equation for heterosexual couples.[14]

Notice that all parents in our sample eventually have children, so that the event dummies are identified from comparisons of same-aged parents with a youngest child aged j to parents of children at other ages in the same calendar year. Thus, if the exact timing of birth is as good as randomly assigned conditional on gender-specific age profiles and calendar-year shocks, our estimates can be given a causal interpretation as the impact of children on earnings. Kleven *et al.* (2018) show that the event study approach we use here performs well at identifying both short and long run child penalties compared to alternative approaches such as using instruments for first birth.

Our objects of interest are α_{jk}, the change in the outcome for a parent of type k at child age

[13]This means that the effect of age and year on income is the same for all women, be they in heterosexual or lesbian couples.

[14]While it is possible to estimate equation (4) separately for heterosexual mothers and fathers, lesbian mothers and co-mothers and gay fathers and co-fathers, estimating the equation jointly allows us to exploit the large number of heterosexual couples to help identify these control variables for the same sex couples as well as heterosexual couples.

j compared to the earnings the year before birth. Notice that these child penalties include the impact of subsequent children that may appear in later years. Ideally, we would use a log-linear specification of equation 4 so that we could interpret the coefficients as percentage changes in earnings, but the presence of zeros in the outcome complicates matters. To convert these absolute estimates to percentage child penalties, we follow Kleven *et al.* (2018) and construct the following measure of the child penalty.

$$C_{jk} = \frac{\hat{\alpha}_{jk}}{\mathbb{E}(\hat{y} \mid t = j, K_i = k)} \tag{5}$$

The interpretation of C_{jk} is the percentage drop in the outcome for parent type k at child age j relative to the predicted outcome absent children. When computing confidence intervals or standard errors for these estimates, we use a bootstrap, clustering at the couple, to account for the fact that the denominator is an estimated object.

3.1 Comparing heterosexual and same sex couples

The simple event study identifies the causal effect of having children on labor market outcomes of mothers and fathers in heterosexual couples, mothers and co-mothers in lesbian couples, and fathers and co-fathers in gay couples. These results are interesting on their own, so we highlight them below. However, any differences across couples types are only informative regarding the cause of the heterosexual child penalty if the distribution of other factors that may determine changes in labor income around the time of the arrival of children are also identical across couple types. In addition to the differences highlighted by our model, the way to get children is clearly different between same sex and heterosexual couples. In particular, it is reasonable to suspect that the preferences for children is stronger among same sex couples, because the procedure for most of them will involve more costs in the form of money and time.[15]

[15]One might argue that a more natural comparison group for lesbian couples getting children is heterosexual couples getting children through IVF. This is not necessarily the case, however, because heterosexual couples doing IVF have fertility problems, while lesbian couples do not necessarily have any fertility problems. Therefore, one might speculate whether heterosexual couples doing IVF might have even stronger preferences for children than lesbian couples doing the same.

Our model predicts that the relative productivities in labor market and home production of the two spouses, $\frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}}$, will determine the changes in labor income following birth due to household specialization, and these relative productivity differences may not be identical across couple type. To rule out specialization driven by comparative advantage differences across couples, we use the model to motivate two approaches. First, we investigate whether there are still differences in child penalties across couple types conditional on the relative productivity in the couple by adding interactions of $\frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}}$ and the event time dummies to the specification in equation (4).

Unfortunately, we observe neither wages nor home productivity. We observe pre-child incomes, y_{it} and y_{-it} . This is sufficient if $k_i = k_{-i}$, or if one of the following conditions hold. First, our general household model includes household production before and after the child arrives. In that case, specialization will occur before the child arrives and will be captured by pre-market income gaps. Provided the household productivity parameters are unchanged or linearly related over time, then $y_{it} - y_{-it}$ controls for $\frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}}$. Second, if k is instead identical for all women and smaller than k for all men , then controlling for $y_{it} - y_{-it}$ should also be sufficient.

To control for specialization, we flexibly control for the differences in own and spouse's earnings prior to birth interacted with event dummies, by adding $\sum_j \theta_j \mathbb{1}[t = j](y_i - y_{-i})$ to equation (4), with income differences measured at the start of our panel, 4 years prior to birth. To the extent that comparative advantage is captured by the relative income levels of the two spouses, these flexible event dummy controls will pick it up and we can attribute the remaining child penalties from α_{jk} to the other possible mechanisms highlighted by the model. Notice that these controls capture more than the intended comparative advantage. In particular, they also capture the autocorrelation in earnings over time. When presenting these results, we scale by the predicted earnings from the baseline estimates in equation (4), and bootstrap confidence intervals for the scaled results clustering on couple. We interpret any remaining child penalties in earnings as coming from sources other than specialization. As an alternative, we control for the differences in years of education interacted with event time dummies, another measure related to labor market productivity.

Second, in case the (untestable) assumptions required for the first approach to work do not hold, we also report results using propensity score matching to construct samples of heterosexual couples that are similar to lesbian couples based on pre-birth observables. To this end, we estimate a logit model for the probability of being a lesbian couple in the sample of lesbian and heterosexual couples, using as covariates a full set of municipality dummies to capture urban/rural differences, both spouses' age at birth and their interaction, indicators for number of children and both spouses' years of education and their interaction. We do not match on pre-birth earnings, as this is our outcome and could lead to over matching. We then re-estimate equation (4) using the propensity score estimates as weights to get a sample of heterosexual couples that are more similar to lesbian couples based on pre-birth observables. We bootstrap the entire procedure clustering on couple.

4 Institutional context, data and sample selection

Norway was the second country in the world to legally recognize same sex partnerships in 1993 through the Partnership Act, and Figure 2 documents the number of new same sex male and female partnerships in Norway in the following years.[16] Under this act, a partnership was legally equivalent to marriage in most respects. However, partnerships were restricted regarding children. Same sex couples were not eligible for domestic adoptions, were not eligible for publicly subsidized assisted fertility treatment, and the registered spouse of a woman giving birth was not automatically registered as the second parent (as the *pater est* principle established for married heterosexual couples). It wasn't until 2002 that a change to the rules for adoptions allowed same sex couples to formally adopt the children of their spouse. This change to the guidelines allowed same sex couples to be considered for adoption of stepchildren just like heterosexual couples. The guidelines required a stable relationship and having had a *de facto* parenting role for the child in question for some period of time, most often five years, as well as consent from the existing parent. If the child was already registered with two parents, the other parent was given the right to

[16] Aarskaug Wiik *et al.* (2014) investigates the stability of these same sex marriages and partnerships, and find that they are less stable than heterosexual marriages.

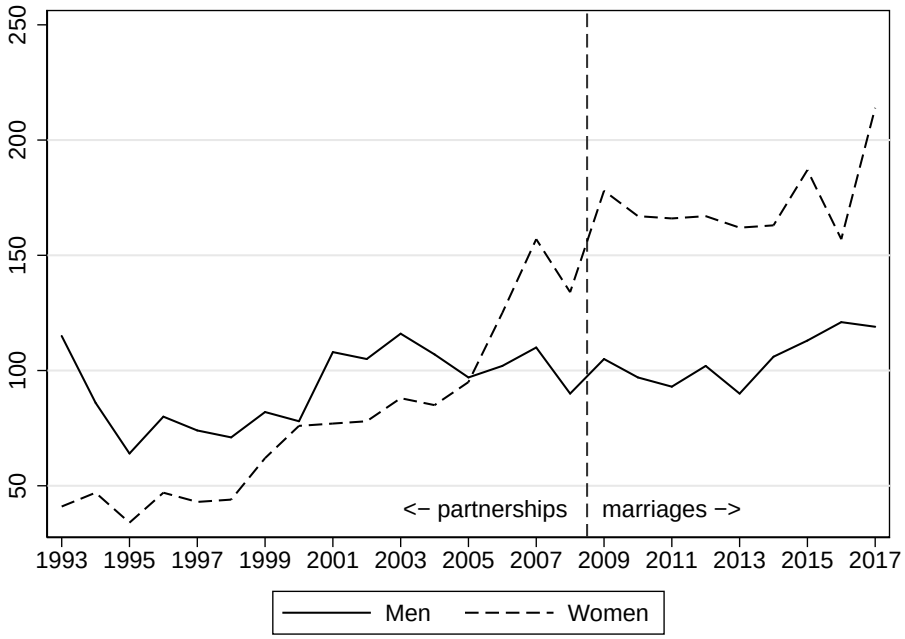


Figure 2: Number of new same sex partnerships and marriages in Norway, 1993-2017

Source: Statistics Norway *Statistikkbanken*, tables 10160 and 05713.

express his opinion on the adoption, but the case was ultimately decided by the adoption agency.

In practice the increasing use and availability of assisted fertility treatments among lesbian couples challenged this five-year rule, as planned children of lesbian couples conceived through assisted fertilization abroad became increasingly common. Therefore, in 2006 the Norwegian government clarified the rules so that the five-year rule would not apply in cases where the fatherhood cannot be established, such as with IVF treatment using an anonymous donor. In 2009, a new marriage act was introduced which equalized same sex and heterosexual marriages in all but one respect: A same sex spouse cannot later adopt the child of his/her spouse that was in turn adopted from a country that does not allow adoptions to same sex couples. The new marriage law from 2009 also gave lesbian couples the right to IVF treatment in Norway, but only when using non-anonymous donor, as the law requires all children conceived through IVF in Norway to have the possibility of knowing the identity of the donor father at age 18. Before this, lesbian couples often traveled abroad to get IVF treatment, most often in Denmark. Even after the new law was passed, many couples still travel abroad either to speed up the process or

because they want to use an anonymous donor. If conception happens through IVF treatment with a non-anonymous donor in a recognized (private or public) fertility clinic, co-mothership can now be registered at birth, but otherwise the couple must go through an adoption process in order for the partner to be formally registered as the co-mother.

For gay couples, getting children is naturally more complicated. Surrogacy is illegal in Norway, but some gay couples still enter into surrogacy agreements with surrogate mothers from abroad. No special rules apply to these children, and parenthood must be established according to the law when returning with the child. Typically, this means that the (most often biological) father will declare fatherhood upon returning to Norway and be registered as the father, and that the other spouse will then have to start the adoption process to be registered as co-father. Alternatively, gay and lesbian couples have formally been eligible for adoption since 2009 just like heterosexual couples, but this possibility is typically limited by the lack of donor countries willing to adopt children to these couples.[17] Domestic adoption at birth is very rare in Norway,[18] but some children are adopted by their foster parents after a number of years in foster care. This typically happens at much later ages and we would not expect this to have an impact on labor market status around the birth of the child.

We observe registrations of legal parent status in the population registers. In practice, we therefore observe children appearing in same sex couples at various times following birth. When identifying births to same sex couples in the administrative data, we try to be as certain as possible that we capture planned arrivals of children by a same sex couple that happens in the year of birth of the child, without losing too many observations because children often aren't legally registered with both parents until the following year.

Following birth, Norwegian parents have been entitled to a generous paid parental leave since

[17]The first adoption from abroad to a same sex couple in Norway happened in the fall of 2017, when Colombia became the first donor country to approve an adoption to a Norwegian same sex couple following a controversial Supreme Court ruling from 2015. In the empirical analysis, we restrict attention to children born in 2014 at the latest, so that foreign adoptions to same sex couples will not be relevant for this paper.

[18]Ruling out adoptions by near family and adoptions of foster- and step-children, as few as two to three children are adopted away at birth or right thereafter per year in Norway. In addition, the biological parents are given a say on prospective adoptive parents, and their opinion is given considerable weight in the decision among potential adoptive parents. This makes matters worse for same sex couples if the biological parents prefer a heterosexual couple. In practice, this means that this option is not very relevant for same sex couples.

1977. Total parental leave is currently 49 weeks at 100% replacement or 59 weeks at 80% replacement rate, but the length of leave has been steadily increased since the mid 1980's, reforms that we exploit and describe in more detail in Section 6. Benefits are capped at around 600,000 NOK or 70,000 USD. The leave is split in three with a quota for the mother, one for the father (since 1993) and the rest to be distributed among the parents. Leave spells can also be graded, allowing parents to combine work and leave for a longer period of time. A parent must be legally registered as a parent to the child at the time of leave start.

In order to qualify for leave, a parent must have been employed for at least 6 of the 10 months prior to birth, and the annual earnings must exceed a low threshold of around 50,000 NOK or 6,000 USD. Benefits from sickness absence or some other benefits may qualify as earnings for meeting this requirement. Mothers who do not qualify for parental leave are entitled to a one-time-benefit of 63,000 NOK or approximately 7,600 USD. In addition to paid leave, all parents have job protection for another year if they want to take additional unpaid leave. Taken together, this means that the total leave uptake is a much better measure of the time the father spends off work with the child than the mother, because mothers more often stay home with the child on unpaid leave than fathers and also stay home using the one-time benefit when they are not eligible for parental leave.

Following parental leave, Norway has a well developed, regulated, and highly subsidized child care sector with high coverage, as documented in figure 13a. The alternative to sending children to formal care is mostly home care by the parents, for which there is a cash for care benefit given to the parents of young children[19] who do not use the subsidized formal care system. Because of the heavy subsidies for formal care, the market for paid child care outside this system is very small, but subsidies are available for both private and public suppliers of formal care.

[19]The age eligibility criteria has varied somewhat over the period, but cash for care is now available for children aged 13 - 24 months only. The benefit is relatively generous at 7,500 NOK or 900 USD per month, assuming no formal care use.

4.1 Data and sample selection

Our data comes from Norwegian administrative registers covering the entire resident population. Through unique identifiers we link individuals over time and to family members such as parents, enabling us to identify couples around the time of the arrival of a child. Data on residency status, date of birth, gender, municipality of residence and links to mothers and fathers comes from the official population register, and is provided on January 1st every year from 2000 onward. We obtain data on education for the years 1980 - 2016 from official education registers on the level, field and length of education as well as whether or not an individual is enrolled in a study program by October 1st each year.

Our labor market outcomes come from two sources. The primary data on annual labor market earnings comes from the tax records. Importantly, these are wage incomes that include taxable benefits such as sickness and parental leave and benefits.[20] We also observe employment spells from the FD-Trygd database. These cover most important employment spells from 1992 - 2003 and all employment spells (not self-employment) from 2003 - 2014. To create comparable measures across most of the sample period, we exclude spells of self-employment from the pre-2003 data and include only the employment spell with the most contracted hours for the post-2003 data.[21] From these spells, we construct the following measures of *monthly* labor supply, measured for the spell that covers the 15th and 16th of each month: Dummies for the employment spell exceeding 4, 20 and 30 contracted hours per week, whether the primary employment is in the public sector (2003 - 2014 only) and a proxy measure of the family friendliness of the firm. The latter measure is the leave-out-mean of mothers with children below 15 years that work in the firm. In addition, we measure the total working hours of all employment spells for the years 2003 - 2014.

For parental leave and sickness absence spells we also pull data from FD Trygd, the register of the Norwegian Public Insurance system. For sickness absence, we measure the number of sickness days due to physician-certified spells of leave that exceed 16 days in a given month, scaled by the grade in the case of graded sickness absence to measure efficient days lost. For

[20] We set negative incomes to 0, comprising less than 0.2% of the observations.

[21] In more than 95% of the cases, the spell considered most important in the pre-2003 data is the one with the longest contracted hours.

parental leave spells we measure how many weeks of leave were taken for a particular child, which we infer from the start and stop dates of the leave spells and birth dates of the children. Details on this measure is provided in Appendix B.1.

Finally, we exploit data on child care use and availability. For the measure of child care slots, we use administrative data from the child care centers on the number of slots for children of different ages by December 15th each year. At the individual level, however, we can measure the exact use of child care at ages 13 - 36 months for the years 2000 - 2011. For these years, a cash for care benefit was given to children who did not attend formal care in a given month. If we assume that all children who do not use child care apply for the benefit, which is relatively generous,[22] we know exactly which children attended how much care for each month. From these data, we construct precise measures of full-time equivalent years of child care use from ages 13 - 36 months.

For our *main sample* of same sex and heterosexual couples, we want to be as certain as possible that we capture the arrival of planned children in a household with two parents. This is more challenging given that the formal adoption process to the other parent in some cases may take time. We therefore start with the universe of children born in Norway in the years 2001-2014. We assign the parents to be the first parents ever registered to the child, which gives us a large number of heterosexual parents and a small number of same sex parents. This approach allows for one of the parents to be missing for a year or two until the legal adoption procedure is completed. We restrict attention to children where both parents were legally registered as parents at the latest in the year the child turns 3 in order to minimize the risk of capturing partners not present at birth, and also to avoid getting an unbalanced sample of children even in the year of birth.

We furthermore keep only first-born children to both parents. In case of multiple births, we keep the couple in the sample only once. We drop a handful of lesbian couples who receive multiple kids in the same year and register different parent status for each child, and keep only couples where both spouses reside in Norway the year before birth. Lastly, we keep in both samples only couples where the first child appears at ages 22 to 60 for both parents, giving us

[22]Throughout 2001-2009, which is the period we exploit, the benefit was around 3,500 NOK or 420 USD per month, but varied somewhat.

some time before and after birth to observe earnings.

For *the long sample*, which we use only in the long-run analysis of changes to the child penalty over time, we start with all children born 1971 to 2010 where both mother and father are registered. We restrict attention to first-born children of both parents, and in cases of multiple births we include the parents only once. We drop a small number of couples where one of the parents (most often the father) had several children with different people in the same year, and drop kids with same sex parents. Unfortunately, we do not observe residency status or changes of legal parent status before the year 2000, which means that we may be allocating a very small number of later adoptees to their adoptive parents even before the adoption happens.[23]

This leaves us with a main sample of 250,296 heterosexual couples, 634 lesbian couples and 32 gay couples, and a long sample of 721,291 heterosexual couples. We match these mothers and fathers to their labor market earnings in all years from $t-4$ to $t+5$ or $t+15$, centered around the birth of the first child, to investigate labor market responses to the child's birth. Note that for children born after 2002, we will not see a full 15 years of income after birth because our data ends in 2017. Since most children born to same sex couples are born late in the sample period, we see later labor market outcomes less frequently for same sex couples relative to heterosexual couples. For the main sample we therefore restrict the window of interest to be between $t-4$ and $t+5$ to limit this imbalance. Summary statistics for these samples are given in Table 1. The population of lesbian couples is reasonably large. In contrast, the number of gay couples with children is very small, which corresponds to very imprecise estimates for this group in the next section. As expected, the population of heterosexual couples with children is very large. We can also see that same sex couples have much higher pre-birth labor earnings relative to heterosexual couples. This suggests that it might be important to control for income and initial income gaps in order to compare the child penalty between similar heterosexual, lesbian and gay couples as described in Subsection 3.1. Lesbian couples are slightly older than heterosexual couples at first birth, and are also slightly more educated. Reflecting the rules on establishing legal co-parent status, the age

[23] In the main sample of heterosexual couples we see that the average age at which both parents are first registered is 0.02 years, indicating that this problem should be extremely minor, and probably smaller back in time due to higher marriage rates.

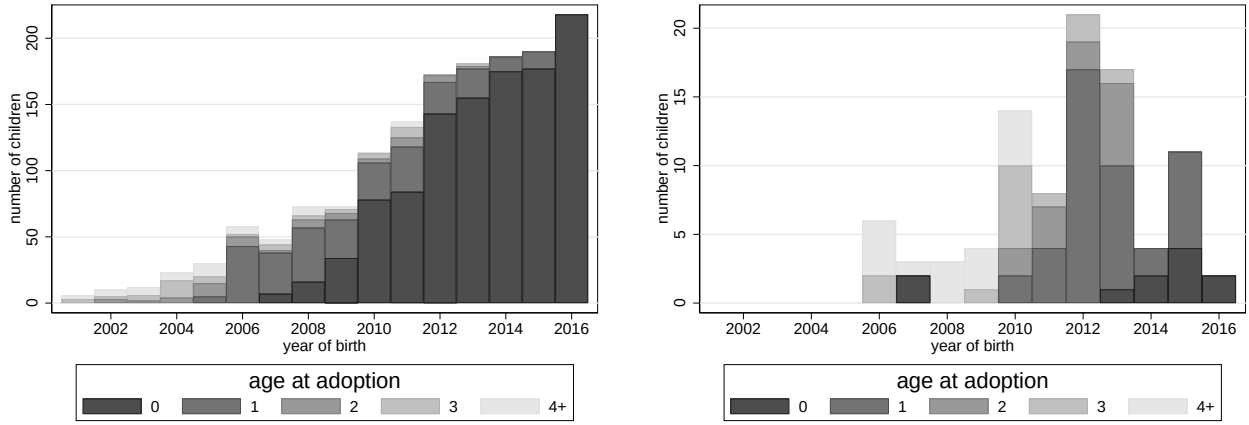
Table 1: Summary statistics by couple type

	Heterosexual couples		Lesbian couples	Gay couples
	Long sample	Main sample	Main sample	Main sample
Birth year (first child)	1971-2010	2001-2014	2001-2014	2001-2014
A: Child characteristics				
Birth year	1992.0	2007.7	2010.7	2011.9
	(11.5)	(4.00)	(2.87)	(1.60)
Multiple birth	0.015	0.020	0.069	0.25
	(0.12	(0.14)	(0.25)	(0.44)
Female child	0.49	0.49	0.48	0.52
	(0.50)	(0.50)	(0.49)	(0.48)
Age at adoption		0.022	0.48	1.34
		(0.17)	(0.81)	(0.94)

B: Parent characteristics, year before birth

	Mother	Father	Mother	Father	Mother	Co-mother	Father	Co-father
Parent type (K)	1	2	1	2	3	4	5	6
Age at first birth	26.3	28.7	27.8	30.3	32.2	32.8	38.4	38.2
	(4.04)	(4.83)	(4.23)	(5.00)	(4.12)	(5.64)	(5.21)	(6.05)
Labor income	250.3	346.3	362.7	487.3	488.9	480.0	737.4	813.2
(1,000s of 2017 NOK)	(152.4)	(243.6)	(188.9)	(314.6)	(196.7)	(308.3)	(264.3)	(378.7)
Years of education†	14.2	14.0	15.2	14.6	16.4	16.0	17.2	17.1
	(2.94)	(3.00)	(2.89)	(3.01)	(2.42)	(2.65)	(2.44)	(2.57)
N couples	721,291		250,296		634		32	

Note: Summary statistics on estimation samples constructed as described in Section 4. Standard deviations in parentheses. †Available from 1980 and onward only.



(a) Lesbian couples

(b) Gay couples

Figure 3: Registered children to same sex couples, by year of birth and age at adoption

Notes: Own calculations, based on sample and data described in Section 4. Age at adoption refers to the age of the child in the year we first observe both parents registered.

at adoption is slightly delayed for lesbian couples compared to heterosexual couples, as it takes some time for the co-mother to be legally registered.

5 Results

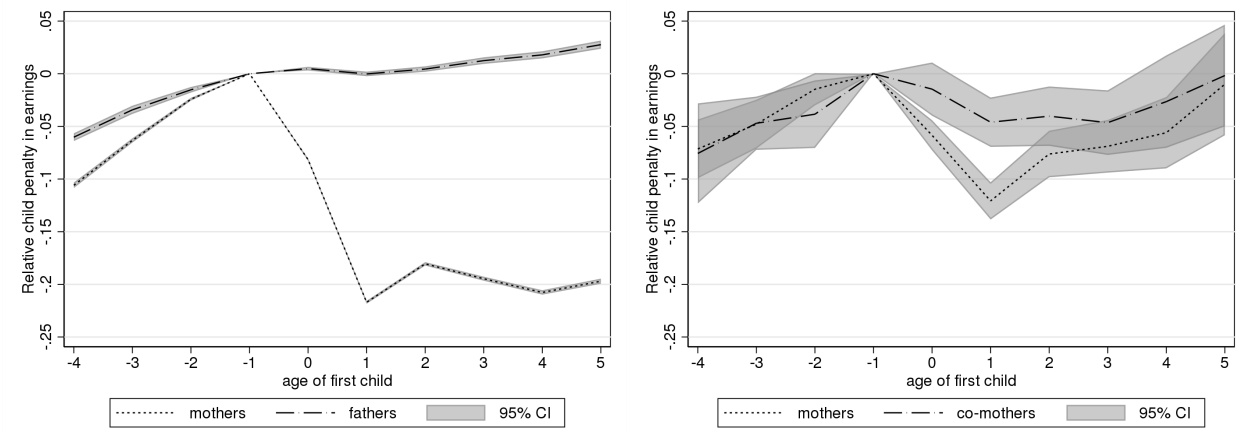
In this section we provide the main results on child penalties. First, Section 5.1 provides the main estimates of child penalties across couple types, including household child penalties and our attempts at controlling for specialization using and pre-birth earnings and education differences. Next, Section 5.2 decomposes the child penalties into a series of determinants of income that allow us to investigate how lesbian parents' labor supply choices differ from heterosexual parents'. Lastly, Section 5.3 provides some evidence that the different parenting style that lesbian parents' adopt, as evident by their different child penalties, does not seem to come at the expense of children's long-term schooling outcomes.

5.1 Heterosexual and same sex child penalties

In Figure 4 we present the main results. The graphs report estimates of C_{jk} (see equation (5)) generated by the simple event study in equation (4). Starting with the first row, the results for heterosexual couples are shown on the left and lesbian couples on the right. Results for gay couples are shown in the second row.[24] As has been shown in many other papers, we also find that mothers in heterosexual couples experience large income penalties in the range of 20% of their counterfactual earnings in the absence of children upon the birth of their first child. Fathers experience no income penalty upon the birth of the first child. The graph for lesbian couples is strikingly different. We find that both mothers experience a child penalty the year after the child is born, but initially the woman who gives birth has a child penalty more than double the size of her partner. The drop in income, however, is much smaller than that of heterosexual mothers, at around 13% and 5% of counterfactual earnings for mothers and co-mothers, respectively. Moreover, 2 years after birth the woman who gives birth catches up and her penalty is no longer statistically significantly different from her partner's. By five years after birth, the child penalty for both women has largely disappeared.

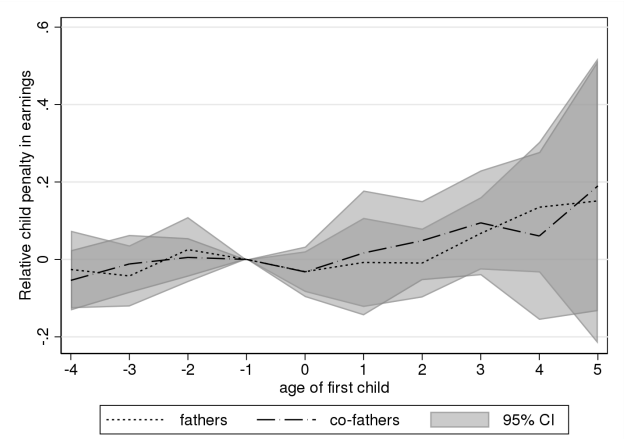
The fact that the lesbian partner who gives birth initially experiences a larger child penalty than her partner suggests that biology plays a role in the child penalty, but only in the first year after birth. The fact that both partners experience child penalties, and that those penalties are statistically indistinguishable from 2 years after birth onward, suggest that women have a preference for time with children over career. Note that an alternative formulation of the model might assume that this preference η is larger for the mother who gives birth within a lesbian couple than the mother who does not, given that which mother gives birth is endogenous in lesbian couples. However, if this is the case then we would expect to see a persistent gap between lesbian mothers in later years, in contrast to the catch up that we find. The last graph in Figure 4 corresponds to gay couples. Consistent with the small population size, the estimates are very imprecise. However, the patterns are consistent with a gender norms, preferences, and biology

[24] In Appendix Figure A2 we also report the raw mean earnings by event time for each couple type, without imposing any of the structure from equation (4), and the results are quantitatively similar.



(a) Heterosexual couples

(b) Lesbian couples



(c) Gay couples

Figure 4: Estimated child penalties across couples types

Note: Figures show the estimated child penalties from equation 4, scaled as described in eq. 5. Sample construction and data as defined in section 4. Bootstrapped 95% confidence intervals in gray using 200 replications and clustering by couple. Note that the scale of the y-axes are separate for gay couples compared to heterosexual and lesbian couples.

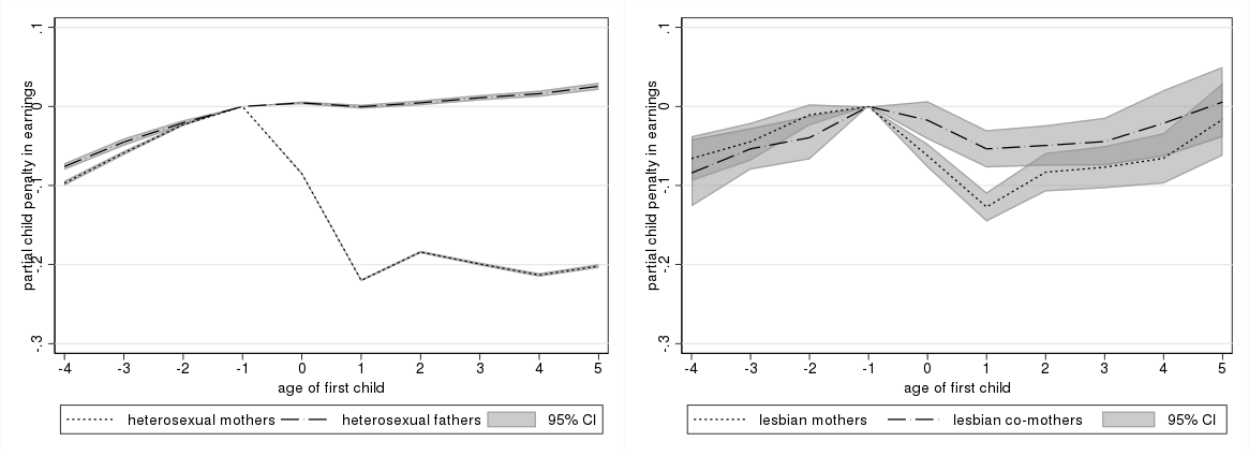
story. In the event study, neither partner experiences a child penalty.

These results are suggestive, but without removing the contribution of specialization we cannot definitively pinpoint mechanisms. The impact of specialization might differ across couple types and we observe quite different distributions of earnings and education before birth between couple types. To address this, in Figure 5 we report estimates controlling for household specialization as measured by income or education differences before birth interacted with event dummies, as discussed in Subsection 3.1. Note that this figure presents the remaining child penalty after removing the portion of the penalty explained by the productivity differences. Except for some differences in the impact before birth that is likely caused by autocorrelation of incomes over time for the income differences specification, the figures are remarkably similar to the baseline estimates for both the income and education measures of relative productivity. This suggests that specialization alone cannot explain the differences across couple types that we see. Note that we only report results for heterosexual and lesbian couples, given the large imprecision in the estimates for gay couples.

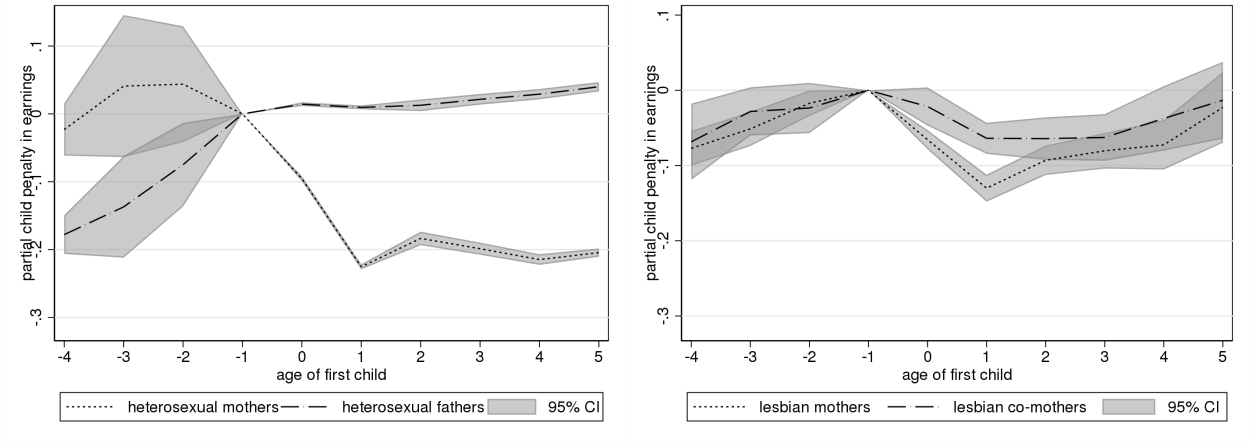
Finally, Figure 6 presents results from the matching exercise, where we have constructed a sample of heterosexual couples similar to the lesbian couples on pre-birth observables such as education, age and municipality of residence. Panel (a) of this figure reveals that the child penalty for the matched sample of heterosexual couples looks very similar to the baseline child penalty for heterosexual women, indicating that the different child penalties between the couple types documented so far are not driven by differences in the pre-birth covariates we include in the propensity score model. Although precision is lower in the sample of lesbian couples, the results are roughly comparable to the baseline estimates.

Last, we point out that these results do not appear to be driven by differential fertility. First, in Appendix Figure A4 we show that lesbian and heterosexual couples have almost identical completed fertility during this period. Second, in Figure A5 we repeat the main exercise but restrict to heterosexual and lesbian couples who do not have an additional child five years after the birth of their first child. We find that the results are unchanged.

The child penalty experienced by women in heterosexual couples is so large, it would seem

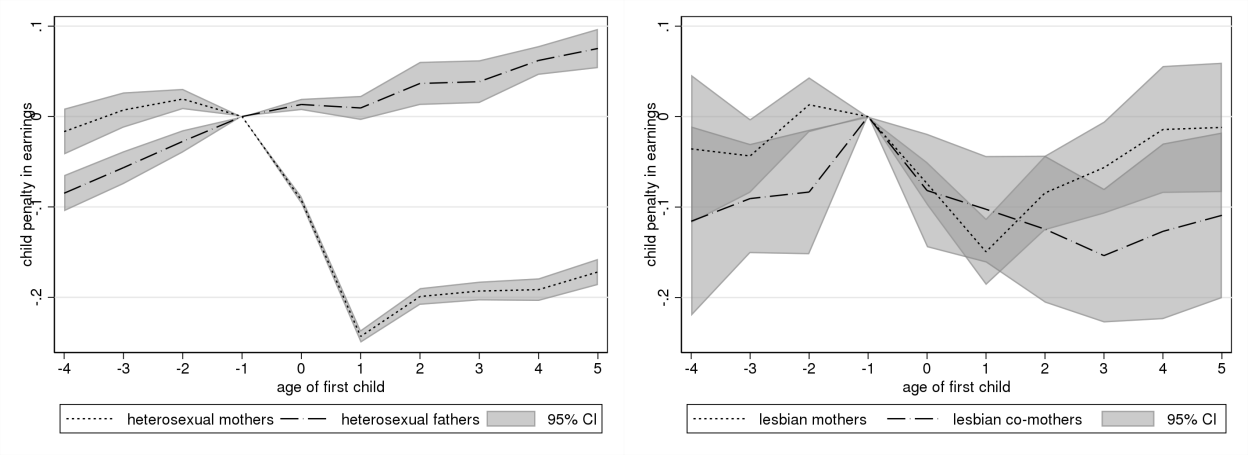


(a) Using years of education differences in $t-1$



(b) Using labor market income differences in $t-4$

Figure 5: Partial child penalties, controlling for comparative advantage



(a) Heterosexual couples

(b) Lesbian couples

Figure 6: Child penalties in a matched sample

Note: Figure shows child penalties estimate from the baseline model in a sample matched to the lesbian couples on pre-birth characteristics. In the sample of heterosexual and lesbian couples before birth, we estimate a logit model for being a lesbian couple. Covariates include both spouses age and their interaction, both spouse's years of education and their interaction and a full set of municipality dummies. We then weight the baseline model with the propensity score from this model, so that heterosexual couples who look more like lesbian couples on observables are given higher weight. The entire procedure is bootstrapped, clustering on couple.

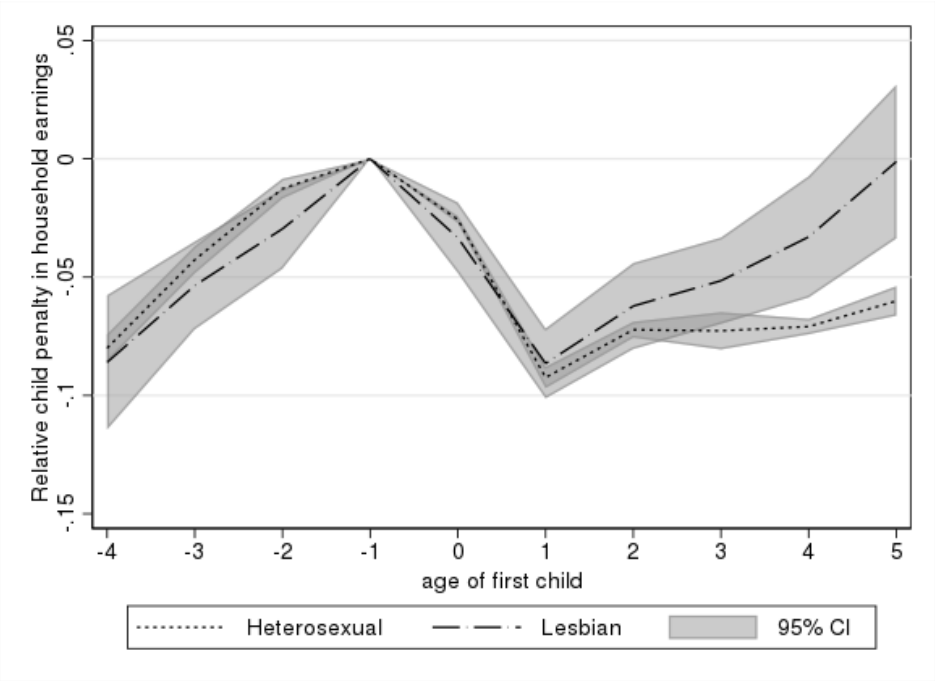
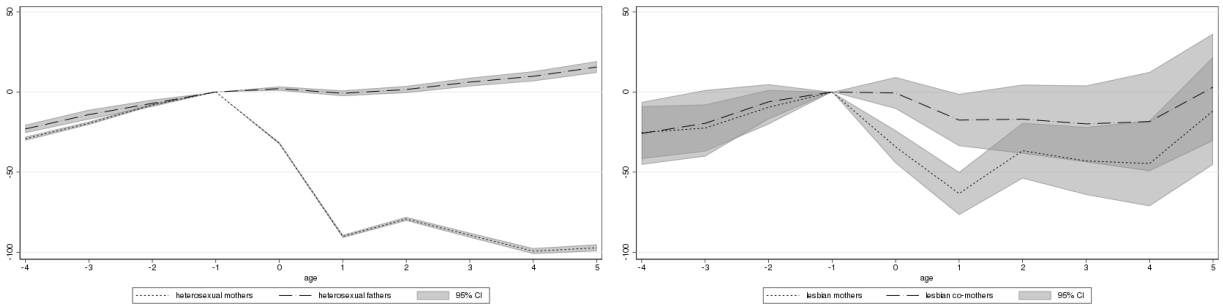


Figure 7: Child penalty, total household income

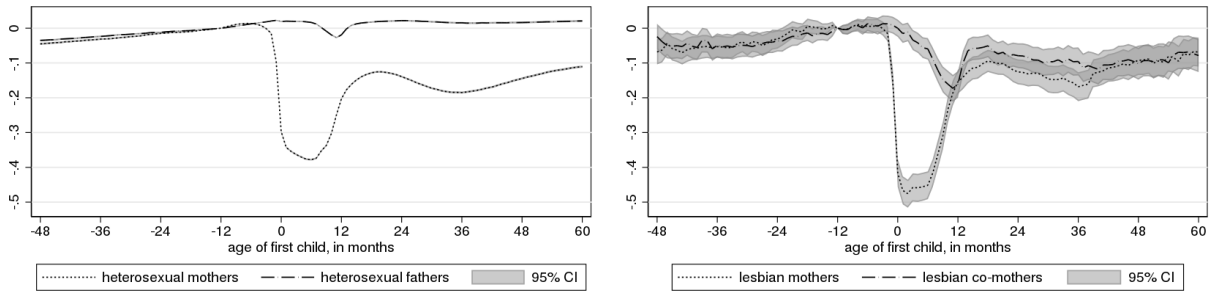
to imply an overall household income penalty. In Figure 7 we show this is the case by using the total income of the two spouses as the outcome. What is particularly interesting is that both lesbian and heterosexual couples experience the same initial income decline on the birth of the first child. However, this drop in income persists for heterosexual couples while it decreases over time for lesbian couples. Note that this is despite the fact that lesbian and heterosexual couples have relatively similar completed fertility, as we show in Figure A4 in the Appendix. We again exclude gay couples from this analysis due to the small sample size, but as you would expect based on the previous figures, gay couples experience even smaller household income penalties compared to lesbian and heterosexual couples.

5.2 Decomposing the child penalty

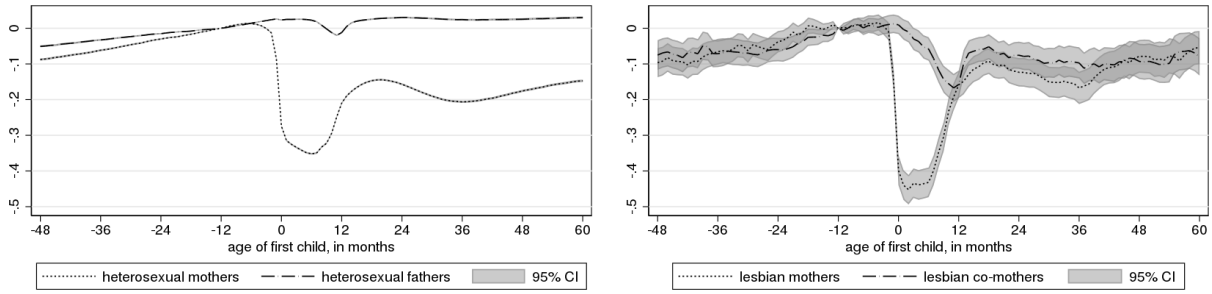
To further understand the anatomy of the child penalty and what lesbian couples do differently than heterosexual couples, we estimate the child penalty separately for the following determinants of income: extensive margin participation, an indicator for full time work, weekly contracted hours of work, family friendliness or public sector status of the firm, and days of sick leave. Just like the baseline event study, we construct a panel from 48 months before birth to 60 months after birth, and regress the outcomes on parent type-specific event time dummies and gender specific age profiles (in months) and monthly shocks. Unlike the baseline, to ease interpretation of the various mechanisms, we do not scale the estimates like in equation 5. Therefore, the estimates are interpretable as the effects of children at age (in months) j, relative to the effect 12 months before birth. Results are presented in figures 8 and 9. We begin in figure 8 by repeating the baseline estimates, but unlike in Figure 4 these are unscaled. As expected, the child penalties look largely the same as the baseline results with an immediate drop of around 100,000 NOK (approximately 11,600 USD) for mothers in heterosexual couples that persist over the period we investigate and a smaller and decreasing penalty for lesbian mothers. In panel (b) we plot effects on the extensive margin of having any active employment relation. Unlike the baseline outcome of labor earnings, we see a strong dip in employment around the time of child birth for mothers, driven by employment spells not being active when mothers are on leave in contrast to mater-



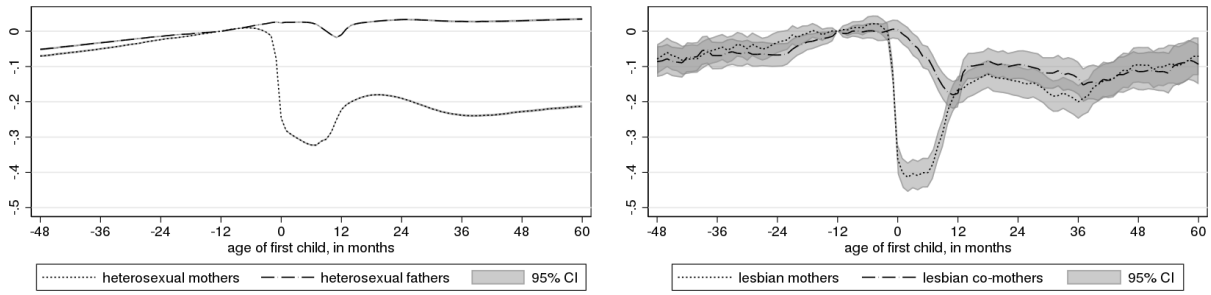
(a) Total labor income, 1,000 NOK (baseline outcome)



(b) Main employment relation at least 4h/week contracted

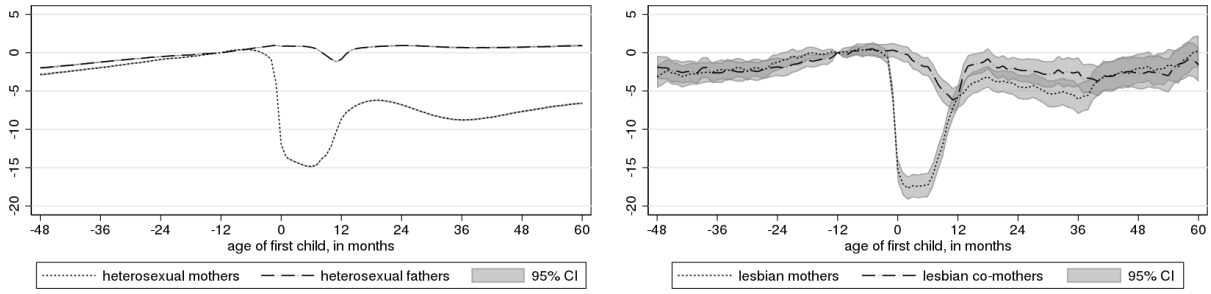


(c) Main employment employment relation at least 20h/week contracted

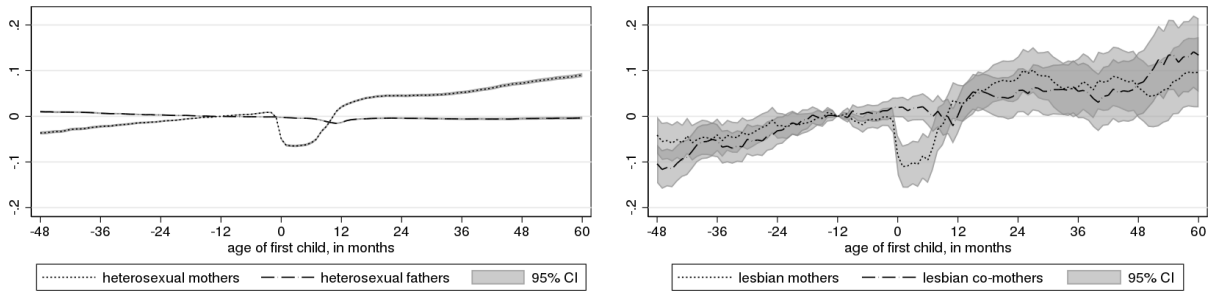


(d) Main employment relation at least 30h/week contracted

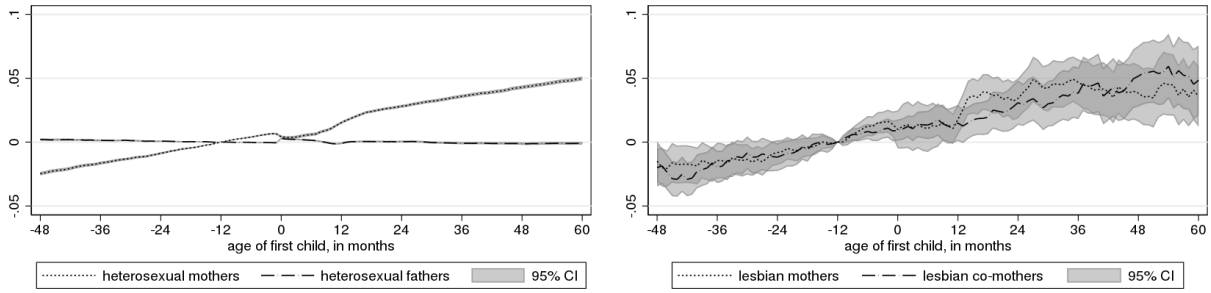
Figure 8: Decomposition I: Child penalties for heterosexual (left) and lesbian (right) couples



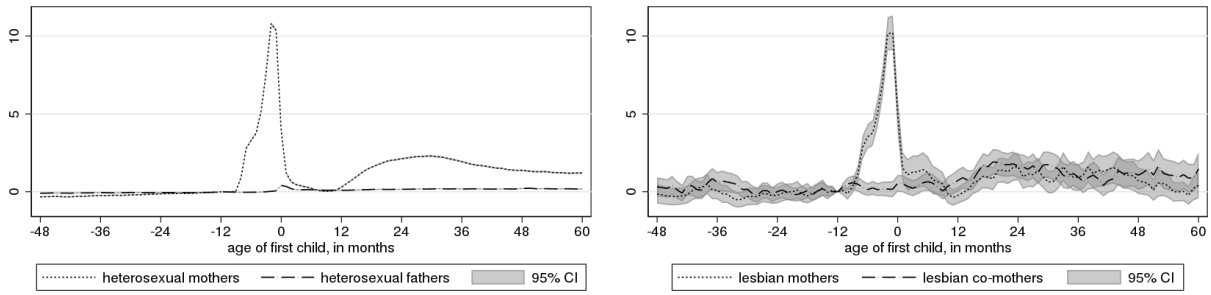
(a) Weekly contracted hours in all employment relations, 2003 - 2014



(b) Main employment relation in public sector, 2003 - 2014, conditional on working



(c) Family friendliness of employer, conditional on working



(d) Days of sickness absence for spells exceeding 16 days, conditional on working

Figure 9: Decomposition II: Child penalties for heterosexual (left) and lesbian (right) couples

nity leave benefits that replace earnings and are included in our income measure. Following the initial dip, employment bounces back but stays below -0.1 for the period under study, indicating 10 percentage points lower probability of being employed compared to the baseline employment rate 12 months before birth. In panel (c) we estimate impacts on a dummy indicating a full time job, as defined by contracted weekly hours above 30. The fact that the impact on this measure is larger than on the employment measure, at around a 20 percentage points reduction, indicates that there is response both on the extensive and intensive margins of labor force participation: some mothers drop out of the labor force entirely while others reduce labor supply and work part time following child birth. As before, we find little response among heterosexual fathers for these measures.

For lesbian mothers, the response on the extensive margin of labor supply is slightly smaller, but largely in line with the results for heterosexual mothers. Furthermore, when excluding the immediate dip in employment that is cause by parental leave directly, lesbian co-mothers behave similarly to their partners, reducing labor force participation by around 10 percentage points in response to child birth. For the full time measure, however, the reduction is markedly smaller for lesbian mothers than heterosexual mothers, indicating that part of the differences in income patterns are driven by more mothers working full time in lesbian than heterosexual couples following child birth. This difference is mirrored in the outcome for total hours on top of Figure 9, which we can measure for 2003 - 2014 only. Here we see reductions of total contracted hours of around 10 hours for heterosexual mothers, while the response among lesbian mothers is smaller and fully recovers 4-5 years after birth. Lesbian co-mothers behave much like their partners after the first year of leave, while heterosexual fathers increase total contracted hours. Summing up, the differences in the child penalties between heterosexual and lesbian mothers seem to be driven by differences in the response on the intensive, not the extensive margin.

Following Kleven *et al.* (2018), we also estimate the impact on two measures of workplace flexibility. The first is a dummy for whether the employer is in the public sector, which is known for its flexibility and well regulated working conditions. The second is a measure of family friendliness that we construct at the firm-month level. It represents the share of mothers of children

below 15 years of age among the other workers who have their primary employment relation with the firm. Both of these measures, however, are defined only for employed people; since we have shown that employment is endogenous to child bearing, these should be interpreted with care. That caveat aside, the child penalties for these outcomes are plotted in panel (b) and (c) of figure 9. We see strong positive trends in public sector employment for mothers in heterosexual couples around child bearing. Ignoring the dip in the year of birth that is likely caused by the very low employment rates of new mothers, mothers move into the public sector in anticipation of - and following - child birth, while this trend is flat for men. The trend in this outcome is relatively similar for both partners in lesbian couples. Our measure of family friendliness suggests that all types of mothers move to more family friendly firms in the period up to and following birth. The fact that lesbian mothers do not experience long term child penalties, but are just as likely as heterosexual mothers to move into family friendly firms, suggests that occupational selection in response to children cannot fully explain the gender income gap post children.

Finally, we use a measure of days of sickness absence to see if childbirth may cause longer term health shocks that impact income. The measure counts the full-time equivalent days of absence due to sickness from physician-certified spells of leave that exceed 16 days, so will generally not include short term illness such as seasonal cold or flu. It also include sickness absence spells for dependents that require the employee to be absent, in particular young children. As with the measures of family friendliness, this measure is conditional on employment.[25] Results indicate an unsurprising spike in sickness absence for heterosexual and lesbian mothers who will eventually give birth during pregnancy. The results during the maternity leave period for most of the first year should be interpreted with care, as the measure of sickness absence is conditional on employment, but sickness absence eventually stabilizes at a higher rate than before birth.[26] The pattern is relatively similar for both partners in lesbian couples. Heterosexual fathers also take slightly more sickness absence after the birth of children than before.

[25]Despite this, we occasionally see non-employed individuals in these data. We exclude the few non-employed individuals who are registered with absence spells.

[26]Note that some of this could be caused by subsequent pregnancies.

5.3 Child test scores

We have shown that individuals in same sex couples share the burden of child rearing more evenly, and experience less severe household income penalties compared to heterosexual couples. It is natural to ask if this reduction in the relative (and total) child penalty comes at the cost of worse outcomes for children. If sharing the burden of child care is simply more efficient, then same sex couples and their children could be better off than heterosexual couples and their children. Alternatively, same sex couples could be choosing to substitute purchased child care for home production, in which case their children could be equally well off, depending on the quality of care. Last, same sex couples could be investing less in their children, in which case their children would be worse off.

In Table 2 we present results from a simple regression of test scores at age 10 for the children of heterosexual and same sex couples on a dummy for having same sex parents and an increasing set of control variables across columns. Standard errors are clustered by both parents using two-way clustering. The results in the first column, corresponding to no controls, indicate that children of same sex couples do much better than children of heterosexual couples, in the range of 0.4 to 0.6 standard deviations in the three subjects. Moving right, we gradually add more controls for observable pre-birth differences between same sex couples and heterosexual couples. Education level in particular reduces the differences quite a lot, but children of same sex couples still do around 0.2 standard deviations better in both reading and English even when controlling for our large range of observable characteristics. These results suggest that while same sex parents appear to parent more equally and experience smaller costs to overall household income, their alternative approach to child rearing does not come at the cost of child outcomes.[27]

[27] Although a further analysis of the relative performance of children from same sex and heterosexual couples is beyond the scope of this paper, these results might also indicate stronger positive selection into child bearing among lesbian couples that is not accounted for by our rich set of controls.

Table 2: Impact on children: Test scores at age 10

Outcome variable	(1)	(2)	(3)	(4)	(5)
Math	0.395***	0.363***	0.283***	0.0893	0.0766
	(0.0858)	(0.0853)	(0.0853)	(0.0835)	(0.0838)
Reading	0.410***	0.352***	0.263***	0.146*	0.170**
	(0.0832)	(0.0833)	(0.0836)	(0.0821)	(0.0810)
English	0.565***	0.529***	0.433***	0.248***	0.235***
	(0.0800)	(0.0794)	(0.0803)	(0.0773)	(0.0777)
Pre-birth controls					
Child gender	✓	✓	✓	✓	✓
Birth year dummies	✓	✓	✓	✓	✓
Age dummies (mother × father)		✓	✓	✓	✓
Municipality dummies			✓	✓	✓
Education level dummies (mother × father)				✓	✓
Income (mother, father, interact)					✓
Observations (min. over course type)	316,039	315,880	315,880	315,879	302,468
Children of lesbian couples	134	134	134	134	133
Children of gay couples	4	4	4	4	4

Note: Separate cross sectional regressions of test scores by course on couple type, including controls as indicated. Sample consists of all children born 2001-2007 in the main sample described in Section 4, before conditioning on the first child or the age of the parents at first birth. Standard errors in parentheses are clustered at both parents using two-way clustering. Test scores are normalized within course and year to have mean zero and standard deviation 1. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Singleton observations are dropped.

6 The impact of family friendly policies

The results thus far suggest that the relative child penalty experienced by heterosexual couples is primarily driven by female preferences and gender norms, and that the alternative shared parenting approach taken by same sex couples increases household income and may improve child outcomes. Despite the persistence of the relative child penalty within heterosexual couples, history suggests that decreases in the relative child penalty are possible. In Figure 10 we graph the child penalty of women and men in heterosexual couples from 1971-2010 using the long sample of heterosexual couples (see Section 4.1). Note that each line represents the child penalty for children born during a five year interval, estimated using the event study approach from the previous section (see equations (5) and (4)). The figure shows that the child penalty for women has declined substantially over time. In the 1970's and 1980's, fathers not only didn't experience a child penalty, but actually obtained an increase in income on the birth of their first child. However, over time this child premium for fathers has decreased, and currently fathers largely experience no change in income following the birth of their first child. Combining the two graphs, while the reduction in the relative child penalty has been substantial from the 1970's until today, the remaining gap is still large, and largely driven by the penalties experienced by mothers. In the remainder of this paper we estimate the impact of two important policy tools aimed partly at decreasing this gap: Paid paternity leave which targets fathers and access to high quality childcare.

6.1 Paternity leave

As means for increasing fathers' involvement in raising children, the so called daddy quotas of the Scandinavian countries have attracted considerable interest. Starting as early as 1993, Norway mandated a four week period of parental leave for fathers. If not taken, this leave period could not be transferred to the mother. A number of other countries have introduced similar quotas, including Ireland (14 weeks), Slovenia and Iceland (13 weeks), Germany (8 weeks), Finland (7 weeks), and Portugal (6 weeks), see OECD (2014). Paternity leave, by forcing fathers to spend more time with their children, might increase the value fathers place on time with chil-

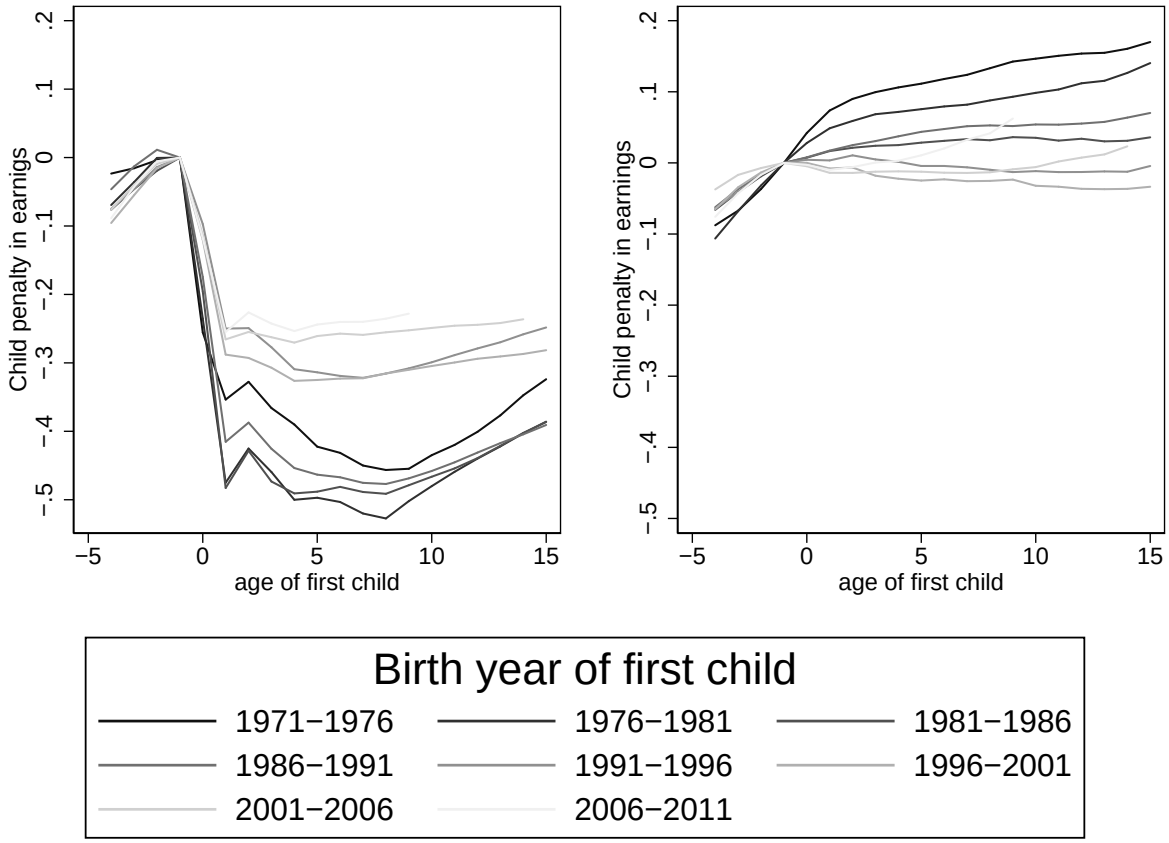


Figure 10: The child penalty in income over time for mothers and fathers in heterosexual couples

Note: Child penalties estimated separately by birth cohort of first child in 5-year intervals. Estimated using the event study framework from equations 5 and 4. Separate plots by birth cohort, including confidence intervals, is found in Figure figure A7 in the appendix.

dren (increasing β in equation (1)) and might also decrease the distaste fathers have for mothers working outside the home (reducing α in equation (2)). Paternity leave could also increase the productivity of fathers in home production (increasing k in equation (2)). Within the framework of our model, all of these effects could decrease the relative child penalty. In Table 3 we report every leave reform in Norway from 1992 - 2014. The maternal and paternal quota columns report the amount of parental leave in weeks that is reserved exclusively for the mother and father. The remaining leave can be shared among parents however they choose and is reported in column 6. The reforms were generally announced in October the year before implementation as part of the budgeting process, making it nearly impossible to plan conception in response to the announcement of the quota change in order to manipulate birth dates around the cutoff in April or July. In Figure B1 in the Appendix, we verify that there is no statistically significant change in the density of births around the cutoff for each reform.

In this paper, we exploit the 2005, 2006, 2009, 2011, 2013 and 2014 reforms using a regression discontinuity design. As in all regression discontinuity designs, identification relies on continuity in the underlying regression functions at the cutoff. Our identification strategy exploits the fact that parents of children born just before the reforms were not subject to the changes in parental leave quotas, while parents of children born right after each reform were subject to the changes. For this exercise, we draw on heterosexual couples from the main sample with first children born in 2005, 2006, 2009, 2011, 2013 and 2014. Because we want to capture mothers and fathers exposed to the leave reforms, we include in the sample only couples where the mother took some leave, indicating that she is eligible, because users of the alternative one-time benefit would not be affected. We do not believe there is reason to think that the extensive margin of maternity leave use is affected by the reforms, because these changed the maternity leave quota at very high margins of leave. Furthermore, the one-time benefit is so small that it is highly unlikely that any eligible mother would choose this over maternity leave. For fathers, we set leave to zero for fathers where we observe no leave take-up. Appendix B.1 provides additional details on the construction of our parental leave measure. We further restrict the sample to births in a window around the reform date using the optimal bandwidth, see below. We begin by estimating

Table 3: Parental leave reforms in Norway

Reform date	Leave in weeks	Compensation	Maternal quota in weeks	Paternal quota	Shared leave	Max weeks mother
April 1st, 1992	35 (44.3)	100% (80%)	8 (2 before birth)	0	27 (36.3)	35 (44.3)
April 1st, 1993	42 (52)	100% (80%)	9 (3 before birth)	4	29 (39)	38 (48)
July 1st, 2005	43 (53)	100% (80%)	9 (3 before birth)	5	29 (39)	38 (48)
July 1st, 2006	44 (54)	100% (80%)	9 (3 before birth)	6	29 (39)	38 (48)
July 1st, 2009	46 (56)	100% (80%)	9 (3 before birth)	10	27 (37)	36 (46)
July 1st, 2011	47 (57)	100% (80%)	9 (3 before birth)	12	26 (36)	35 (45)
July 1st, 2013	49 (59)	100% (80%)	17 (3 before birth)	14	18 (28)	35 (45)
July 1st, 2014	49 (59)	100% (80%)	13 (3 before birth)	10	26 (36)	39 (49)

Source: NOU 2017:6 (2017)

the impact of each reform separately. We estimate a fuzzy RD separately for both mothers and fathers and for each year using the following specification:

$$\begin{aligned} y_{it} &= \beta_t L_i + f_t(x_i) + \epsilon_{it} \\ L_i &= \gamma \mathbb{1}(x_i \geq 0) + g(x_i) + \eta_{it} \end{aligned} \tag{6}$$

Where x_i, the running variable, is the number of days after the reform date that the child was born. For the 2014 reform, which decreased the leave quota, the running variable is instead coded as days before the reform. $f_t(x_i)$ and $g(x_i)$ are local linear polynomials that are separate on either side of the cutoff. We use the optimal bandwidth that minimizes the mean squared error of the RD estimate to define the sample, and a triangular weighting function in order to obtain estimates local to the cutoff. Because the reforms differ in the quota change implemented, we scale the first stage and reduced form estimates to represent the impact of one additional week of quota so that the reforms are comparable. We estimate and report robust bias-corrected confidence intervals (Calonico *et al.*, 2014) together with the conventional, heteroskedasticity-robust confidence intervals.[28] For details, see Cattaneo *et al.* (2018a,b).

[28]Many models in this section are estimated using the robust RD commands for Stata written by Matias D. Catta-

The critical assumption for the validity of our RD approach is that the underlying regression functions are continuous at the threshold. This implies that the population of couples around the discontinuity are identical. In Table B1 in the appendix we report estimates that show that on observables, individuals around the cutoff are statistically indistinguishable from each other with only a few exceptions. The exception is maternity leave use, which is no surprise given the reform details in Table 3. For roughly the reforms where the total length of the maternity quota and the shared leave was reduced, we see reductions in maternity leave use. Because these reductions come from very high margins of maternity leave use, we think it is more likely that any impact on long-term labor supply is driven by the much larger effects on paternity leave use than these effects on maternity leave take-up. In Appendix B.3, we exploit the fact that some of these reforms expanded the paternity leave use at the expense of the maternity leave, while others expanded the total leave length. This allows us to instrument for both the maternity and paternity leave use, confirming the baseline results of the effects of paternity leave.

Furthermore, if parents were able to manipulate either conception or birth at the cutoff in order to qualify for reforms, then we would expect a statistically significant change in the density of births around the cutoff. In Figure B1 in the appendix, we show graphically that this is not the case. The p-values reported in each panel are for a test of equal densities on either side of the cutoff, using methods from Cattaneo *et al.* (2017, 2018c). None of the tests reject that the densities are the same.

We report first stage estimates for these specifications in Table 4, separately for each reform. Results have been scaled to reflect one week of quota expansion. We see clear and significant effects of all reforms , whether using robust bias-correcting inference or conventional inference that only accounts for heteroskedasticity. In the appendix, we plot the reduced form and first stage estimates together for each of the six reforms in Figure B2, scaling by the quota increase to get estimates that reflect one additional week of quota. Despite the strong first stages, the reduced form estimates are relatively flat for both mothers and fathers and we find no significant differences between couples where the father is exogenously exposed to greater paternity leave

neo and coauthors, whom we owe thanks. These include rdrobust, rddensity, rdbwselect and others. These packages are documented in Calonico *et al.* (2018) and Cattaneo *et al.* (2018c)

Table 4: RDD first stage estimates

Reform year	2005	2006	2009	2011	2013	2014	Pooled	Stacked
RD estimate per week	0.83***	1.09***	0.96***	0.88***	0.74**	0.78***	1.26***	0.88***
conventional standard error	(0.34)	(0.33)	(0.094)	(0.28)	(0.24)	(0.11)	(0.13)	(0.066)
robust standard error	0.42	0.40	0.11	0.33	0.28	0.13	0.14	
conventional p-value	0.015	0.001	0.000	0.002	0.002	0.000	0.000	0.000
robust p-value	0.024	0.006	0.000	0.004	0.010	0.000	0.000	
Observations	14,658	15,138	16,556	16,558	16,268	14,330	93,508	93,508
Optimal bandwidth	60.7	55.2	74.4	45.3	63.0	43.3	35.3	
Efficient observations	5006	4,901	6,993	4,467	6,120	3,815	19,751	31,302
Weights in pooled	0.15	0.16	0.17	0.18	0.18	0.16		
Weights in stacked	0.16	0.16	0.22	0.14	0.20	0.12		
Quota increase	1	1	4	2	2	-4		

Notes: Robust semiparametric sharp RD estimates of the effect of paternity leave reforms on paternity leave takeout using optimal bandwidths, triangular kernel and local linear polynomials on either side of the cutoff. All estimates are scaled to reflect one week of quota increase. Stacked estimates are the stacked individual cutoffs, allowing polynomials to vary over cutoffs and using the cutoff-specific bandwidths and weights. Conventional standard errors are heteroskedasticity-robust, but not bias-corrected. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, using conventional, heteroskedasticity-robust standard errors.

and couples who are not. These results suggest that paternity leave does not cause fathers to parent more equally with mothers, at least not in such a way that mothers experience less severe child penalties. Estimates are, however, somewhat imprecisely estimated.

In order to move beyond these separate reforms and increase the precision of our estimates, we next stack all the reforms from above. The common way of doing this in RD studies is to re-center the running variable to be zero at the relevant cutoff for all individuals and run semi-parametric RD estimates in the pooled sample. We call this the pooled estimate, and report the first stage specification for this procedure in Table 4. This estimate, however, restricts the functional form of the local linear polynomials to be the same for all cutoffs, potentially increasing the approximation error and lowering precision. An alternative and more straightforward way to stack the estimates is to allow the local polynomials of the running variable to vary by cutoff and use the cutoff-specific optimal bandwidths and kernel weights from the individual specifications. Scaling is secured by using an indicator of the number of weeks of quota increase rather than a dummy at the cutoff. Unfortunately, we cannot calculate bias-corrected standard errors for this specification, but we argue that the problem should be relatively minor. First, notice that the difference between the conventional and the robust standard error estimate for the pooled specification is very small, indicating that the variance coming from the approximation error is relatively minor. Second, the approximation error should be smaller for the stacked than the pooled specification because we allow the local polynomials to differ between cutoffs and thus approximate the unknown functions better. Nonetheless, inference from this specification is only correct if the model is well specified, so that approximation error vanishes asymptotically.

The last two columns of Table 4 report the first stage results from these specifications. The pooled estimate is - somewhat surprisingly - larger than most of the cutoff-specific estimates, indicating more than a one week increase in leave use per week increase in the paternity leave quota.[29] Second, although the estimate is highly significant, notice that the standard error of the pooled estimate is still larger than the most precisely estimated individual reform. In contrast, the stacked specification delivers improved precision over any of the individual estimates, finding a

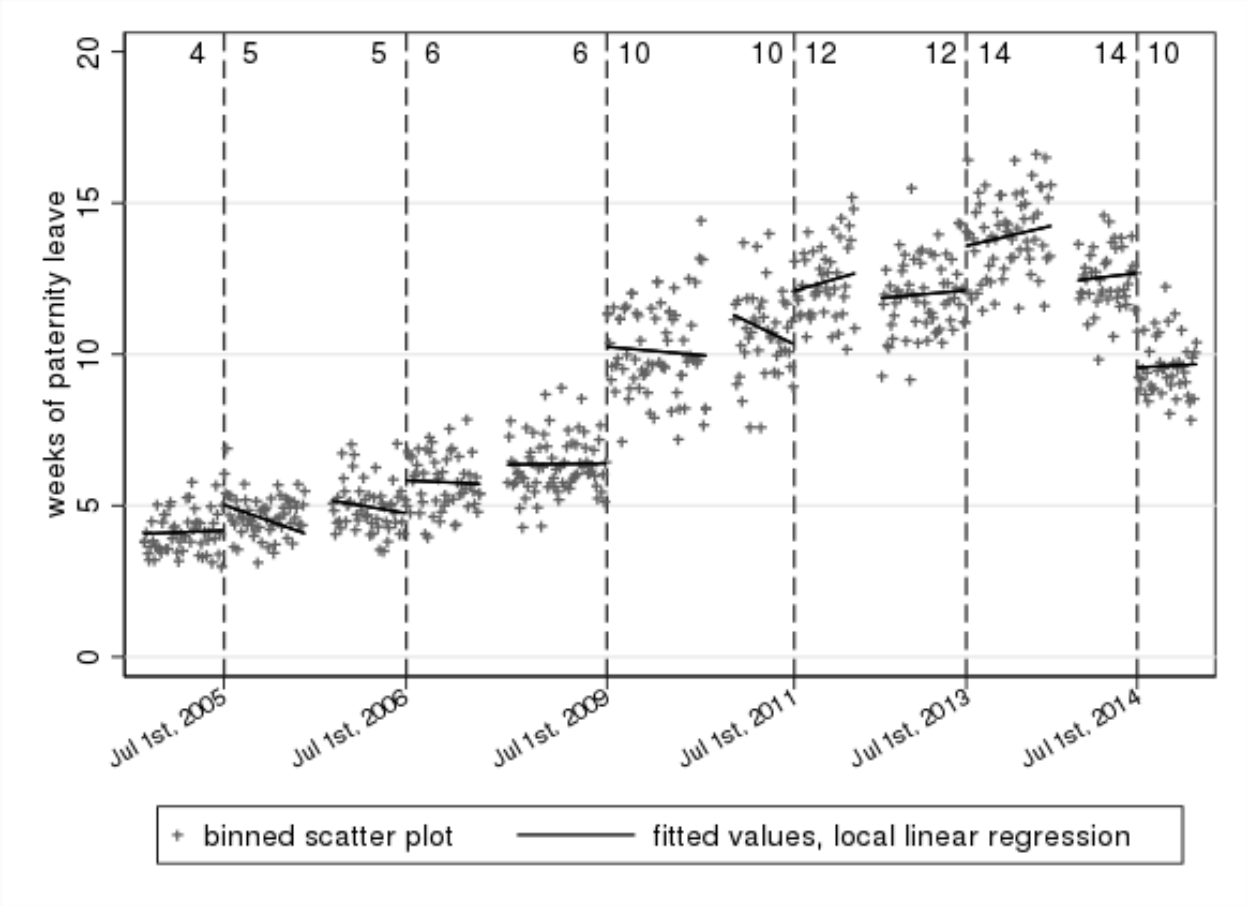
[29]For this specification, we scale the estimated parameter with the average number of weeks of quota increase in the sample.

more reasonable .88 weeks increase in leave use per week of quota increase.

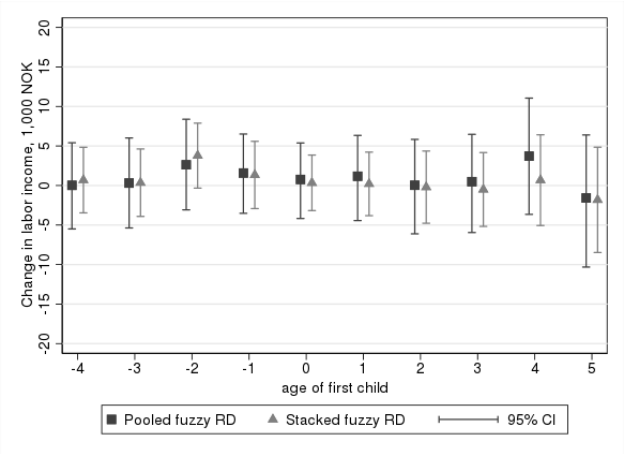
Informed by this, we move to estimate fuzzy RD specifications of the impact of paternity leave use on mothers' and fathers' labor supply using the pooled and stacked models described above. For the stacked estimates, we revert to the cutoff-specific treatment indicators as instruments because the fuzzy RD takes care of the scaling. This specification exactly reproduces the cutoff-specific first stage estimates reported in Table 4 and so is a natural way to stack the reforms. When interpreting these fuzzy RD estimates, it is important to keep in mind that these estimates are local average treatment effects: they capture the effects of additional leave use on earnings for people induced to use more leave because they were exposed to the reform. In our case, the compliers represent *unwilling users* of paternity leave, because these couples were free to distribute more leave than the quota to the father irrespective of the reform. In case of heterogeneous treatment effects, the average effect for the compliers need not be the same as the average effect in the population. Despite this, we argue that the LATE is a particularly policy relevant treatment effect in our case, because it reflects the effects of paternity leave use for fathers induced to take more leave by the policy instrument, which is arguably the population of interest to policy makers.

The results from the stacked and pooled fuzzy RD estimates for mothers and fathers are presented together with the combined first stages in Figure 11. The top panel illustrates how the various reforms affected paternity leave takeout, mirroring the estimates from table 4 and showing clear discontinuities at the cutoffs. The bottom two figures presents the impacts on mothers' and fathers' yearly incomes over time. Notice first that there is no effect of paternity leave use on pre-birth outcomes. This is a reassuring result, which can be interpreted as a balancing exercise or placebo test. Following birth, we see no impact of paternity leave use at years 0 and 1 when most of the leave takeout happens. Neither do we see any impact in the following years; the estimates are flat and centered at zero, confirming the findings from before of little impact of paternity leave use on child penalties. The stacked specification does, however, provide more precise estimates than the results using separate reforms, ruling out positive impacts larger than around NOK 5,000 on mother's earnings per week of paternity leave use for all years post-birth.

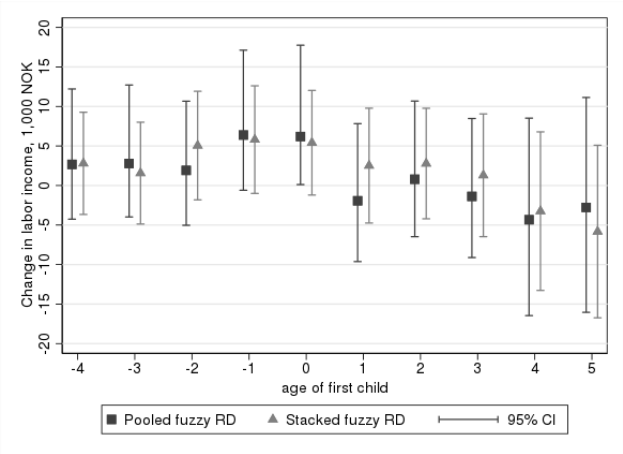
To provide a sense of the potential percentage change in the child penalty, we re-scale the



(a) First stage estimates



(b) Mothers



(c) Fathers

Figure 11: Main RD estimates of paternity leave use

Note: Top panel shows first stage estimates around each reform date, using local linear polynomials, triangular weights and optimal bandwidths. Bottom panels show fuzzy RD estimates of the impact of an additional week of paternity leave use on earnings over time, using all six reforms. Pooled estimate refers to the simple reentered estimate, while the stacked estimate stacks the cutoff-specific specifications. Robust bias-correcting inference reported for the pooled estimate, conventional, heteroskedasticity-robust inference for the stacked estimate.

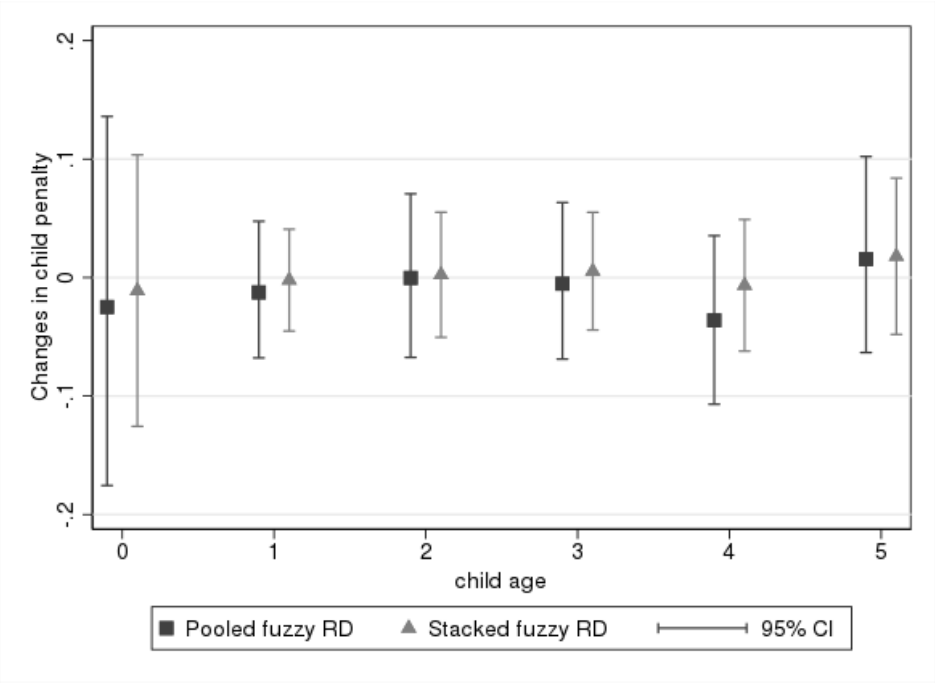


Figure 12: Scaled stacked and pooled fuzzy RD estimates, mothers

Note: Stacked fuzzy RD estimates for mothers, scaled by the estimated child penalties from the baseline so that the estimates can be interpreted as the relative increase in the child penalty per week increase of paternity leave use.

figures so that the y-axis represents the percentage of the child penalty, as estimated from the event studies from the first half of the paper. We present these results in Figure 12. While point estimates remain close to zero, the lower bound of the effect is still informative. We can rule out reductions larger than around 5 - 7% of the female child penalty per week of paternity leave use for ages 1 through 5. Similar estimates for fathers are too imprecise to draw firm conclusions, in part because the initial child penalty for fathers is very close to zero.

Results so far provide clear evidence that paternity leave use does not have a causal effect on the distribution of market work within the couples induced to use paternity leave by the change in quota. Paternity leave might, however, influence gender norms or preferences around the distribution of home work in ways that do not influence labor market earnings. One possible measure of such norms is the use of paternity leave itself for future children. To investigate whether paternity leave use has a direct effect on the father's choice of spending time with children, we exploit the fact that many of the fathers that havea child around the time of the reforms subsequently go on to have more children. We therefore estimate our fuzzy RD model using as an outcome the father's leave takeout for subsequent children for all children born up until and including 2014 in a setup similar to the peer effects estimates from Dahl *et al.* (2014).[30] We cluster standard errors on the father to account for the fact that each father may have multiple treated kids and may get multiple subsequent kids for which we can measure outcomes. Notice that we cannot use the 2014 reform for this exercise, as we cannot reliably measure paternity leave use for kids born after 2014.

Table 5 provides the results of this exercise, both for each reform separately and the pooled and stacked estimates for all reforms. Across the rows of Table 5, we see little evidence of any permanent impact on norms as measured by takeout of paternity leave for later kids: Except for the 2013 reform, where the efficient sample size is only 150 children and we find a marginally significant effect, none ofof the reforms provide any statistically significant results, and point estimates are negative. Focusing on our preferred stacked estimates, the results indicate non-

[30] Notice that if fertility is endogenous to the parental leave reforms, this might constitute an endogenous sample selection criteria. Hart *et al.* (2019) investigates fertility response to the 2009 reform, finding no evidence of such effects.

Table 5: Paternity leave norms: Fuzzy RD of paternity leave use on leave for subsequent kids

	2005	2006	2009	2011	2013	Pooled	Stacked
RD estimate	-0.15	-0.25	-0.064	-0.47	0.71*	-0.21	-0.11
conventional standard error	(1.49)	(0.38)	(0.15)	(0.38)	(0.34)	(0.16)	(0.13)
robust standard error	1.81	0.45	0.18	0.45	0.40	0.19	
conventional p-value	0.92	0.51	0.67	0.21	0.037	0.20	0.39
robust p-value	0.95	0.57	0.72	0.15	0.039	0.18	
Observations	13,476	13,019	11,531	7,918	746	46,690	46,690
Optimal bandwidth	58.0	66.0	65.2	63.9	42.3	70.3	
Efficient observations	4,455	4,955	4,288	3,052	150	18,896	16,900

Notes: Fuzzy RD estimates of the impact of one more week of paternity leave for a child on the weeks of paternity leave use for subsequent children. Standard errors are clustered by father. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

significant effect of .11 *less* weeks of leave for subsequent children for each week of leave for the first child. This relatively precise estimate allows us to rule out effects larger than about .14 more weeks of leave. An important caveat for these results is that some fathers might be constrained to corner solutions also for later kids. If paternity leave use affects preferences for future paternity leave in a way that would make fathers prefer to take more leave, but still not prefer to take more than the quota, this RD specification would not be able to detect the effect. Nonetheless, we conclude that there is little evidence for paternity leave quotas to permanently affect fathers preferences for staying home with children as measured by their leave taking behavior.

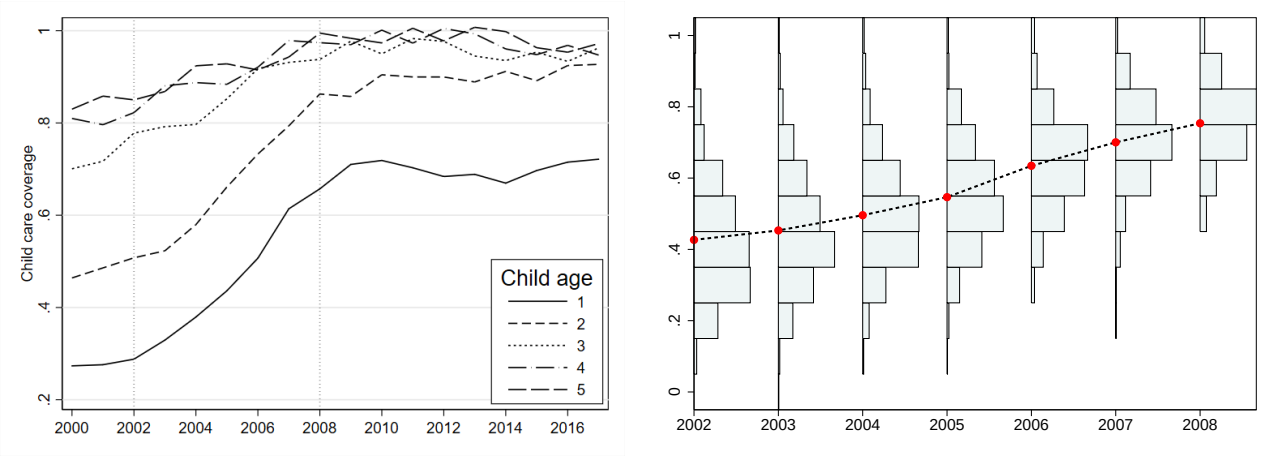
The results in this section cover a variety of different paternity leave expansions. Despite the number of reforms we study and the strength of the first stages, we never find a statistically significant impact of paternity leave on income, child penalties or leave use for subsequent children. Based on these results, we conclude that paternity leave does not appear to reduce the relative child penalty.

6.2 Improved access to early child care

An alternative approach to reduce the relative child penalty is for the government to provide a high quality substitute for mother's time. Figure 13a shows the child care coverage rates over time in Norway, separately by age of the children. These figures show that the formal care sector for preschoolers was well developed in Norway by the early 2000's, with more than 80% of Norwegian 3 - 5 year olds attending care.[31] For toddlers, however, coverage was much lower at less than 50% and 30%, respectively, and the market was strongly rationed. These facts are documented in greater detail in Andresen and Havnes (2019), including additional evidence from surveys on the actual and preferred modes of child care for children at these ages. The underrepresentation of children between ages 1 and 2 in formal care was the impetus for the Child Care Concord in 2002, a broad, bipartisan agreement to increase the availability of care for toddlers. Following this reform, coverage increased rapidly for 1-2 year olds over the next years as shown in Figure 13a. However, the expansion varied considerably between municipalities and over time, as shown in Figure 13b. This is the variation exploited to estimate the effects of formal care use on parents' labor supply in Andresen and Havnes (2019). We use the same variation to estimate the impact of increasing access to high quality formal child care on the child penalties experienced by mothers and fathers in this section. Andresen and Havnes (2019) shows that the exact timing of expansion was subject to a range of constraints that were hard to predict, and the timing of expansion was not necessarily easy to predict even for the municipalities themselves. Furthermore, the exact timing of expansion only to a minor extent seems to be predictable by pre-reform characteristics of the municipalities, as documented in Figure 14, making the expansion of care availability a potential instrument for the endogenous choice of how much child care to use. For this application, we start with all children from the main sample born in the years 2000-2006. This restricts the sample to children who are two years old around the time of the reform-induced expansions of care.[32] We assign children to their municipality of residence at the age of 1 and look at couples where both parents reside in that municipality when the child is 1. While much of the literature restricts the sample to

[31]The prevalence of care is the result of a reform and gradual expansion of formal care for these children in the 1970's (Havnes and Mogstad, 2011).

[32]This includes a few thousand twins. Clustering at the municipality level accounts for within-family clustering.

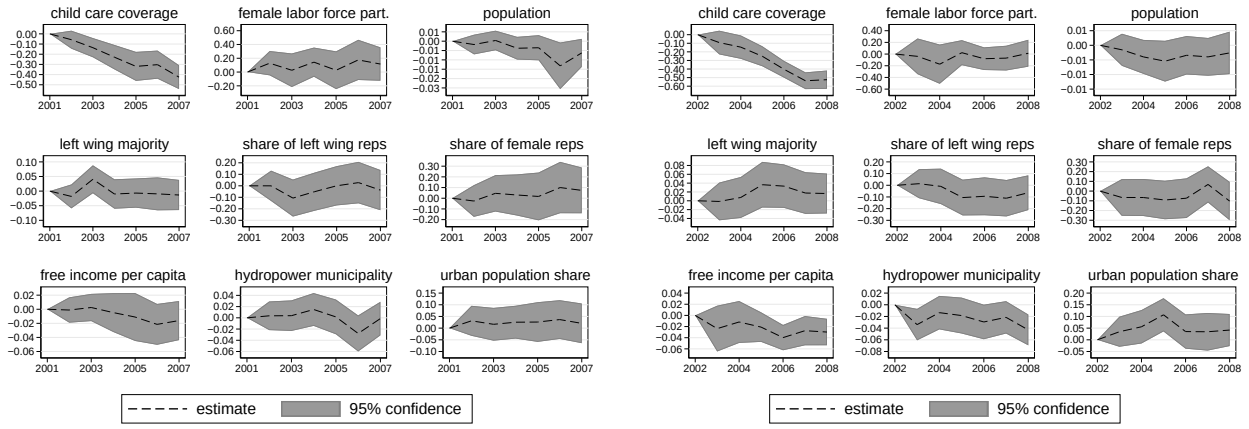


(a) Child care coverage rates

(b) Distribution of municipal coverage rates for 1-2 year olds

Figure 13: Child Care Coverage

Source: Statistics Norway Statistikkbanken, tables 09169 and 07459.



(a) Child care coverage for 1-year olds

(b) Child care coverage for 2-year olds

Figure 14: Predicting expansion of slots from pre-reform characteristics

Note: Results from regression of our two instruments, child care coverage at age 1 and 2, on municipality- and year fixed effects and an interaction of pre-reform characteristics interacted with year dummies, in a sample of municipalities over time. Plotted are the year-specific impact of the pre-reform characteristics on expansion of care in a particular year. 95% confidence intervals in grey, clustered at the municipality level.

children without younger siblings, we view future fertility as a potentially endogenous outcome of the reform, and therefore do not restrict the sample to youngest children. To be consistent with the results we have presented thus far, we look at the effects on maternal and paternal income when the child is between the ages 0 through 5, and use the years before birth as placebo outcomes. This leaves us with a sample of 116,480 couples.[33]

For this sample, we take our baseline event study specification separately for mothers and fathers and separately at each event time and see how adding the measure of individual child care use at ages 13 to 36 months affects the child penalty. Because child care is endogenous to labor supply, we instrument care use with the expansion of slots for 1-year olds at age 1 and for 2-year olds at age 2 in the following IV model:

$$
\begin{aligned}
y_{it} &= \pi_k + \gamma_{T_{it}} + \beta_{a_{it}} + \phi m_i + \epsilon_{it} \\
m_i &= \tilde{\pi}_k + \tilde{\gamma}_{T_{it}} + \tilde{\beta}_{a_{it}} + \phi_1 CC_k^1 + \phi_2 CC_k^2 + \tilde{\epsilon}_{it}
\end{aligned} \tag{7}
$$

Where $\gamma_{T_{it}}$ are calendar year fixed effects, π_k are municipality fixed effects, $\beta_{a_{it}}$ are age fixed effects for the parent (in years) and m_i is our measure of child care use from ages 13 - 36 months from the cash for care data. The instruments are CC_k^1, the share of slots for 1-year olds in the municipality at age 1 to the population of one year olds, and CC_k^2, the same share for 2-year olds, measured at the relevant age of the child.

The variation we exploit thus comes from the variation in expansion of care across municipalities and over time. As long as the exact timing of expansion of care is uncorrelated with other drivers of parents' labor supply, our approach recovers the causal effect of an extra year of early child care on labor supply for the compliers: the mothers who take up the newly expanded slots. Because child care was strongly rationed before the reform, it is natural to think of the compliers

[33]Notice that because we restrict to first born children, the sample size in this paper is a little less than half the size of the samples of cohabiting mothers and fathers in Andresen and Havnes (2019). This gives us less precision but is consistent with the rest of the paper. Because of the inherent focus on labor supply over time, we also measure child care use throughout the full 13 - 36 months period we can measure, in contrast to the preceding paper that is mostly concerned with child care use and labor supply during the calendar year the child turns 2.

as the mothers of children who wanted child care before the reform, but were restricted by the low supply. Figure 14 provides some support for the idea that expansions did not systematically vary across municipalities with different pre-reform characteristics (except, of course, the initial coverage rate), while Andresen and Havnes (2019) provide a range of specification checks that demonstrate the robustness of the instrument.

First stage estimates from this specification are presented in Table 6, column 1, where we see that the availability of slots in care has a strong influence on years of early care use. Expansions of care both at age 1 and at age 2 have a strong impact on child care use between 13 and 36 months, with an additional slot in care at age 1 increasing care use by around 0.8 years and at age 2 by about 0.6 years. Because our endogenous variable captures the intensity of use throughout the full period, these coefficients are not 1; as additional slots are generally opened in August, children may not have the chance to exploit them to capacity the whole year. The IV strategy thus scales the reduced form estimates to reflect a full year of early child care use. The F-statistic is around 200, indicating a very strong first stage.

The second stage estimates from the baseline specification are presented in Figure 15a and 15b, where the baseline model discussed so far is indicated with diamonds. Focusing first on the years of treatment, ages 1-3, we see that the estimates increase in this period up to point estimates of around 21,000 NOK at ages 2 and 3, where most of the treatment happens, only to return to zero the last two years of the panel.[34] Estimates are significant at the 5% level at age 3 and 10% level at age 2, and thus indicate that there is some immediate effects of child care use on earnings during the years of treatment, perhaps driven by allowing mothers to return to work earlier after child birth. Results for fathers are noisy, but point, if anything, to negative impacts on earnings, which could also reduce relative child penalties. The pre-birth outcomes, which we can think of as placebo outcomes, indicate small and insignificant impacts of future child care use on past earnings, supporting the estimation strategy.

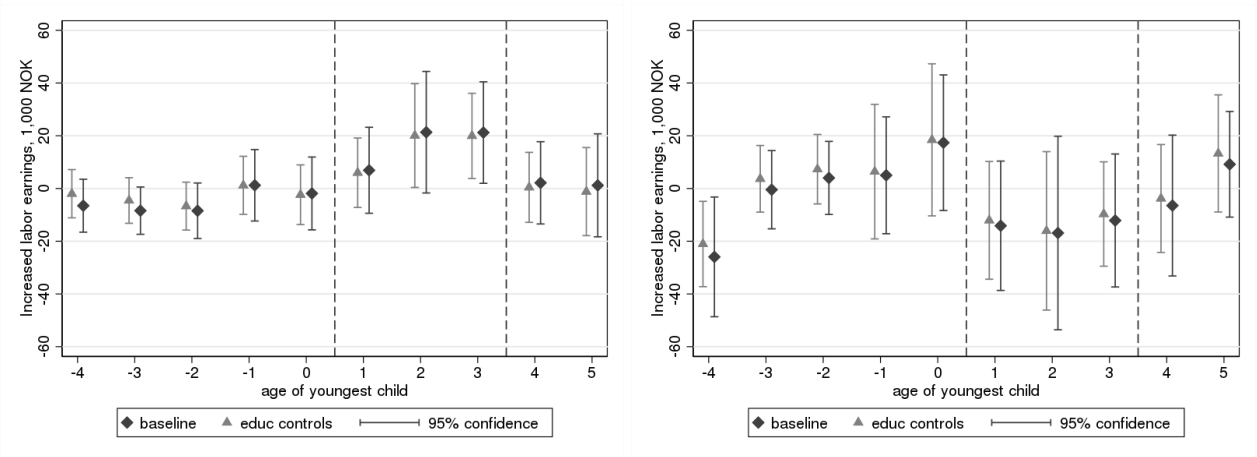
As a robustness check, we include the education level-specific age profiles in equation 7. The

[34]This estimate is smaller than the baseline estimate in Andresen and Havnes (2019), but a number of differences in the sample and specification may explain this, as well as the lower level of precision in our study due to a sample size than half the size because of the focus on first born children only.

Table 6: First stage estimates, formal care use

Years of child care use at ages 13 - 36 months				
Coverage rate at age 1	0.891***	(0.0503)	0.879***	(0.0516)
Coverage rate at age 2	0.589***	(0.0516)	0.588***	(0.0500)
Municipality fixed effects	✓		✓	
Year fixed effects	✓		✓	
Age profiles	✓			
Education-specific age profiles			✓	
N	116,478		116,461	
mean dep. var	1.034		1.034	
F	215.5		198.3	

Note: First stage estimates of eq. 7 for mothers. Point estimates for fathers (not shown) are very similar. Standard errors in parentheses, clustered at municipality.



(a) Mothers

(b) Fathers

Figure 15: Impact of a year of child care use at ages 13-36 months on income

Note: IV results from equation 7 reflecting the impact on labor earnings in 1,000 NOK across child age for an extra year of early child care use at ages 13-36 months.

first stage from this specification is hardly affected by this, as documented by column 2 in table 6. The second stage results are also very similar.

The peak impacts on maternal labor supply are in the range of 21,000 NOK for ages 2-3. Like in the paternity leave application, it is natural to ask how big this impact is in light of the estimated child penalty from Section 5. In Figure 16, we therefore scale the IV estimates for mothers with the estimated baseline child penalties to present the relative effect of a full year of child care use on the child penalty. Results show that the child penalty is reduced by a little more than 25% for mothers when their children are between the ages 2-3. We conclude that early child care shows more promise as a policy tool for reducing child penalties than paternity leave, although it does not appear to have a permanent impact.

6.3 Conclusion

In the first half of this paper we show that same sex couples experience a very different child penalty compared to heterosexual couples. Based on our household model, we conclude that while biology plays a small role in explaining the relative child penalty experienced by heterosexual couples, some combination of preferences and gender norms explain the vast majority of the relative child penalty experienced by heterosexual couples. Moreover, the large child penalty experienced by heterosexual mothers translates into a significant household income penalty for heterosexual couples that persists over time. In contrast, while lesbian couples experience the same sized household income penalty initially (albeit shared more evenly between the two partners), the overall household income penalty decreases over time until five years after birth lesbian couples no longer experience a household income penalty from having children. This is despite the fact that lesbian couples have a similar number of children compared to heterosexual couples and our descriptive evidence suggests that children of lesbian parents outperform children of heterosexual couples in test scores at age 10.

In the second half of the paper we examined two possible policy responses to the relative child penalty. First, government policy might aim to address the behavior of fathers, and the most commonly proposed such policy is paternity leave. Second, governments might provide a

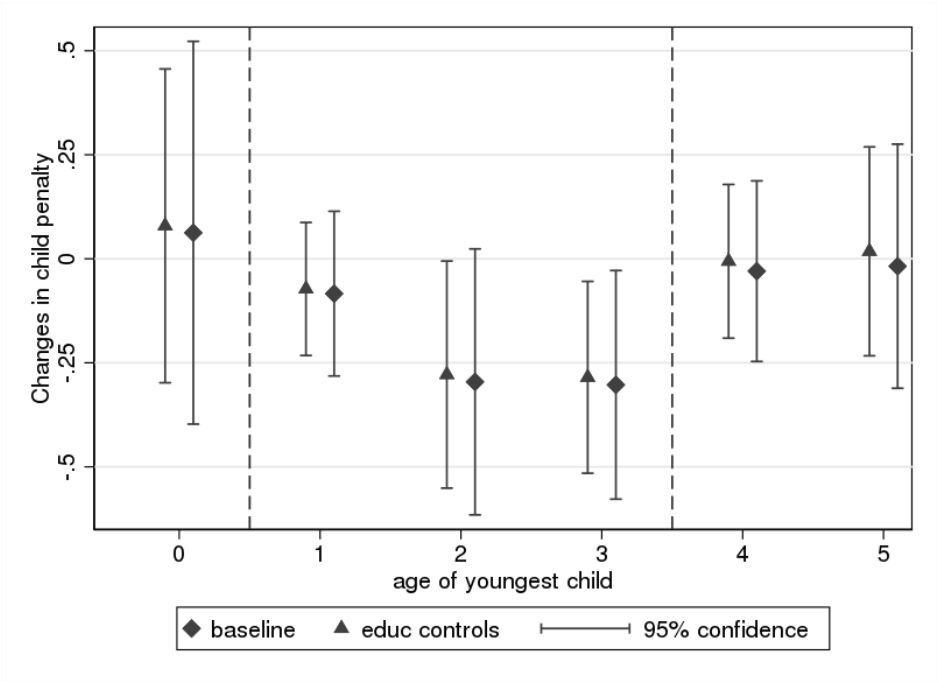


Figure 16: Effects of a year of early child care use on the child penalty

high quality child care substitute for households to utilize in place of the mother's time. Using a series of adjustments to paternity leave in Norway and a regression discontinuity framework, we show that while fathers take paternity leave (the first stage is strong), paternity leave has no impact on the relative child penalty. In addition, paternity leave has no impact on whether the father takes additional leave for future children. Next, we use an instrumental variables approach exploiting the staggered expansion of care following a Norwegian child care reform and show early child care use reduces the child penalty for mothers by around 25% per year of use in the years of treatment. These results suggest that if policy makers wish to decrease the relative child penalty, they should focus on providing better child care to families, not on offering paternity leave to fathers.

Our paper sheds light on both why the child penalty occurs and how policy might impact the relative child penalty. While we have focused on two of the most commonly proposed policies to reduce the child penalty, there are a number of additional policy changes that could impact the child care penalty differently and that would be productive avenues for future research.

References

Aarskaug Wiik, K., Seierstad, A. and Noack, T. (2014). Divorce in Norwegian same-sex marriages and registered partnerships: The role of children. *Journal of Marriage and Family*, **76** (5), 919–929.

Akgunduz, Y. E. and Plantenga, J. (2018). Child care prices and maternal employment: A meta-analysis. *Journal of Economic Surveys*, **32** (1), 118–133.

Alden, L., Edlund, L., Hammarstedt, M. and Mueller-Smith, M. (2015). Effect of registered partnership on labor earnings and fertility for same-sex couples: Evidence from Swedish register data. *Demography*, **52** (4), 1243–1268.

Andresen, M. E. and Havnes, T. (2019). *Child Care, Parental Labor Supply and Tax Revenue*. Working paper, available at `http://bit.ly/childcarenorway`.

Angelov, N., Johansson, P. and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, **34** (3), 545–579.

Antecol, H., Bedard, K. and Stearns, J. (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies? *American Economic Review*, **108** (9), 2420–41.

— and Steinberger, M. D. (2013). Labor Supply Differences Between Married Heterosexual Women And Partnered Lesbians: A Semi-Parametric Decomposition Approach. *Economic Inquiry*, **51** (1), 783–805.

Baker, M. and Milligan, K. (2015). Maternity leave and children's cognitive and behavioral development. *Journal of Population Economics*, **28** (2), 373–391.

Baumle, A. K. (2009). The cost of parenthood: Unraveling the effects of sexual orientation and gender on income. *Social Science Quarterly*, **90** (4), 983–1002.

Bergsvik, J., Kitterød, R. H. and Wiik, K. A. (2019). *Parenthood and couples' relative earnings in Norway 2005-2014*. Discussion paper, Statistics Norway, no. 873.

Black, D. A., Sanders, S. G. and Taylor, L. J. (2007). The economics of lesbian and gay families. *Journal of Economic Perspectives*, **21** (2), 53–70.

Blau, D. and Currie, J. (2006). *Pre-School, Day Care, and After-School Care: Who's Minding the Kids?*, Elsevier, *Handbook of the Economics of Education*, vol. 2, chap. 20, pp. 1163–1278.

Blau, F. D. and Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic perspectives*, **14** (4), 75–99.

Borusyak, K. and Jaravel, X. (2016). Revisiting event study designs. SSRN.

Bursztyn, L., Fujiwara, T. and Pallais, A. (2017). 'Acting Wife': Marriage market incentives and labor market investments. *American Economic Review*, **107** (11), 3288–3319.

Calonico, S., Cattaneo, M. D., Farrell, M. H. and Titiunik, R. (2018). RdrobUSt: Stata module to provide robust data-driven inference in the regression-discontinuity design.

—, — and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82** (6), 2295–2326.

Carneiro, P., Løken, K. V. and Salvanes, K. G. (2015). A flying start? Maternity leave benefits and long-run outcomes of children. *Journal of Political Economy*, **123** (2), 365–412.

Cattaneo, M. D., Idrobo, N. and Titiunik, R. (2018a). A practical introduction to regression discontinuity designs: Volume I.

—, — and — (2018b). A practical introduction to regression discontinuity designs: Volume II.

—, Jansson, M. and Ma, X. (2017). Simple local polynomial density estimators.

—, — and — (2018c). Manipulation testing based on density discontinuity. *Stata Journal*, **18** (1), 234–261.

Chung, Y., Downs, B., Sandler, D. H. and Sienkiewicz, R. (2017). *The Parental Gender Earnings Gap in the United States*. Working Papers 17-68, Center for Economic Studies, U.S. Census Bureau.

Cools, S., Fiva, J. H. and Kirkebøen, L. J. (2015). Causal effects of paternity leave on children and parents. *The Scandinavian Journal of Economics*, **117** (3), 801–828.

Dahl, G. B., Løken, K. V. and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, **104** (7), 2049–74.

Del Boca, D. and Flinn, C. (2012). Endogenous household interaction. *Journal of Econometrics*, **166** (1), 49–65.

Dobkin, C., Finkelstein, A., Kluender, R. and Notowidigdo, M. J. (2018). The Economic Consequences of Hospital Admissions. *American Economic Review*, **108** (2), 308–352.

Druedahl, J. and Martinello, A. (2016). *Long-Run Saving Dynamics: Evidence from Unexpected Inheritances*. Working Papers 2016:7, Lund University, Department of Economics.

Ekberg, J., Eriksson, R. and Friebel, G. (2013). Parental leave: A policy evaluation of the Swedish daddy month reform. *Journal of Public Economics*, **97**, 131 – 143.

Evertsson, M. and Boye, K. (2018). The transition to parenthood and the division of parental leave in different-sex and female same-sex couples in Sweden. *European Sociological Review*, **34** (5), 471–485.

Fadlon, I. and Nielsen, T. H. (2017). *Family Health Behaviors*. Working Paper 24042, National Bureau of Economic Research.

Fernández, R., Fogli, A. and Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, **119** (4), 1249–1299.

Goldberg, A. E., Smith, J. Z. and Perry-Jenkins, M. (2012). The division of labor in lesbian, gay, and heterosexual new adoptive parents. *Journal of Marriage and Family*, **74** (4), 812–828.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, **104** (4), 1091–1119.

Halrynjo, S. and Kitterød, R. H. (2016). Konsekvenser av arbeid-familietilpasning og velferd i Norge og Norden. en litteraturstudie.

Hart, R. K., Andersen, S. N. and Drange, N. (2019). *Effects of extended paternity leave on union stability and fertility*. Discussion paper, Statistics Norway, no. 899.

Havnes, T. and Mogstad, M. (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, **95** (11–12), 1455 – 1465, Special Issue: International Seminar for Public Economics on Normative Tax Theory.

Johansson, E.-A. (2010). *The Effect of Own and Spousal Parental Leave on Earnings*. Tech. Rep. 4, IFAU - Institute for Labour Market Policy Evaluation.

Kleven, H., Landais, C., Posch, J., Steinhauer, A. and Zweimuller, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*.

—, — and Søgaard, J. E. (2018). *Children and gender inequality: Evidence from Denmark*. Tech. rep., National Bureau of Economic Research.

Kotsadam, A. and Finseraas, H. (2011). The state intervenes in the battle of the sexes: Causal effects of paternity leave. *Social Science Research*, **40** (6), 1611 – 1622.

— and — (2013). Causal effects of parental leave on adolescents' household work. *Social Forces*, **92** (1), 329–351.

Kuziemko, I., Pan, J., Shen, J. and Washington, E. (2018). *The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?* Tech. rep., National Bureau of Economic Research.

Lalive, R., Schlosser, A., Steinhauer, A. and Zweimüller, J. (2014). Parental leave and mothers' careers: The relative importance of job protection and cash benefits. *Review of Economic Studies*, **81** (1), 219–265.

— and Zweimüller, J. (2009). How does parental leave affect fertility and return to work? Evidence from two natural experiments. *The Quarterly Journal of Economics*, **124** (3), 1363–1402.

Lundborg, P., Plug, E. and Rasmussen, A. W. (2017). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review*, **107** (6), 1611–37.

Martell, M. E. and Roncolato, L. (2016). The homosexual lifestyle: Time use in same-sex households. *Journal of Demographic Economics*, **82** (4), 365–398.

Moberg, Y. (2016). *Does the gender composition in couples matter for the division of labor after childbirth?* Working paper 2016:8, IFAU.

Morrissey, T. W. (2016). Child care and parent labor force participation: A review of the research literature. *Review of Economics of the Household*, pp. 1–24.

NOU 2017:6 (2017). Offentlig støtte til barnefamiliene.

OECD (2014). OECD family database: PF2.1 - key character of parental leave systems. available at `http://www.oecd.org/els/soc/PF2_1_Parental_leave_systems_1May2014.pdf`.

Olivetti, C. (2006). Changes in women's hours of market work: The role of returns to experience. *Review of Economic Dynamics*, **9** (4), 557–587.

— and Petrongolo, B. (2016). The evolution of gender gaps in industrialized countries. *Annual Review of Economics*, **8** (1), 405–434.

Patnaik, A. (2019). Reserving time for daddy: The consequences of fathers' quotas. *Journal of Labor Economics*, **0**, null, forthcoming.

Rege, M. and Solli, I. F. (2013). The impact of paternity leave on fathers' future earnings. *Demography*, **50** (6), 2255–2277.

Rosenbaum, P. (2019). *The Family Earnings Gap Revisited: A Household or a Labor Market Problem?* Tech. rep., SSRN.

Rudlende, L. and Lima, I. (2018). Medmødre tilpasser seg også fedrekvoten. *Arbeid og velferd*, (3).

Schneebaum, A. (2013). *Motherhood and the Lesbian Wage Premium.* Economics department working paper series, University of Massachussets, Amherst, number 2013-4.

A Additional results, child penalties

Table A1 summarizes the predictions of the theoretical model from Section 2, while simulations of various mechanisms are found in Figure A1.

Mean earnings over time for the three couple types are found in Figure A2 together with the simple event study estimates that omit age- and year fixed effects. Figure A3 allows the age-gender profiles to be different for parents with different education levels, without this substantially changing the baseline estimates

In Figure A4 we plot the mean number of children over time for each parent type, suggesting that fertility by 5 years later is very similar across heterosexual and lesbian couples. Figure A5 performs the baseline event study approach using only the couples where neither of the spouse has additional kids by age 5. Although this is an endogenous sample selection, so we should be careful in interpreting these estimates, results are relatively similar to the baseline estimates.

Figure A6 provides subsample analysis by (birthing) mother's education, revealing relatively similar effects across groups.

Finally, Figure A7 provides estimates of child penalties separately by time period, as measured in 5-year intervals from 1971 to 2010.

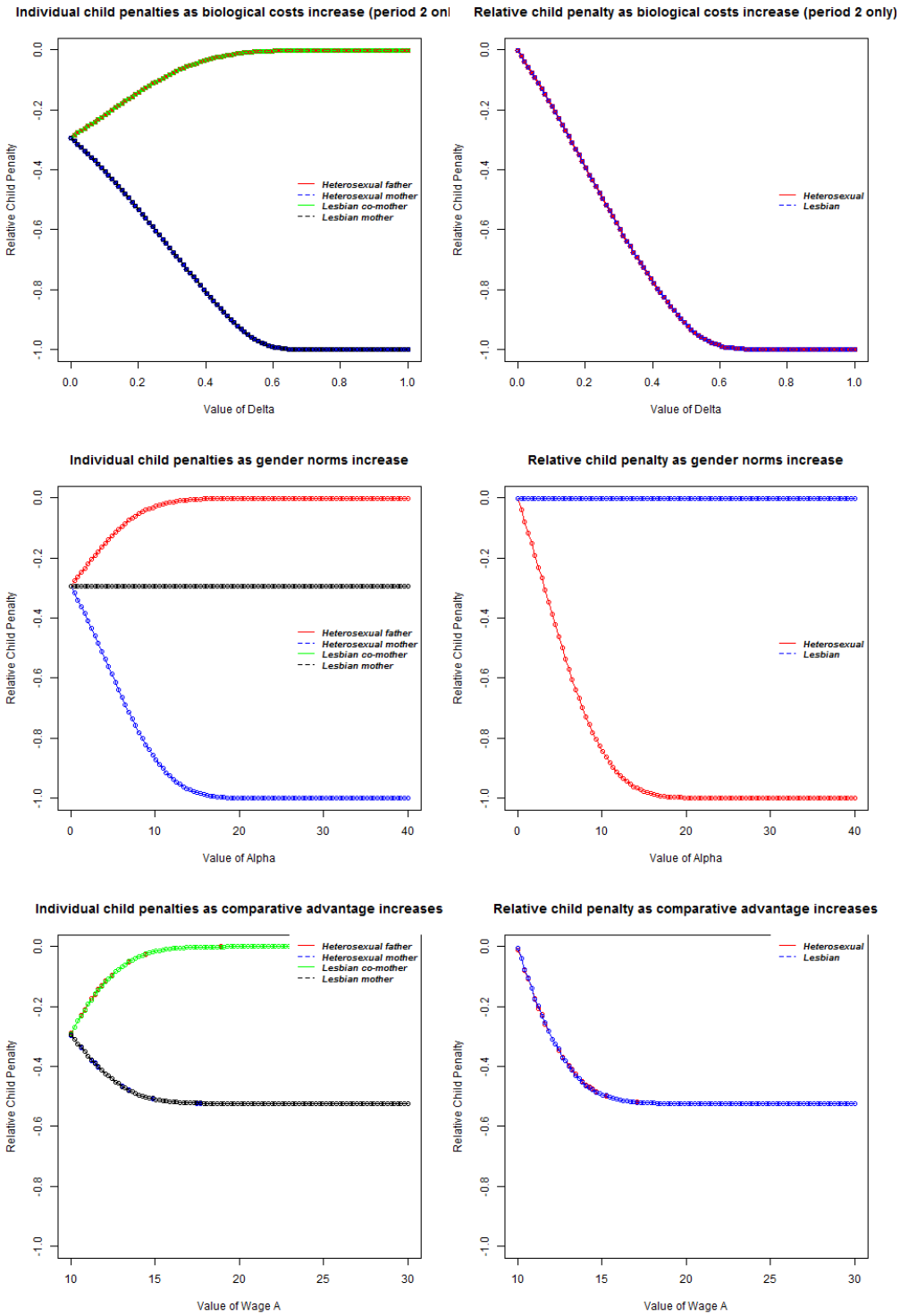
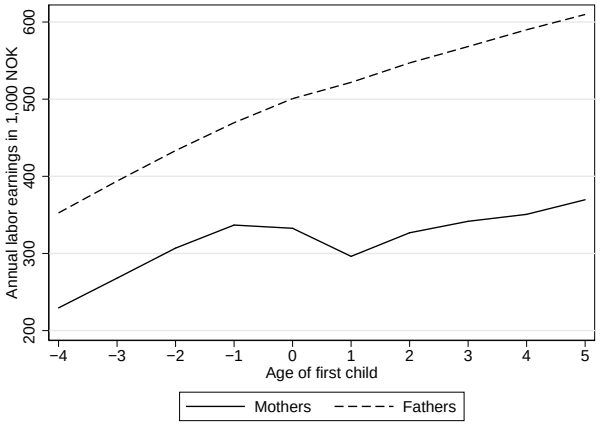
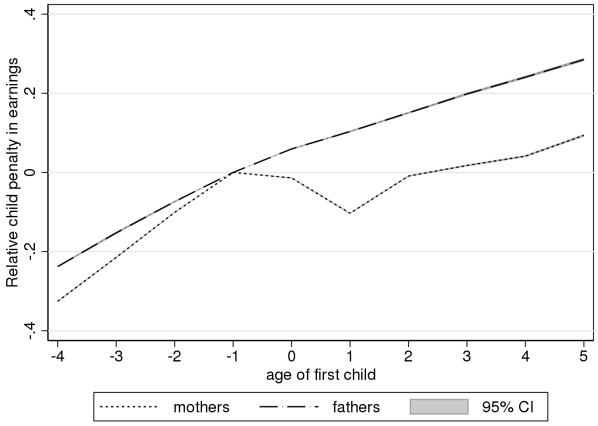


Figure A1: Model Predictions: Simulations for Gender Norms and Specialization

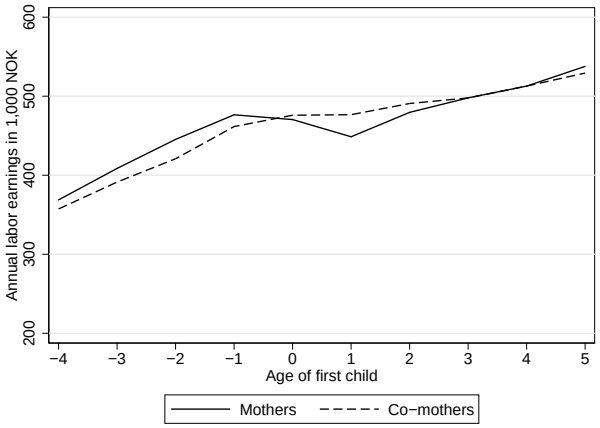
Note: Left panels show individual income penalties relative to full time income in period 1, and right panels show child penalty by couple type. To produce the simulations we set $h(1-t_i) = 1-t_i$. The baseline parameter values are: $k_a = k_b = 1$, $\lambda_a = \lambda_b = .5$, and $\beta = 5$. At baseline, wages of both partners are normally distributed with mean 10 and standard deviation 1. At baseline $\alpha = \eta = \delta = 0$. In panel 1, we solve for 100 equally spaced grid points of $\delta \in [0, 1]$. In panel 2 we solve for 100 equally spaced grid points of $\alpha \in [0, 40]$. In the last panel, we vary the mean of w_a between 10 and 30.



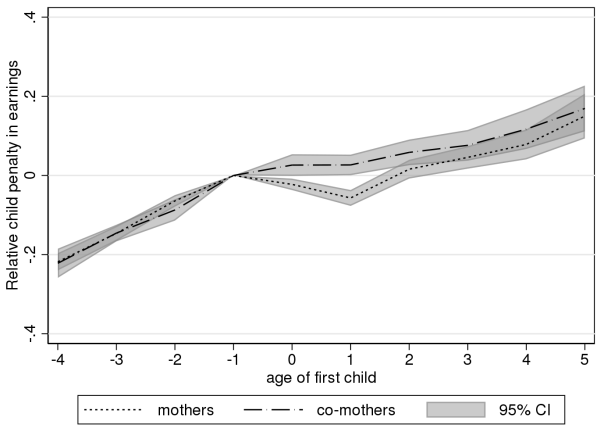
(a) Heterosexual couples



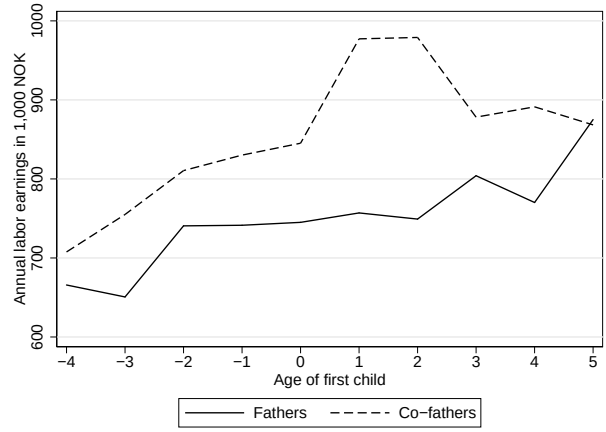
(b) Heterosexual couples



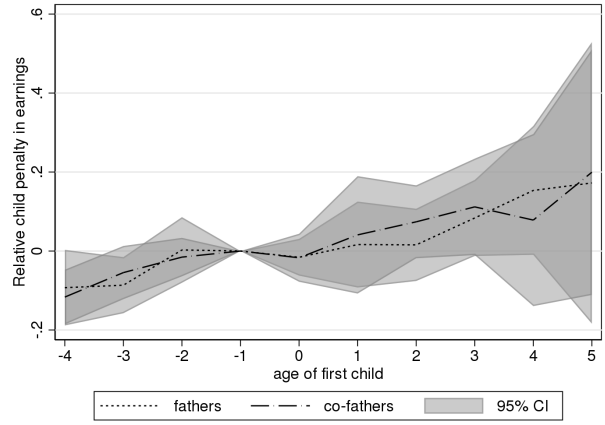
(c) Lesbian couples



(d) Lesbian couples



(e) Gay couples



(f) Gay couples

Figure A2: Mean earnings by event time (left) and raw child penalties (right)

Note: Left panels show show means of annual labor earnings for the years before and after birth of the first child. Right panels show simple event study estimates without year and age profiles. Sample construction and data as defined in section 4. Note that the scale of the y-axes are separate for gay couples compared to heterosexual and lesbian couples.

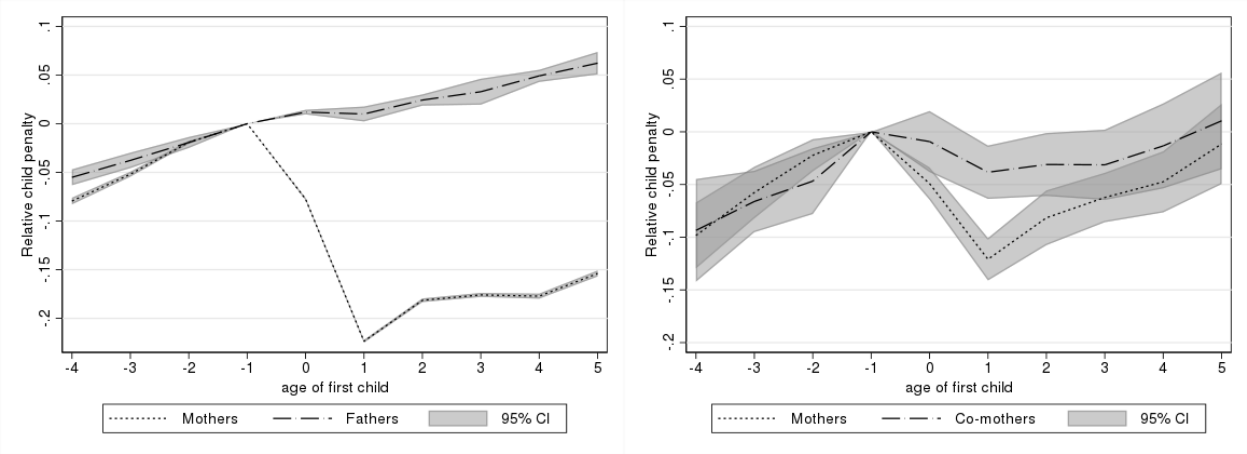


Figure A3: Controlling for education- and gender-specific age profiles

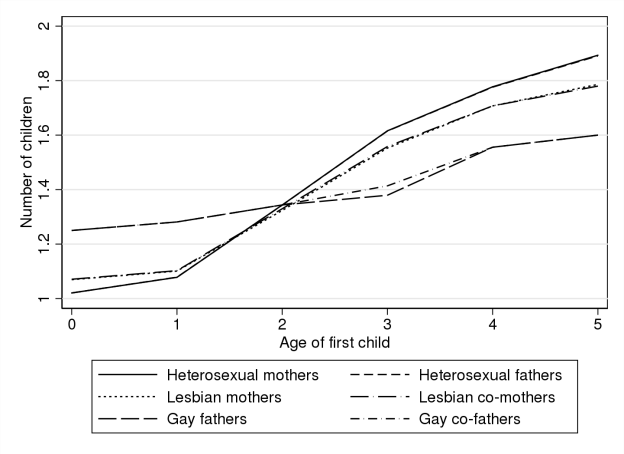


Figure A4: Number of children over time by parent type

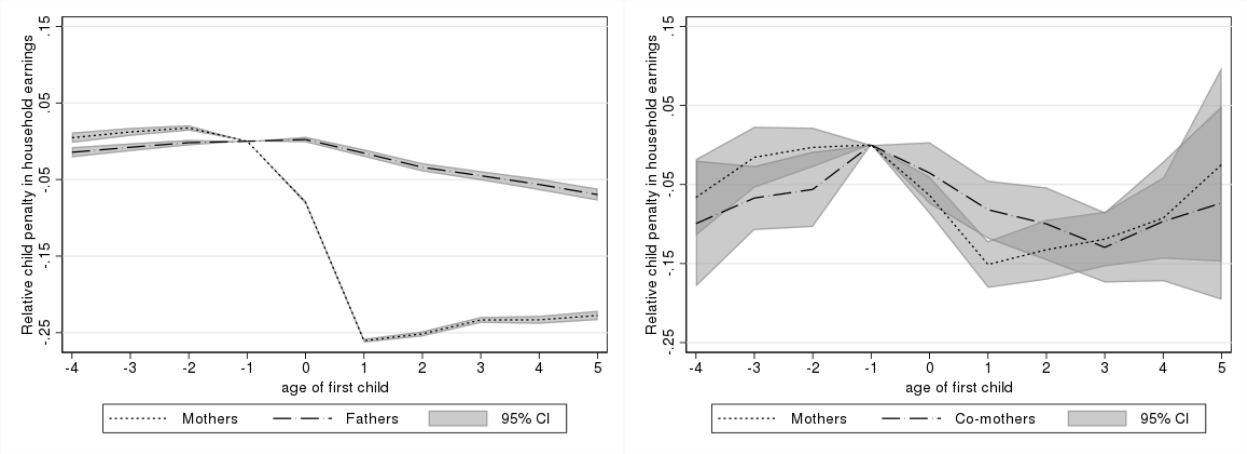
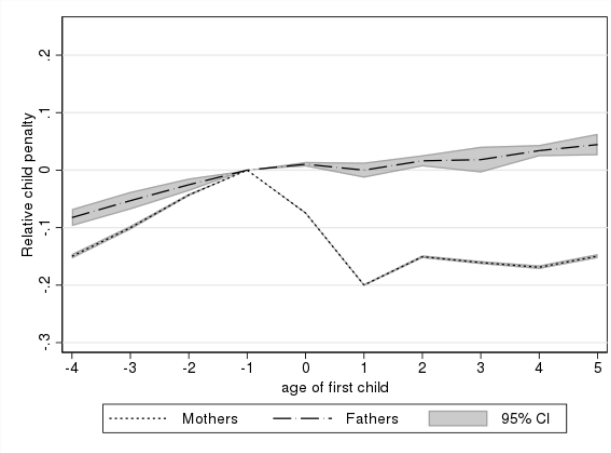
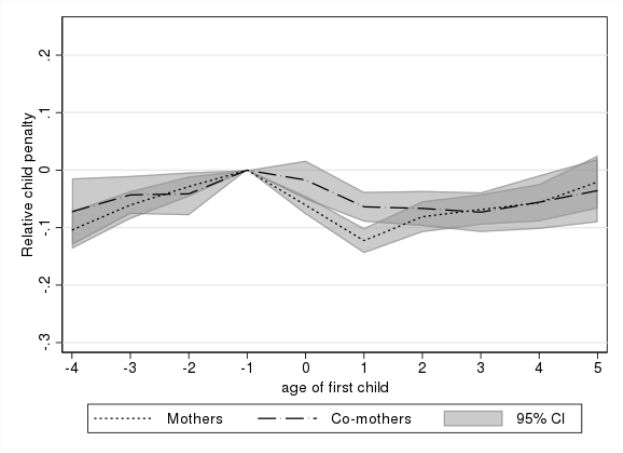


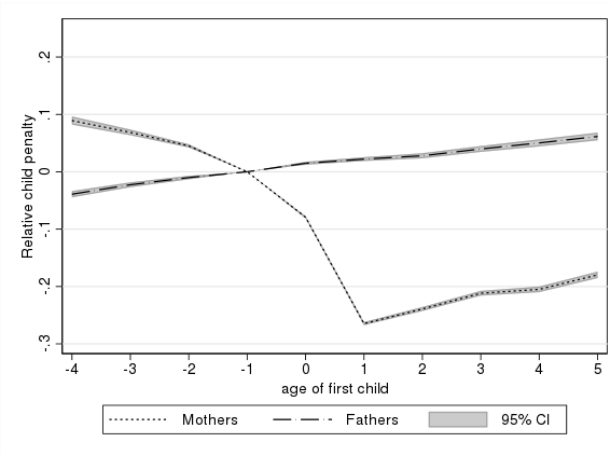
Figure A5: Only couples where no partner has additional kids until $t + 5$



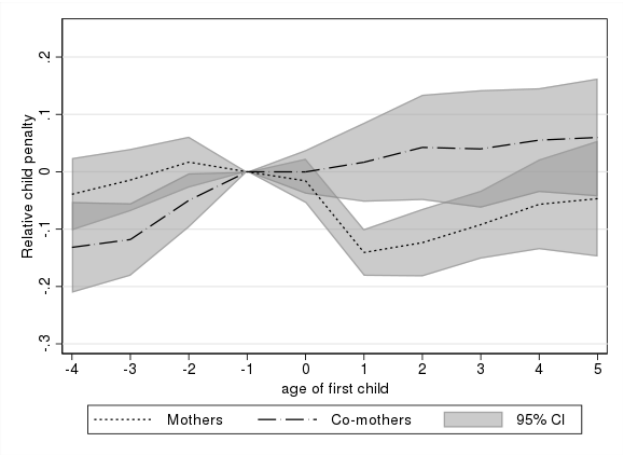
(a) Heterosexual couples, high ed. mother



(b) Lesbian couples, high ed. mother



(c) Heterosexual couples, low ed. mother



(d) Lesbian couples, low ed. mother

Figure A6: Subsample analysis by level of mother's education: High school or below vs. more than high school

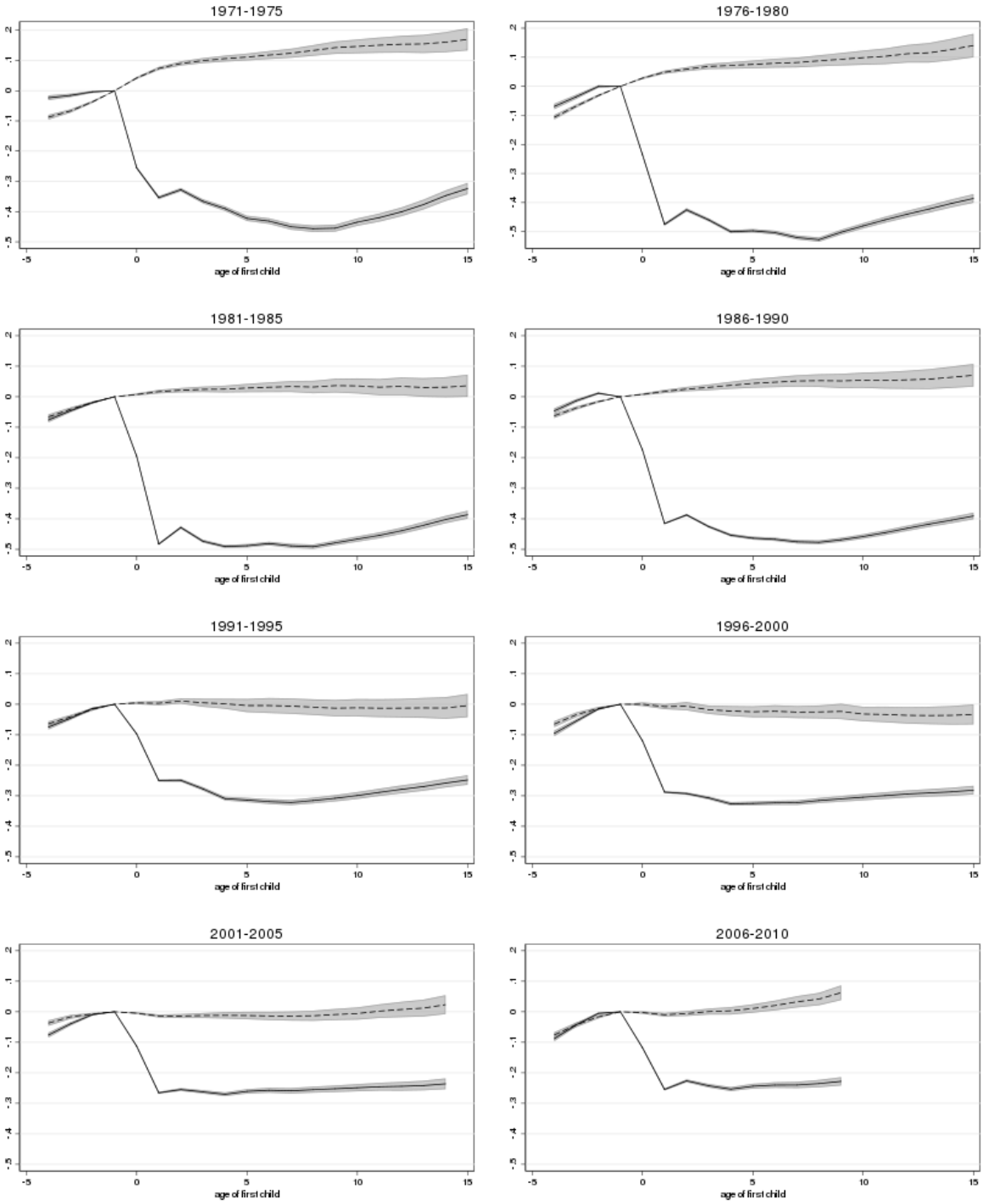


Figure A7: The child penalty over time, heterosexual couples

Note: Child penalties estimated from equation 4 for heterosexual couples, separately by birth year of the first child in five year bins. Mothers in solid lines, fathers in dashed lines, 95% confidence intervals (gray area) calculated using bootstrap, clustering at couple.

Table A1: Summary of the predictions of the model

	Individual Child Penalty		
	Heterosexual	Lesbian	Gay
Preferences (η)	Female spouse	Both spouses (<hetero)	Neither spouse
Biology (α)	Female spouse	One spouse	Neither spouse
Gender norms (δ)	Female spouse	Neither spouse	Neither spouse
Specialization ($\frac{w_a}{k_a} - \frac{w_b}{k_b}$)	Female spouse	One spouse	One spouse
	Relative Child Penalty		
	Heterosexual	Lesbian	Gay
Preferences (η)	Yes	No	No
Biology (α)	Yes	Yes	No
Gender norms (δ)	Yes	No	No
Specialization ($\frac{w_a}{k_a} - \frac{w_b}{k_b}$)	Yes	Yes	Yes

B Parental leave measures and robustness

B.1 Measuring parental leave

The FD-Trygd database provides data on all spells of leave for Norwegian parents. Technically, there are five types of leave spells recorded. In addition to the regular parental leave spells, there are pregnancy leave spells, available for mothers with jobs that impose health risks to the unborn child, such as chemicals or heavy lifts, leave spells for adopted children, combined leave spells and other leave spells. In practice, more than 97% of the leave spells recorded are for regular leave spells , and we focus on these.

Unfortunately, the data does not contain direct links to the child or children for which the leave is taken, only to the individual who takes leave. We therefore have to infer the relevant child

from the birthdates of the children. To this end, we assign a parent's leave spell to a particular child if it

- starts no earlier than 60 days before the birth of the child, and
- starts no later than 3 years after the birthdate of the child, and
- starts no later than 60 days before the birthdate of the next child to the same parent

This mirrors the rules for parental leave, which can be taken up to the age of three, but any remaining leave not taken by the time the next child is born, is lost. Using this procedure, we match 99.45% of all leave spells to a particular child.

The data makes no distinction between leave spells with 80% and 100% wage compensation. We are interested in the number of weeks at home with the child, so that this distinction does not matter, but we treat a day of leave at 80% compensation the same as a day of leave at 100% compensation. In contrast, it is possible to take graded leave, meaning that a parent will have a leave spell where he or she works part-time. In these cases, we compute the number of efficient days at home for each leave spell. Following this, we collapse the total length of all spells for a particular time and scale it to represent weeks of total leave.

Finally, we observe a small number of parents who according to this measure takes longer leave than the total leave allowance, even at 80% compensation. We therefore cap 1.15% of mothers and 0.08% of fathers in our sample who are observed with more than 60 weeks of leave to 60 weeks.

B.2 Additional figures and tables, paternity leave

Table B1 provides sharp RD balancing tests for a range of covariates in the baseline RD model. Figure B1 provide robust local polynomial estimates of the density of births around the cutoff. Reduced form and first stage estimates separately by reform is plotted in Figure B2.

Table B1: Sharp RD balancing tests

Variable	Reform year	2005	2006	2009	2011	2013	2014	Pooled	Stacked
Father's	RD estimate	0.060	-0.19	0.034	-0.090	0.087	-0.046	-0.016	0.0060
age	s.e.	(0.23)	(0.25)	(0.067)	(0.12)	(0.13)	(0.071)	(0.058)	(0.039)
	robust p	0.70	0.46	0.72	0.37	0.44	0.62	0.84	
Mother's	RD estimate	0.20	0.14	0.025	-0.19	0.17	0.0043	0.030	0.026
age	robust s.e.	(0.25)	(0.27)	(0.071)	(0.16)	(0.16)	(0.080)	(0.071)	(0.045)
	robust p	0.49	0.59	0.91	0.30	0.30	0.84	0.71	
Maternity	RD estimate	1.01**	0.048	-0.32***	-0.17	-0.088	-0.65***	-0.35***	-0.32***
leave	robust s.e.	(0.48)	(0.50)	(0.11)	(0.23)	(0.21)	(0.12)	(0.10)	(0.069)
	robust p	0.040	0.96	0.019	0.57	0.85	0.00	0.0075	
Father's	RD estimate	0.085	-0.28	-0.027	-0.059	-0.019	0.0017	-0.030	0.0060
years of ed.	robust s.e.	(0.19)	(0.19)	(0.051)	(0.073)	(0.089)	(0.055)	(0.041)	(0.027)
	robust p	0.90	0.18	0.42	0.50	0.84	0.88	0.42	
Mother's	RD estimate	-0.049	-0.34*	0.044	-0.053	0.094	0.045	0.014	0.033
years of ed.	robust s.e.	(0.19)	(0.19)	(0.045)	(0.090)	(0.088)	(0.050)	(0.038)	(0.029)
	robust p	0.74	0.074	0.36	0.51	0.28	0.34	0.78	
Father's	RD estimate	-0.0047	0.0069	-0.0035	0.0030	-0.0087	-0.0041	-0.0033	-0.0030*
ed. missing	robust s.e.	(0.0079)	(0.0079)	(0.0025)	(0.0049)	(0.0067)	(0.0033)	(0.0025)	(0.0016)
	robust p	0.54	0.39	0.23	0.46	0.17	0.31	0.21	
Mother's	RD estimate	-0.019**	-0.0075	-0.0029	-0.0021	-0.0056	-0.0045	-0.0065***	-0.0042**
ed. missing	robust s.e.	(0.0097)	(0.0093)	(0.0027)	(0.0040)	(0.0064)	(0.0036)	(0.0021)	(0.0017)
	robust p	0.073	0.38	0.38	0.63	0.34	0.39	0.005	

Note: Robust semi parametric sharp RD estimates of the effect of paternity leave quotas on balancing variables using optimal bandwidths, triangular kernel and local linear polynomials on either side of the cutoff. All estimates are scaled to reflect one week of quota increase. Pooled estimates are the simple re-centered robust RD estimates across all six cutoffs. Robust bias-corrected inference except for the stacked estimates, where standard errors are robust, but not bias-corrected. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$, based on the robust, but not bias-corrected standard errors (themselves not reported).

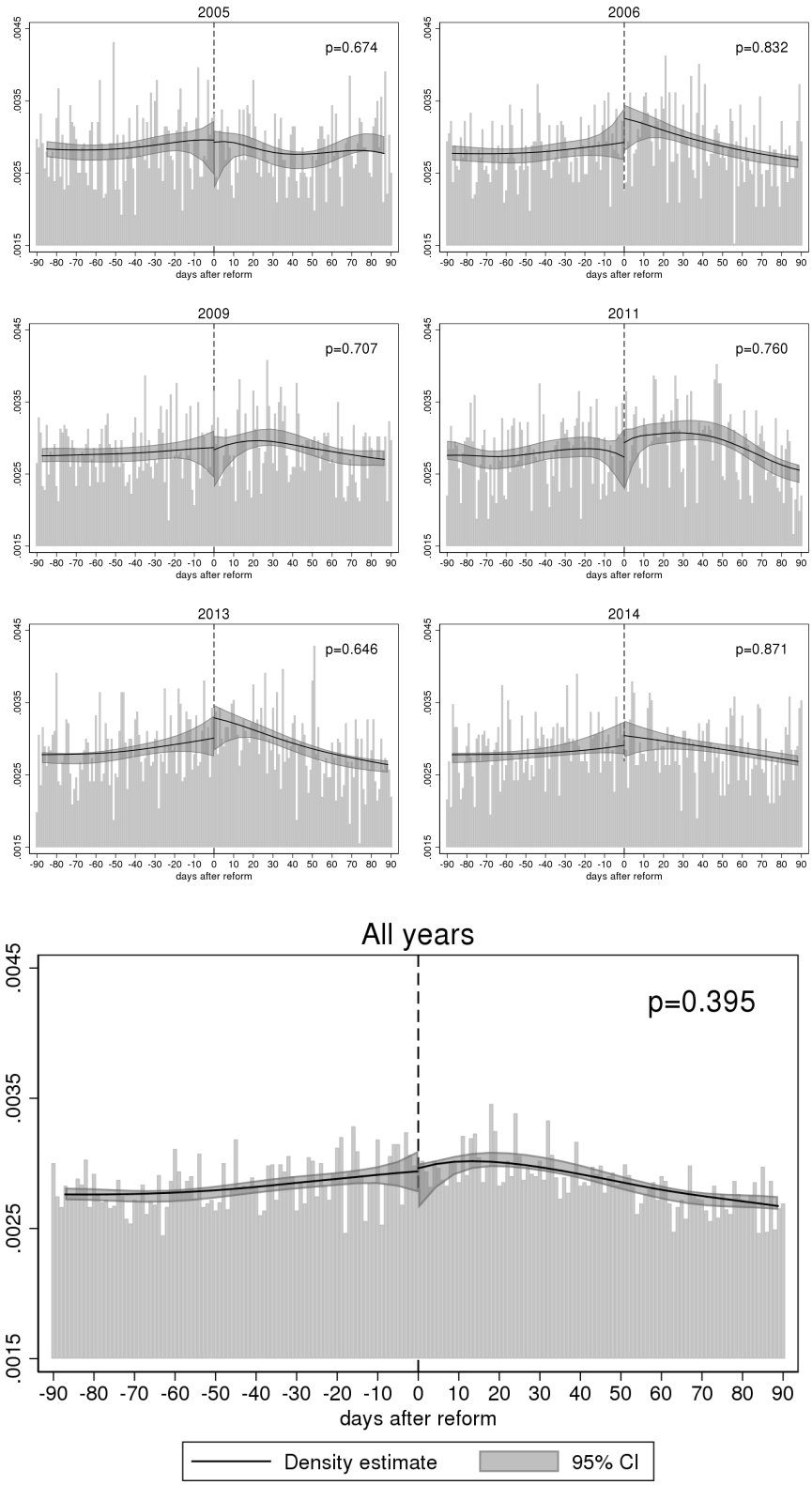
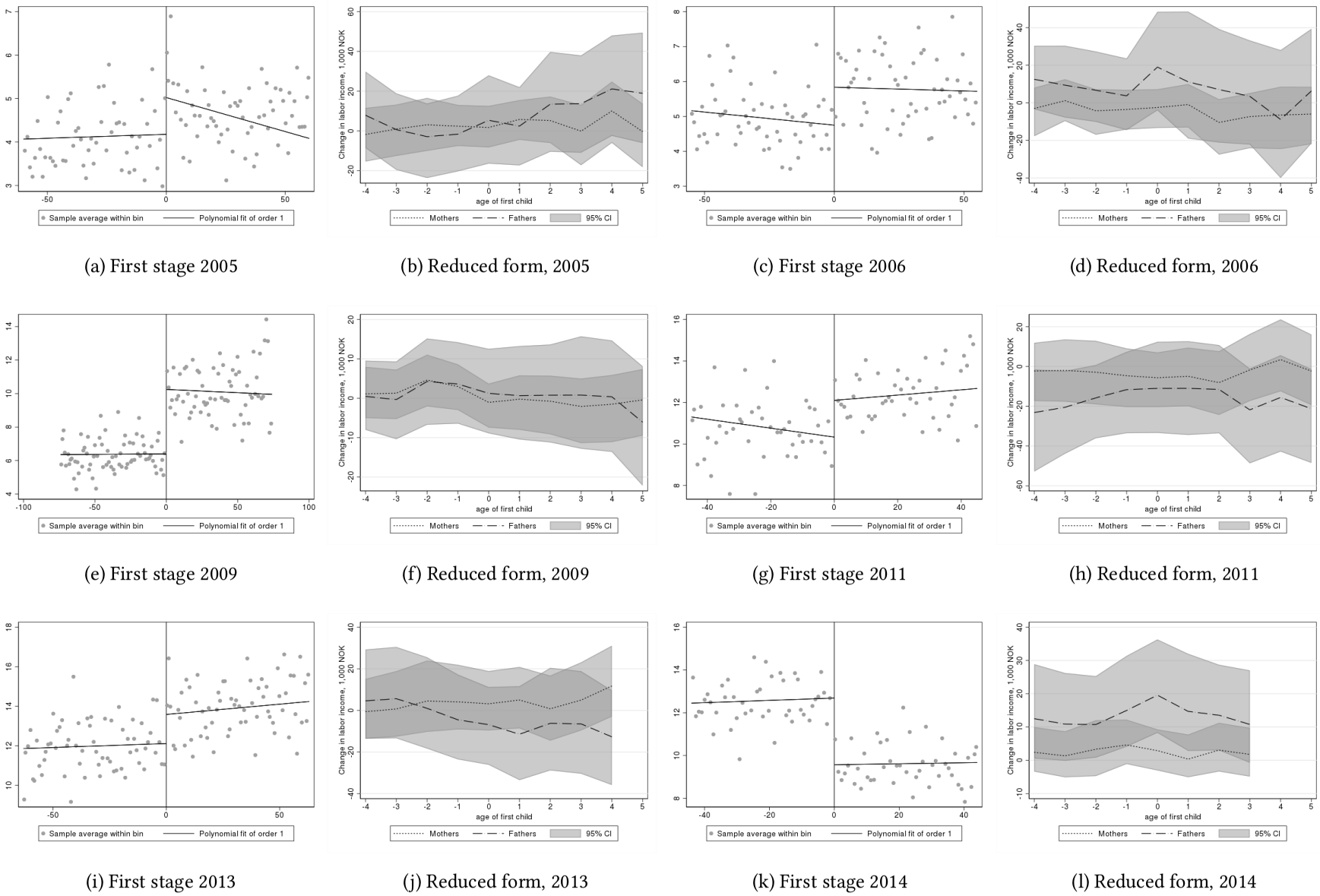


Figure B1: Density plots below and above cutoffs

Note: Graphs show density estimates above and below the cutoff using methods described in Cattaneo *et al.* (2017) and implemented in Cattaneo *et al.* (2018c). p-values reported are for a bias-corrected test of whether the densities at the cutoffs are equal.



(a) First stage 2005 (b) Reduced form, 2005 (c) First stage 2006 (d) Reduced form, 2006

(e) First stage 2009 (f) Reduced form, 2009 (g) First stage 2011 (h) Reduced form, 2011

(i) First stage 2013 (j) Reduced form, 2013 (k) First stage 2014 (l) Reduced form, 2014

Figure B2: Robust RDD estimates, paternity leave reforms

Notes: First and third columns show binned plots of the weeks of paternity leave against birth date of the child in days after the reform, overlaid with the estimated local linear polynomials. Second and fourth panels show sharp RD estimates of the impact of an additional week of quota on maternal and paternal income by year. Optimal MSE-reducing bandwidths, triangular kernel and local linear polynomials on either side of cutoff. Confidence intervals are robust and bias-corrected.

B.3 Accounting for effects of maternal leave

As evident from Table 3, several of the reforms affected not only the paternity leave quota, but also the maternity leave quota and the sum of the maternity leave quota and the shared leave. As documented in Table B1, this resulted in reduced maternity leave take-up roughly for the reforms where the total time a mother could take off work was reduced. Although we argue that this change in maternity leave takeup is relatively minor compared to the change in paternity leave, and at much higher margins, we might worry that it is partly the changed maternity leave that causes any changes in later labor market outcomes, not paternity leave.

To investigate this, we exploit the fact that some of the reforms expanded the paternity leave quota at the expense of maternity leave, while others lengthened the total leave. This means that we can exploit the stacked RD specification to get independent variation in the reform-induced shifts to both maternity and paternity leave use in a 2SLS setup:

$$\begin{aligned} y_{irt} &= \beta_t^L L_i + \beta_t^M M_i + \varphi_r^0 x_i \mathbb{1}(x_i < 0) + \varphi_r^1 x_i \mathbb{1}(x_i \geq 0) + \pi_r + \epsilon_{irt} \\ L_{ir} &= \gamma_{LQ} Q_{ir} + \gamma_{LS} S_{ir} + \varphi_r^{L0} x_i \mathbb{1}(x_i < 0) + \varphi_r^{L1} x_i \mathbb{1}(x_i \geq 0)] + \pi_r^L + \eta_{ir}^L \\ M_{ir} &= \gamma_{MQ} Q_{ir} + \gamma_{MS} S_{ir} + \varphi_r^{M0} x_i \mathbb{1}(x_i < 0) + \varphi_r^{M1} x_i \mathbb{1}(x_i \geq 0)] + \pi_r^M + \eta_{ir}^M \end{aligned} \tag{8}$$

where L_{ir} and M_{ir} are paternity and maternity leave takeup for couple i who is exposed to reform r. Rather than a dummy at the cutoff, the instruments are now Q_{ir}, the paternity leave quota, and S_{ir}, the sum of shared leave and maternity leave quota. Notice that the variation in these two instruments are determined solely by the cutoff in birthdates, and that we have independent variation to separate the effects of both instruments because we stack all six reforms to parental leave. As before, we use local linear polynomials that are separate on either side of the cutoff for each reform and a triangular kernel to control for the forcing variable. The outcome variable y_{irt} is labor market earnings, measured separately for mothers and fathers. This leaves us with two treatments by two outcomes per year we measure outcomes.

When instrumenting for two endogenous variables in an IV-setup, it is not clear how to de-

termine the optimal MSE-reducing bandwidth as before. We therefore use a) the MSE-reducing optimal bandwidth for the first stage of either of the instruments or b) a fixed 50-day bandwidth. As before, we report robust, but not bias-corrected standard errors for the stacked specification.

First stage results for the two endogenous variables are reported in Table B2. Notice that independent variation to identify both effects rely on stacking all reforms, so that we cannot perform these estimates separately by reform. Notice first that the choice of bandwidth is not of essence: The results are very similar whether we use either of the MSE-reducing optimal bandwidths or a fixed 50-day window. Second, note that the reforms work exactly as we would expect: An increase in the daddy quota of 1 week increases paternity leave uptake by almost exactly 1 week when we control for changes to the remaining quota for the mother. Increasing the remaining leave for the mother (comprising maternal quota and the weeks of shared leave) increases maternity leave take up by 0.7 to 0.8 weeks. In contrast, the instruments do not work across spouses: Weeks of paternity leave quota does not affect maternity leave use when controlling for the remaining share available to the mother, in contrast to the balancing exercise in Table B1, while the remaining share for the mother does not affect leave update for the father when controlling for his own quota. Thus, the stacked specification where we instrument for both parents' leave take up circumvents the problem of the reforms affecting both margins of leave. Because the choice of bandwidth does not seem to matter and because we're primarily interested in the effects of paternity leave, we present fuzzy stacked RD estimates based on this specification using the paternity leave-optimal bandwidth from panel C. As in the base model in the paper we also revert to the reform-specific dummies as instruments when reporting the IV estimates rather than quota measures.

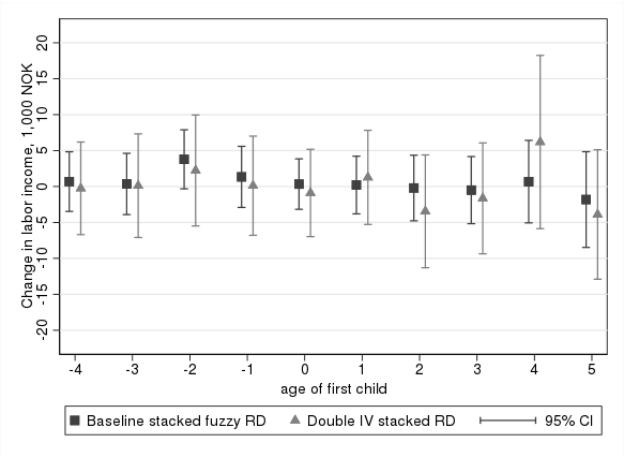
Results from the stacked fuzzy RD model where we instrument for both mothers' and fathers' leave take up is presented in Figure B3. The top panels present effects of paternal leave on mothers' and fathers' earnings by child age, mirroring the estimates from the baseline model. For reference, the coefficients and confidence intervals from the stacked fuzzy RD model where we instrumented for paternity leave use only is added. Except perhaps for the outlier at child age 4, the double IV model provides estimates that are well in line with the baseline model, confirming

Table B2: First stage effects of maternity and paternity leave quotas

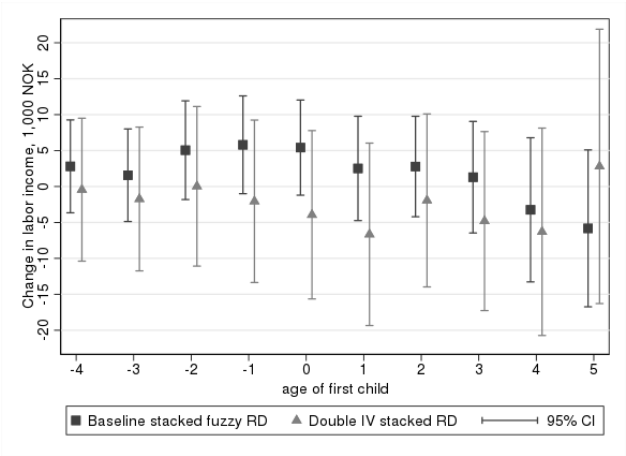
	Weeks of leave		Bandwidth		
	Mother	Father	reform	bw	N
A: 50-day bandwidth					
Paternity leave quota	0.066	1.00***	2005	50	4,037
(Q_{ir})	(0.14)	(0.16)	2006	50	4,303
			2009	50	4,192
Remaining leave for mother	0.77***	0.18	2011	50	3,656
(S_{ir})	(0.21)	(0.21)	2013	50	4,830
joint F	21.1	76.5	2014	50	3,502
N			24,520		
B: Maternity leave-optimal bandwidth					
Paternity leave quota	0.069	0.96***	2005	66.9	5,418
(Q_{ir})	(0.14)	(0.15)	2006	61.3	5,271
			2009	42.4	4,017
Remaining leave for mother	0.79***	0.12	2011	43.8	4,252
(S_{ir})	(0.20)	(0.20)	2013	57.7	5,538
joint F	24.8	79.7	2014	52.4	4,532
N			29,028		
C: Paternity leave-optimal bandwidth					
Paternity leave quota	0.055	0.98***	2005	58.0	4,770
(Q_{ir})	(0.14)	(0.15)	2006	68.4	5,844
			2009	44.6	4,192
Remaining leave for mother	0.73***	0.12	2011	37.4	3,656
(S_{ir})	(0.21)	(0.21)	2013	55.3	5,380
joint F	18.0	74.9	2014	40.8	3,502
N			27,344		

Note: First stage results from stacked specification of all six parental leave reforms, instrumenting for weeks of paternity and maternity leave take up as described in eq. 8. Panel A) uses a fixed 50-day bandwidth, panel B) uses the MSE-reducing optimal bandwidth for each reform if instrumenting for maternity leave only, panel C) the same for paternity leave. Heteroskedasticity robust, but not bias-corrected standard errors. $^{*}p < 0.1,^{**} p < 0.05.^{***}p < 0.01$

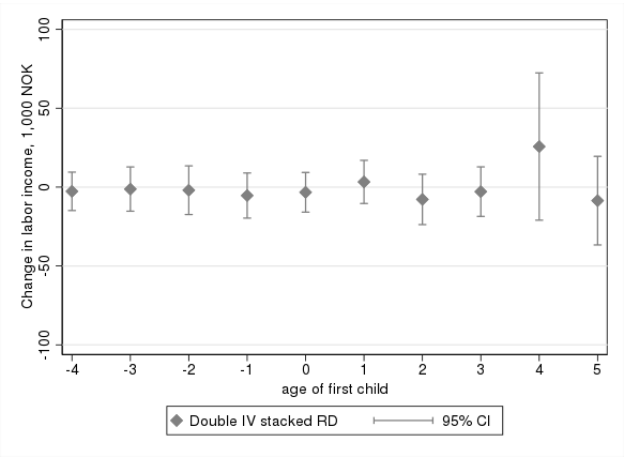
the precise zero effects of paternity leave on mothers' subsequent labor earnings. Just like in the basic model, it does not seem like paternity leave has a potential for reducing the child penalty. The double IV specification inadvertently also estimates the effects of another week of maternity leave on parents later earnings. Results are too imprecise to draw strong conclusions, but provide no evidence of any effects. In short, parental leave policies does not seem like a promising tool for reducing the child penalties in Norway.



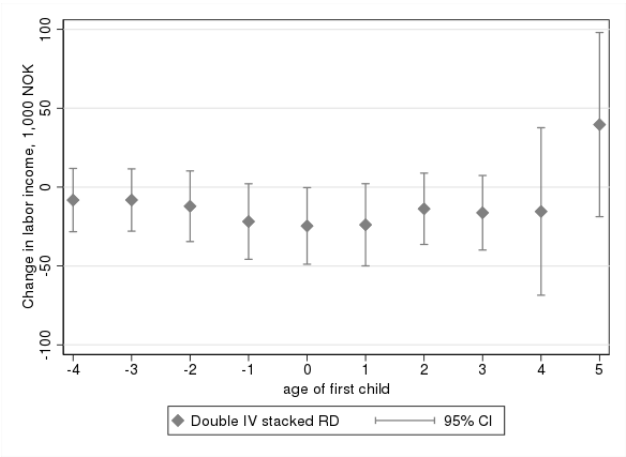
(a) Paternity leave on mother's earnings



(b) Paternity leave on father's earnings



(c) Maternity leave on mother's earnings



(d) Maternity leave on father's earnings

Figure B3: Effects of maternity and paternity leave use on mothers' and fathers labor earnings

Note: Top panels shows the impact of a week of paternity leave use on mothers' (left) and fathers' (right) earnings over time, as estimated from a double IV stacked fuzzy RD as detailed in eq. 8. For comparison we also show our stacked fuzzy RD estimates from the baseline model where we only instrument for the weeks of paternity leave. Bottom panels show the impact of an additional week of maternity leave on mothers' (left) and fathers' (right) earnings.