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

Martin Andresen and Emily Nix

PRELIMINARY DRAFT - PLEASE DO NOT CITE OR CIRCULATE WITHOUT PERMISSION FROM THE AUTHORS

Abstract

While the gender gap in income has narrowed over the past 50 years, women in heterosexual couples continue to experience significant labor market penalties following the birth of children, while their partners experience no such penalty. This "relative child penalty" has been well documented and accounts for the majority of the remaining gender gap in income in many countries. A number of possible explanations for this relative child penalty exist: gender norms around child care, different preferences for child care among men and women, efficient specialization within households, and the impact of giving birth on women. In the first half of this paper, we show that same sex couples do not experience the relative child penalty in the same way as heterosexual couples. We develop a simple economic model that incorporates the main explanations for the child penalty and generates testable predictions. The model, combined with the empirical results, suggest that much of the child penalty experienced by women in heterosexual couples is due to gender norms and preferences, although the costs of giving birth also contribute. Having established that the relative child penalty in heterosexual couples is not inevitable and may not be efficient, in the second half of the paper we provide causal estimates on the impact of two policies aimed at reducing the child penalty: paternity leave and improved access to child care. We find no significant impact of paternity leave on the relative child penalty. The analysis of improved access to child care is in progress.

The gender gap has narrowed significantly over the past 50 years.[1] However, one component of the gender gap has proven to be relatively persistent: the income penalty women in heterosexual

*We thank Kenneth Aarskaug Wiik, Edwin Leuven, and Nina Drange.

[1]Economists have provided evidence on a number of explanations for this decline, such as the narrowing of the gender education gap and the decrease in labor force discrimination.

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 called the "child penalty"[2] and has been documented in a variety of countries such as the United States, Denmark, Norway, and Sweden (see Chung *et al.* (2017), Kleven *et al.* (2018), Bergsvik *et al.* (2018), and Angelov *et al.* (2016)). As other determinants of the gender 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 accounts for 80% of the gender gap in 2013, compared to 40% in 1980.

The stubborn persistence of the relative child penalty among heterosexual couples is a puzzle, particularly given the overall decline in gender wage gaps. If the relative child penalty is largely driven by differences in preferences between men and women, the impact of giving birth, or efficient specialization within households, then the relative child penalty may be an optimal response to the arrival of children. On the other hand, the relative child penalty could be caused by persistent gender norms around child care which may be economically inefficient. In addition to contributing to the gender gap, we also find that the arrival of children in heterosexual couples causes a drop in overall household income. Does this drop represent a necessary cost to households of having children?

To address these questions 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. If the absence of pre-set gender roles lead same sex couples to also split the burden of child care more evenly, the 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, giving

[2]Although commonly called a "penalty", this could very well be driven purely by preferences and not discrimination, which some may associate with the word penalty. This paper aims at disentangling these mechanisms, but we will use the term "child penalty" for the income loss following child birth independently of the mechanism. This is in line with the literature.

birth, household specialization and gender norms around child care in the heterosexual relative child penalty, we build a simple model of the household's 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 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. 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 differential among same sex male partners. If 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.

Similar to previous papers, we find that women in heterosexual couples experience a drop in income of approximately 22% following the birth of the first child, and this drop persists over time. Their husbands experience no income penalty from children. 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 experiencing 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, until there is no longer a child penalty four years after birth. These patterns suggest that both biology and female preferences play a role in the relative child penalty experienced by heterosexual couples, but gender norms likely play a large role as well. 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. In terms of total household income, we find that the drop is larger for heterosexual couples compared to same sex female couples.

From a social planner's perspective, if shared parenting has negative impacts on children, then reducing the heterosexual relative child penalty might not be optimal even if much of the differences in the child penalties experienced by heterosexual men and women is caused by gender norms. We find that despite a lower overall household income penalty within same sex couples, the children of same sex couples outperform children of heterosexual couples on English, reading and math tests at age 10.

Thus, large and persistent differences in the child penalty within a couple are not inevitable, and most of this gap is due to gender norms and women's preferences. The evidence from the first half of the paper helps us better understand the mechanisms behind the relative child penalty in heterosexual couples, but it does not tell us what impact policy might have on the relative child penalty. Policy makers might wish to know how to decrease the relative child penalty for two reasons. First, decreasing the relative child penalty would almost certainly reduce the overall gender income gap. Second, decreasing the relative child penalty could also improve welfare more generally given our findings on household income and children's test scores at age 10.

Broadly speaking there are two possible approaches to try and reduce the relative child penalty: increase father participation in child rearing or provide support to mothers. Successful policies targeting fathers would increase the child penalty for heterosexual fathers while decreasing the child penalty for heterosexual mothers. Successful policies targeting mothers would decrease the child penalty for heterosexual women as a result of the state serving as a "third partner" who shares the burden of child rearing with women. These two approaches are very different, and it is not clear a priori which approach is best.

In the second half of this paper, we compare the impact of a policy targeting fathers on the relative child penalty of heterosexual couples versus a policy targeting mothers. First, we estimate the causal impact of paid paternity leave on the relative child penalty. We use a regression discontinuity design to estimate the impact of 6 different expansions of the paid paternity leave quota in Norway from 1993 to 2013. We estimate a strong first stage: the reforms significantly increased the amount of paternity leave that fathers take. However, we find no significant impact on the relative child penalty. Our results are consistent with Antecol *et al.* (2016) who find that gender neutral tenure clock stopping policies do not help academic women, and may even hurt their careers. We examine a similar shift toward more gender neutral leave policies, and find that they do not help women across the population of professions in Norway.

Next, we estimate the causal impact of a 2002 reform that greatly increased the funding provided for formal child care, with significant variation in the timing of the expansion across municipalities. We use the same IV strategy developed in Andresen and Havnes (2018), but look specifically at the relative child penalty. [IN PROGRESS].

Our paper is most closely related to the literature on child penalties. We use the simple event study approach used in Chung *et al.* (2017), Kleven *et al.* (2018), Bergsvik *et al.* (2018), and Angelov *et al.* (2016) to identify child penalties across couple types. Together, our results and the results from these papers make clear two facts. First, there does not appear to be a sample of heterosexual couples, whether in different countries, educational groups, or socioeconomic class, that does not experience large relative child penalties. Second, the child penalty has become the most important driver of gender gaps. However, by using same sex couples as a comparison, our paper contributes more substantially to this literature by shedding light on why the relative child penalty exists, which is difficult to surmise by looking only at heterosexual couples. 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 the women report more negative opinions toward female employment after giving birth relative to before birth.

We also contribute to a small body of evidence on same sex couples with children. While comparing the outcomes of children born to same sex and heterosexual couples is not the focus of this paper, this topic has been a major point of controversy in the United States and elsewhere. In the landmark 2015 Supreme Court case *Obergefell v. Hodges,* which legalized same sex marriage, the well being of children was a central theme in oral arguments. Justice Scalia raised a concern that not all studies conclude children of same sex parents fare equally well. The empirical evidence has been limited and mostly based in the fields of sociology and psychology. 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[3], mislabeling children from heterosexual couples as children of homosexual couples or vice versa[4], 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. While the population of children born to same sex couples in our sample years is modest when compared to the population of children born to heterosexual couples, it is much larger than the vast majority of existing studies. We find that same sex couples are different from heterosexual couples in terms of observables before birth, so when estimating the impact of having same sex parents on test scores at age 10 we control for income, age, education and other factors that could affect child outcomes (the results are stronger without these controls). We find no evidence of adverse schooling outcomes for children of same sex parents. On the contrary, we find 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.

The remainder of the paper is organized as follows. In Section 2 we present a model for household production 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. Having established that the child penalty is not inevitable, in Section 6 we analyze the impact of paternity leave and access to better child care on the heterosexual child penalty. In Section 7 we conclude.

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

[4]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.

1 A model of household labor supply in the presence of 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, preferences, 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).

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).[5] In the 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 the first period there is no child and no home production, so all adults supply their unit of time inelastically to the market. In the second and third 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:

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

where c is consumption and θ is child quality ($\ln \theta$ is equal to zero in the first period). $\bar{Z}_i$ is an indicator equal to 1 if the individual is a male married to a female in periods 2 and 3, and $\bar{X}_i$ is an indicator equal to 1 if the individual is female. β represents the value of child quality 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.[6]

[5] 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.

[6] 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 US, 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.

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_a and w_b, so that

$$c = w_a t_a + w_b \left(1 - \delta \bar{S}\right) t_b$$

where $\bar{S}$ is an indicator equal to 1 in the second period if the spouse is a woman who gave birth, and δ is the labor market cost of giving birth. We represent total income of each individual in each period as $Y_i = w_i t_i$.

Child quality is produced by the following production function

$$P = k_a h \left(1 - t_a\right) + k_b h \left(1 - t_b\right)$$

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 periods 2 and 3, where household utility is given by

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

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.[7] 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. In the appendix, we show that this approach yields similar predictions to the current model.

There are no dynamics to the problem. This means we can solve the problem sequentially, maximizing t_a and t_b in each period. In period 1, $t_a = t_b = 1$ by assumption. For periods 2 and 3,

[7]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.

the couples solve:

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

The first order conditions for the second period are given below. The first order conditions for the third period are identical, except $\delta = 0$.

$$\begin{aligned}\frac{w_a}{k_a} &= \frac{\beta h'(1-t_a)}{k_a h(1-t_a)+k_b h(1-t_b)} + \frac{\lambda_a \eta \bar{X}_a}{k_a(\lambda_a+\lambda_b)(1-t_a)}\\ \frac{(1-\delta)w_b}{k_b} &= \frac{\beta h'(1-t_b)}{k_a h(1-t_a)+k_b h(1-t_b)} + \frac{\lambda_a \alpha \bar{Z}_a}{k_b(\lambda_a+\lambda_b)} + \frac{\lambda_b \eta \bar{X}_b}{k_b(\lambda_a+\lambda_b)(1-t_b)}\end{aligned} \tag{2}$$

Equations 2 yield the following predictions:

1. **Preferences:** 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 women who gives birth as δ increases, but only in period 2 (the income penalty is decreasing for her partner). The relative child penalty in period 2 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.

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

Table 1: 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; Period 2 only	Yes; Period 2 only	No
Gender norms (δ)	Yes	No	No
Specialization ($\frac{w_a}{k_a} - \frac{w_b}{k_b}$)	Yes	Yes	Yes

heterosexual couples is increasing in α at an increasing rate if $h^{''} < 0$ and at a constant rate otherwise. By construction, η 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.

We also capture the predictions of the model for period 2 income penalties and within couple relative child penalties graphically in Figure 1. Figure 1 plots 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 1 as we vary each parameter (η, δ, α, and

w_a) individually. Note that period 3 graphs are identical, except there is no longer a biological penalty.

Every mechanism leads to a child penalty for 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. While the simulations show the impact of each mechanism separately, the general predictions of the model hold for any combination of parameter values. Based on the model, we can rule out specialization if we compare similar couple types in terms of market and household productivity if we don't also see a relative child penalty for lesbian and gay 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 in period 2, but not in period 3 for lesbian couples. We can rule out preferences 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 will be smaller than the child penalty for heterosexual women, which can also be seen in Figure 1. The intuition is that in heterosexual couples, the husband will decrease labor supply less to compensate for lost income from the mother, while in lesbian couples both spouses will do some of this compensation. This will be an important caveat for our results, and it will be key to understand how much the income penalties might differ in the third period if $\alpha = 0$.

2 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, 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 sharp changes when the child arrives, we can attribute the corresponding discontinuity in income

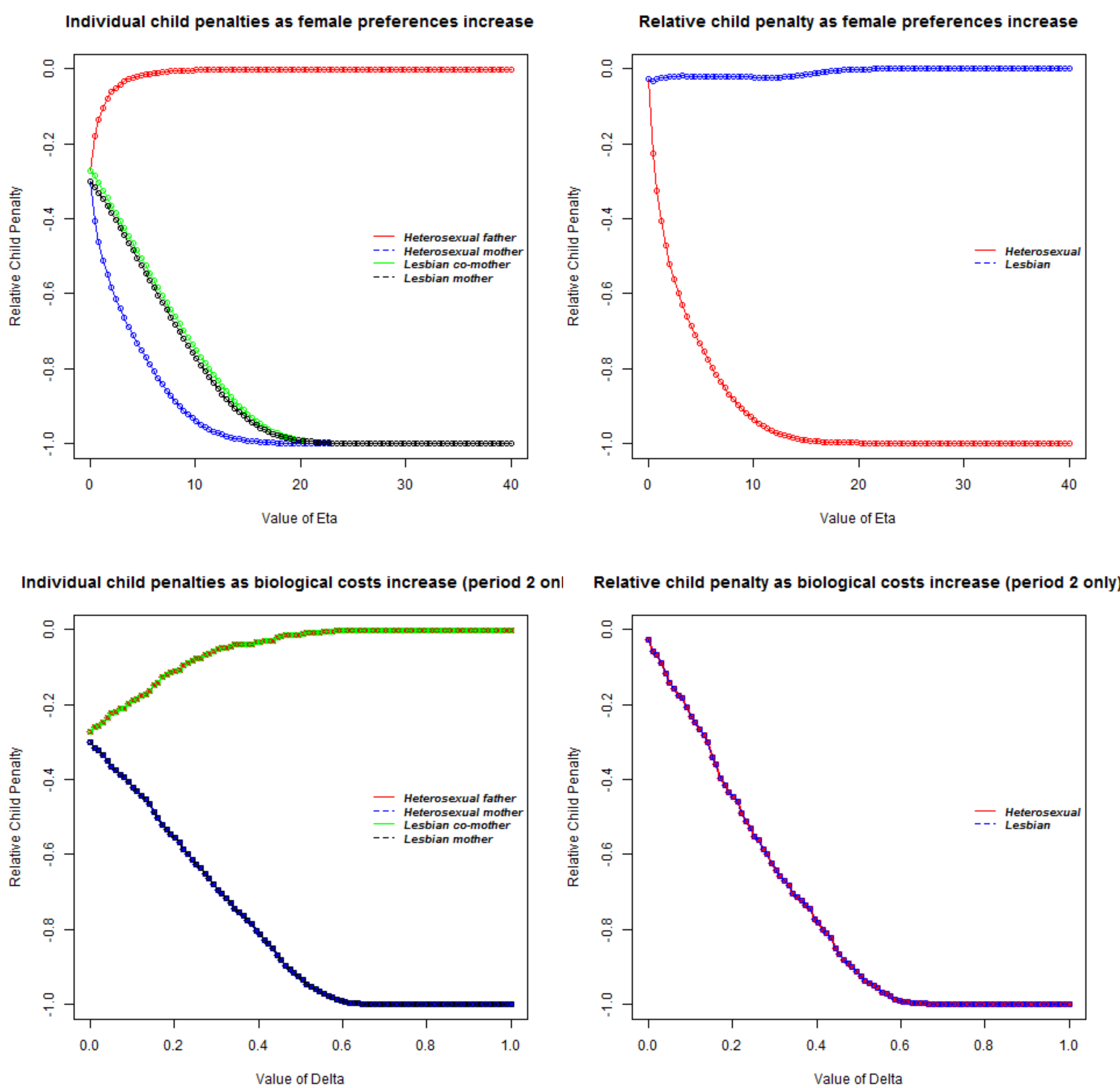


Figure 1: Model Predictions: Simulations for Preferences and Biology

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 $\eta \in [0, 40]$, keeping all other values fixed. Similarly, in panel 2, we solve for $\delta \in [0, 1]$.

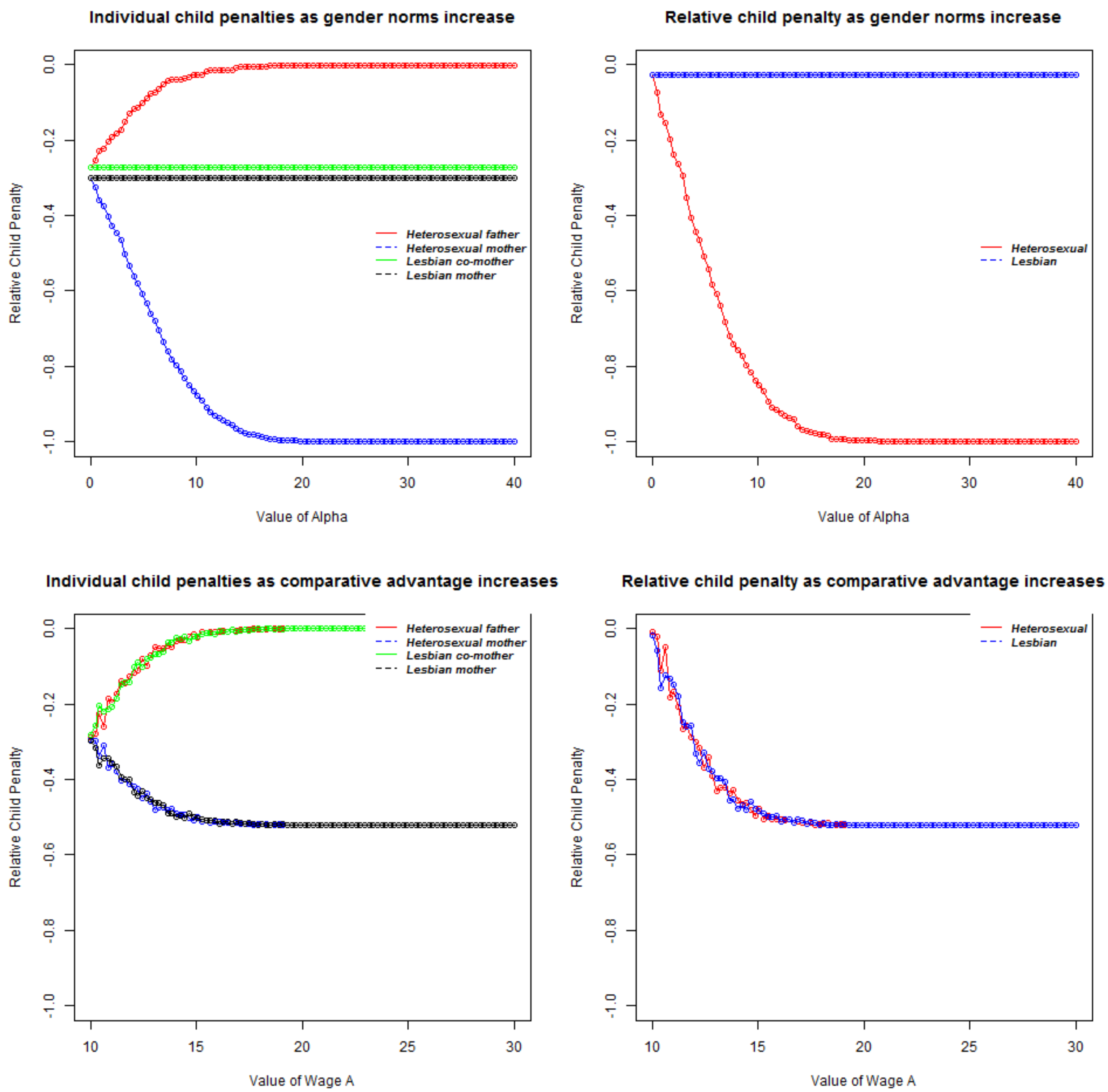


Figure 2: Model Predictions: Simulations for Gender Norms and Specialization
Note: Left panels show individual income penalties relative to full time income in period 1, and right panels show child penalty by couple type. To produce the simulations we set $h(1-t_i) = 1 - t_i$. The baseline parameter 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 $\alpha \in [0, 40]$. In the last panel, we vary the mean of w_a between 10 and 30.

to the arrival of children.

This suggests a simple regression of the outcome of interest on event time dummies. For our main results we also include gender specific age and year dummies which control flexibly for life-cycle and time trends in income. The results with only event time dummies are included in Figure 13 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. 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 time t. We estimate the following equation to identify the child penalty:

$$\begin{aligned} y_{it} = & \overbrace{\sum_{j \neq -1} \sum_{k} \alpha_{jk} \mathbb{1}[t = j, K_i = k]}^{\text{Parent-type-specific event time dummies}} + \overbrace{\sum_{l} \sum_{m} \beta_{lm} \mathbb{1}[age_{it} = l, X_i = m]}^{\text{Gender- specific age profile}} \\ & + \underbrace{\sum_{n} \sum_{o} \gamma_{no} \mathbb{1}[T_{it} = n, G_i = o]}_{\text{Gender-specific calendar year shocks}} + \underbrace{\sum_{p} \eta_p \mathbb{1}[K_i = p]}_{\text{Parent-type specific fixed effects}} + \epsilon_{it} \end{aligned} \tag{3}$$

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, 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. Note that while we allow life-cycle and time trends to vary by gender, we do not allow them to differ within gender. This means that the effect of age and year on income do not depend on whether a woman in a lesbian couple is registered as the primary mother or the co-mother. Equation 3 is equivalent to running the regressions separately for mothers and father if we only estimate the equation for heterosexual couples.[8] Notice that all parents in our sample eventually have children, so that

[8]While it is possible to estimate equation 3 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 controls for the same-sex couples as well as heterosexual couples.

the event dummies are identified from comparisons of 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 year and calendar-fixed effects, 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 j compared to the earnings the year before birth. Ideally, we would use a log-linear specification of equation 3 so that we could interpret the coefficients as percentage change 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.

$$C_{jk} = \frac{\hat{\alpha}_{jk}}{\mathbb{E}(\hat{y} \mid t = j, K_i = k)} \tag{4}$$

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. As an alternative, we also construct

$$P_j = \frac{\hat{\alpha}_{j2} - \hat{\alpha}_{j1}}{\mathbb{E}(\hat{y}_{it} \mid t = j, K_i = 1)} \tag{5}$$

which is a measure of the relative child penalty of mothers relative to men in heterosexual couples, in percentages of the predicted wages of mothers in the absence of children. Notice that spouse type $K_i = 1$ denotes mothers in heterosexual couples and $K_i = 2$ denotes fathers in heterosexual couples. When computing confidence intervals or standard errors for these estimates, we use cluster bootstrap to account for the fact that the denominator is an estimated object.

2.1 Comparing heterosexual and same sex couples

If the exact timing of births is as good as randomly assigned conditional on gender-specific age profiles and yearly shocks, 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 $\frac{w_a}{k_a} - \frac{w_b}{k_b}$ is identical across couple types, according to our model. If the distributions are not identical, we can control for specialization in the event study by estimating:

$$
\begin{aligned}
y_{it} = & \overbrace{\sum_{j\neq -1}\sum_{k}\alpha_{jk}\mathbb{1}[t=j,K_i=k]}^{\text{Parent-type-specific event time dummies}} + \overbrace{\sum_{j\neq -1}\theta_j\mathbb{1}[t=j]\left(\frac{w_i}{k_i}-\frac{w_{-i}}{k_{-i}}\right)}^{\text{time dummies interacted with relative productivity difference}} \\
& + \overbrace{\sum_{l}\sum_{m}\beta_{lm}\mathbb{1}[age_{it}=l,G_i=m]}^{\text{Gender- specific age profile}} + \underbrace{\sum_{n}\sum_{o}\gamma_{no}\mathbb{1}[T_{it}=n,G_i=o]}_{\text{Gender-specific calendar year shocks}} \qquad (6) \\
& + \underbrace{\sum_{p}\eta_p\mathbb{1}[K_i=p]}_{\text{Parent-type specific fixed effects}} + \epsilon_{it} \qquad (7)
\end{aligned}
$$

We observe pre-child income, I_a and I_b. We do not observe productivity in home production (k_a, k_b). However, $I_a - I_b$ is sufficient if $k_a = k_b$, or if one of the following conditions hold. First, consider a more general model where there is 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 $I_a - I_b$ controls for $\frac{w_a}{k_a} - \frac{w_b}{k_b}$. Second, if k is instead identical for all women and smaller than k for all men (for example, women do more household chores as girls then men), then controlling for $I_a - I_b$ should also be sufficient.

We use two approaches to control for pre-birth comparative advantage in market work. First, we flexibly control for the differences in own and spouse's earnings either the year before birth or four years prior to birth interacted with event dummies, by estimating equation 7, replacing $\frac{w_a}{k_a} - \frac{w_b}{k_b}$ with $I_a - I_b$. To the extent that comparative advantage is captured by the relative income levels of the two spouses, these flexible 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. 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 identical to either lesbian or gay couples based on observables. Using this approach, we re-estimate equation 3 using the weighted sample of matching heterosexual couples. In our matching results we control for the following variables: own income at event time -1, income of partner at even time -1, age, gender, own and partner's education, and year [IN PROGRESS]. As before, we scale our estimates by the predicted earnings in the absence of children. Notice that this means that we will plot the partial, or unexplained by comparative advantage, portion of the child penalties when controlling for comparative advantage as a fraction of earnings in the absence of children.

3 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 3 documents the number of new same sex male and female partnerships in Norway following the Partnership Act.[9] Under this act, a partnership was legally equivalent to marriage in most respects. However, the partnerships were restricted regarding children. Same sex couples were not eligible for adoptions within country, 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 implements 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. The change to the guidelines allowed same-sex couples to be considered for adoption of stepchildren just like

[9]Aarskaug Wiik *et al.* (2014) investigates the stability of these same sex marriages and partnerships.

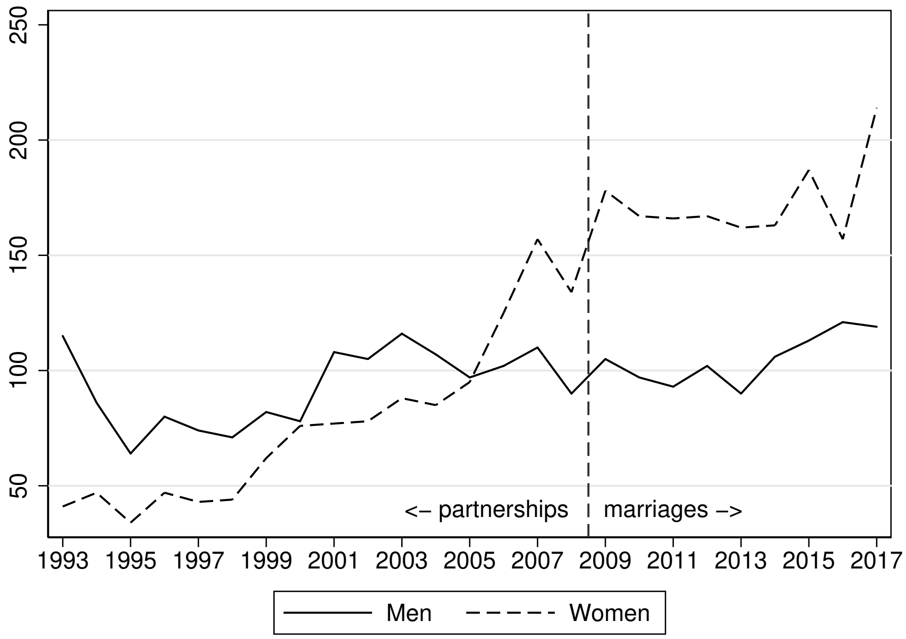


Figure 3: Number of new same-sex partnerships and marriages in Norway, 1993-2017

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

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 5 years, as well as consent from the existing legal parent. If the child was already registered with two parents, the other parent was given the right to express his opinion on the adoption, but the case was ultimately decided by the adoption agency.

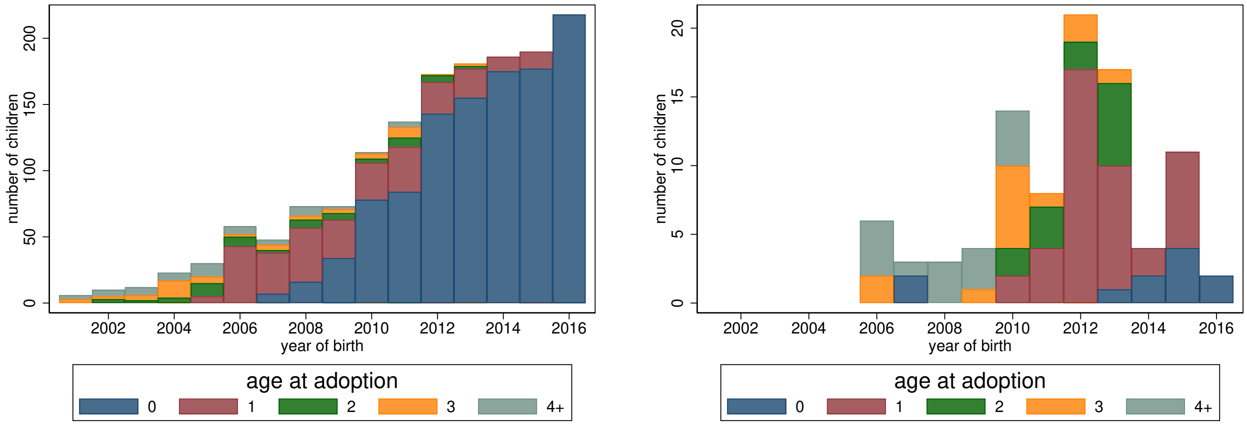
In practice the increasing use and availability of assisted fertility treatments among lesbian couples challenged this 5-year rule, as planned children of lesbian couples conceived through assisted fertilization abroad became increasingly common. Therefore, in 2006 the Norwegian government clarified the rules so that the 5-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 in 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 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.[10] Domestic adoption at birth is very rare in Norway,[11] 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 do not observe births or adoptions directly, only 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

[10]The first adoption from abroad to a same-sex couple in Norway happened in the fall of 2017. 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 2013 at the latest, so that foreign adoptions to same-sex couples should not be a relevant option.

[11]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 after 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 mother prefers a heterosexual couple. In practice, this means that this option is not very relevant for same-sex couples, either.



(a) Lesbian couples

(b) Gay couples

Figure 4: Registered children to same-sex couples, by year of birth and age at adoption

Own calculations, based on sample and data described in section3. Age at adoption refers to the age of the child in the year we first observe both parents registered

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 until the following year.

3.1 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 arrival of a child. Data on residency status, date of birth, gender, municipality of residence and links to mother and father comes from the official population register, and is provided on January 1st every year from 2000 onwards. We also have access to a permanent file of links between children and parents, including all residents ever registered in Norway as of 2016. We obtain data on education for 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 are from two sources. The primary data on annual labor market earnings comes from the tax records.

We also observe the primary employment spell as measured in the matched employer-employee register. This covers all public sector jobs and a large majority of jobs in the private sector, and allow us to link individuals to their primary employer. Contracted hours are provided in bracketed intervals only: from 4-19 hours a week, from 20 to 29 hours a week, or above 30 hours a week. No information on wages is reported in the matched employer-employee register. Starting from 2015, the matched employer-employee register was replaced with detailed and precise data on actual hours and wages based on monthly reports from the employers, called *A-ordningen.* We utilize these unique data for some of the analysis. For parental leave and sickness absence spells, we pull data from FD-Trygd, the register of the Norwegian Public Insurance system that provides parental leave benefits.

We construct two main samples which we use throughout the empirical specifications. For *the long sample* 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 case 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. 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.

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 2001-2013. 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 gay and lesbian couples who gets multiple kids in the same year and register different parent status for each child. Lastly, we keep in both samples only couples where the first child appears at ages 22 to 60 for both parents, giving us some time before and after birth to observe earnings.

This leaves us with a main sample of 231,200 heterosexual couples, 535 lesbian couples and 29 gay couples, and a long sample of 721,291 heterosexual couples. We match these mothers and fathers to their labor market earnings and primary employment relation in all years from $t-4$ to $t+15$, centered around the birth of the first child, to investigate labor market response to child birth. Note that for children born after 2001, we will not see a full 15 years of income after birth because our data ends in 2016. Since most children born to same sex couples are born after 2001, we see later labor market outcomes much 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 2. 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 will be very important to flexibly control for income and initial income gaps in order to compare the child penalty between similar heterosexual, lesbian and gay couples. 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 at adoption is slightly delayed for lesbian couples compared to heterosexual couples (it takes some time for the co-mother to be legally registered in some cases).

Table 2: 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		2001-2013		2001-2013		2001-2013	
A: Child characteristics								
Birth year	1992.0		2007.2		2010.1		2011.7	
	(11.5)		(3.72)		(2.72)		(1.52)	
Multiple birth	0.015		0.020		0.071		0.24	
	(0.12)		(0.14)		(0.26)		(0.44)	
Female child	0.49		0.49		0.47		0.52	
	(0.50)		(0.50)		(0.49)		(0.49)	
Age at adoption			0.022		0.56		1.45	
			(0.17)		(0.86)		(0.91)	
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	30.2	32.3	32.8	38.5	38.6
	(4.04)	(4.83)	(4.22)	(4.98)	(4.14)	(5.72)	(5.36)	(6.21)
Labor earnings (1,000s)	250.3	346.3	343.5	473.2	481.8	476.7	744.6	837.4
(2017 NOK)	(152.4)	(243.6)	(196.3)	(319.3)	(193.3)	(317.7)	(275.3)	(386.4)
Years of education†	14.2	14.0	15.2	14.5	16.4	16.0	17.3	17.4
	(2.94)	(3.00)	(2.88)	(2.99)	(2.42)	(2.60)	(2.42)	(2.46)
N couples	721,291		231,200		535		29	

Note: Summary statistics on estimation samples constructed as described in section 3. Standard deviations in parentheses. †Available from 1980 and onwards only.

4 Main Results

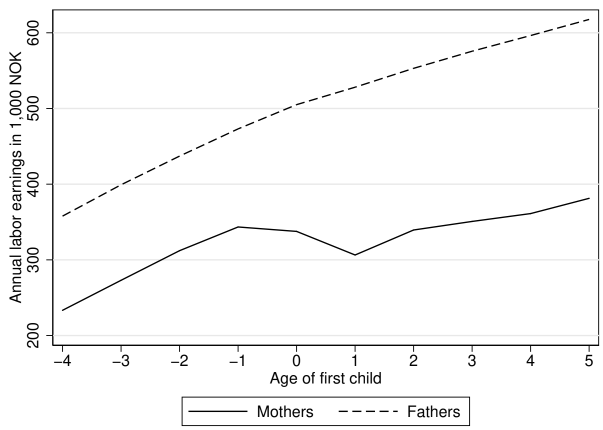
In Figure 3 we present the main results. For each couple type, we present two graphs. The first graph reports raw child penalties. The second graph reports estimates of C_{jk} (see equation 4) generated by the simple event study in equation 3. The results for heterosexual couples are shown in the first row. As has been shown in so many other papers, we also find that mothers in heterosexual couples experience large income penalties upon the birth of their first child while fathers experience no income penalty.

The next row of graphs, corresponding to lesbian couples, is strikingly different. We find that both women experience a child penalty the year after the child is born, but initially the woman who gives birth has a child penalty double the size of her partner. However, 2 years after birth the woman who gives birth catches up and her penalty is no longer statistically significantly different than her partner's. By five years after birth, the child penalty both women experience has largely disappeared.

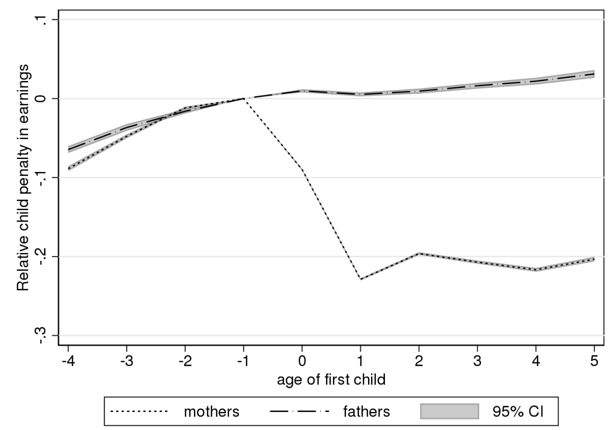
The fact that the lesbian partner who gives birth initially experiences a larger child penalty suggests that biology plays a role in the child penalty for heterosexual couples, but only in the first year after birth. The fact that both partners experience child penalties, and that those penalties are statistically indistinguishable from 2 years after birth onwards, suggest that women do have a preference for children over career. Note that an alternative formulation of the model might assume that 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 both periods, which is not what we find.

The last row of graphs correspond 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 story. In the event study, neither partner experiences a child penalty.

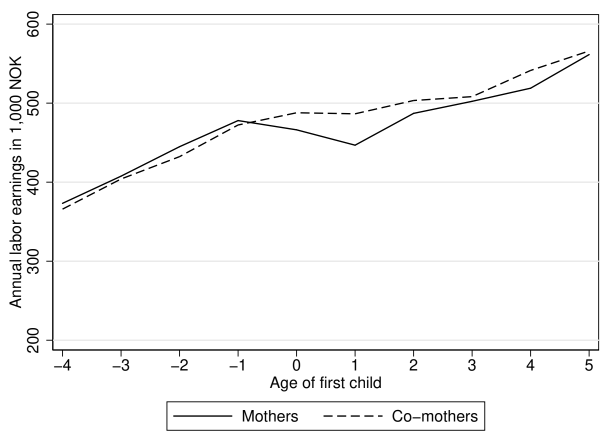
These results are suggestive, but without removing the contribution of specialization we cannot definitively pinpoint mechanisms since the impact of specialization might differ across couple



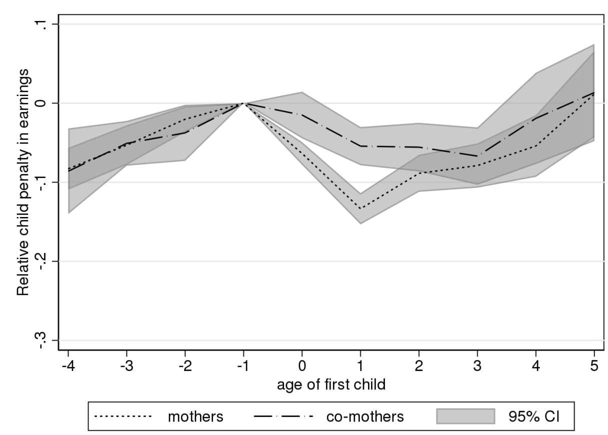
(a) Heterosexual couples



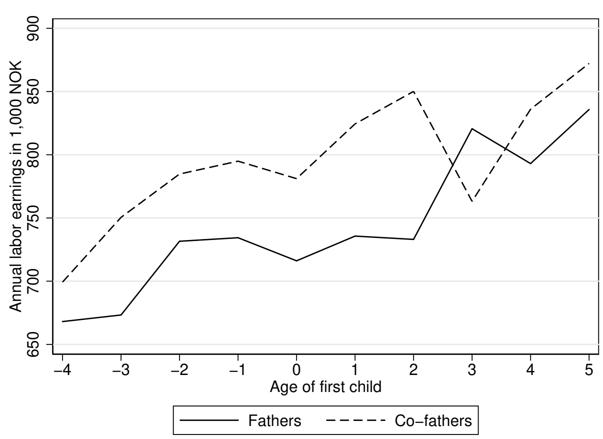
(b) Heterosexual couples



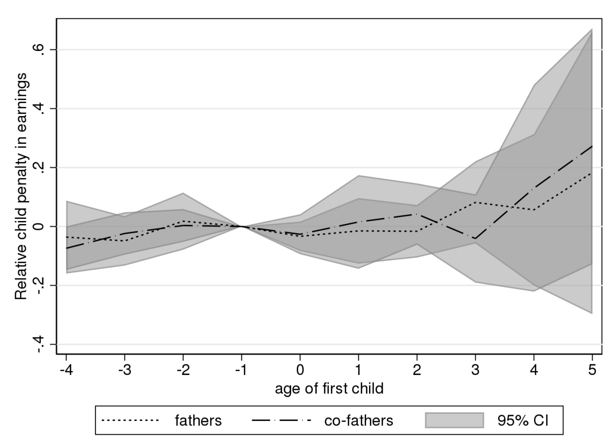
(c) Lesbian couples



(d) Lesbian couples



(e) Gay couples



(f) Gay couples

Figure 5: Raw child penalties (left) and estimated child penalties (right) across couples types

Note: Left panels show means of annual labor earnings for the years before and after birth of the first child. Right panels show the estimated child penalties from equation 3 for a) and b) heterosexual couples, c) and d) lesbian couples and e) and f) gay couples. Samples sizes are 231,200 heterosexual couples, 535 lesbian couples and 29 gay couples. Sample construction and data as defined in section 3. 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.

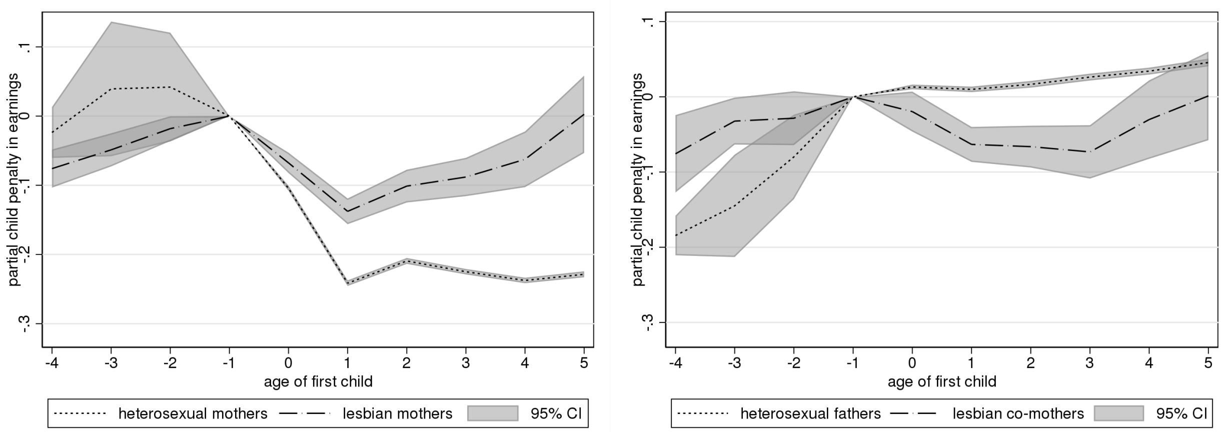


Figure 6: Partial child penalties, controlling for comparative advantage

types. Next, in Figure 6we report estimates conditional on household specialization corresponding to equation 7. Note that this figure represents the remaining child penalty after removing the portion of the penalty explained by comparative advantage. If anything, the differences become even more stark.

4.1 Robustness checks

One major assumption of the model is that η is identical for lesbian and heterosexual women. This is a strong assumption. If η is larger for lesbian women, then the contribution of gender norms is even larger. If η is smaller for lesbian women, then the results could be driven only by preferences, and not be gender norms. This is an untestable prediction. However, if the average η is smaller for lesbian women then for all heterosexual women, then we might also expect η to differ among different groups of heterosexual women. Below we graph the event studies separately for high educated and low educated women, and for high and low income women. [IN PROGRESS].

4.2 Overall impact on household income

The child penalty experienced by women in heterosexual couples is so large, it would seem to imply an overall household income penalty. In Figure 7 we show this is the case. Moreover, we show that the household income penalty experienced by lesbian couples is much smaller in later years. We exclude gay couples from this analysis due to the small sample size of gay couples,

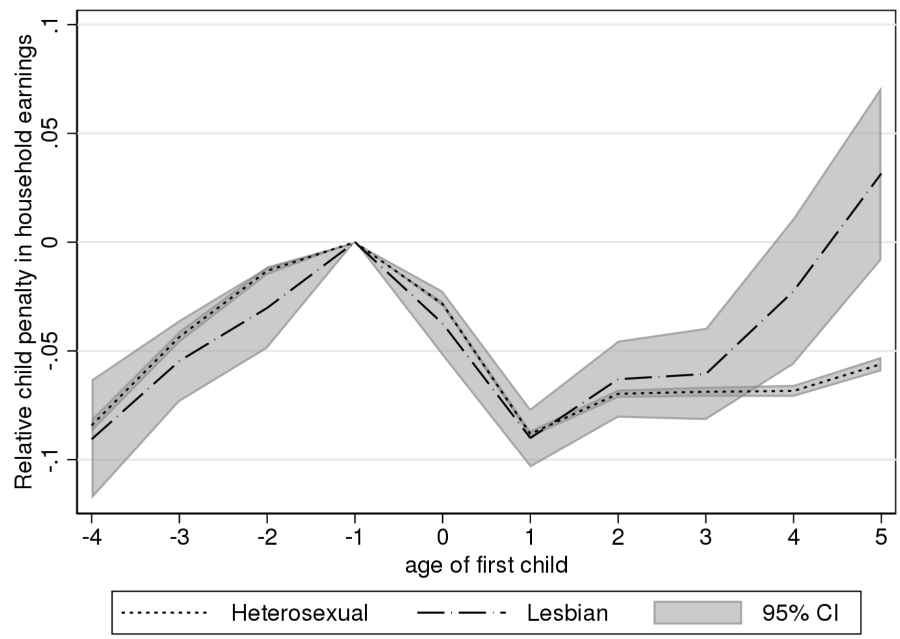


Figure 7: Child penalty, total household income

but as you would expect based on the previous figures, gay couples experience even smaller household income penalties compared to lesbian and heterosexual couples.

4.3 Child outcomes

We have shown that individuals in same sex couples share the burden of child rearing much more evenly, and experience less severe household income penalties compared to heterosexual couples. However, this approach may not be optimal from a social planner's perspective if the reduction in the relative 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. Last, same sex couples could be investing less in their children, in which case their children would be worse off. In Table 3 we present the test scores at age 10 for the children of heterosexual and same sex couples. We find that the children of same sex couples perform better on these tests, and the estimates are significant at the 99th percentile for both reading and

English. 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.

5 The impact of family friendly policies

The results thus far suggest that the 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 improves child outcomes. Despite the persistence of the child penalty within heterosexual couples, history suggests that decreases in the child penalty are possible. In Figure 8 we graph the child penalty of women and men in heterosexual couples from 1980-2009. Note that each line represents the child penalty for children born during a four year interval, estimated using the event study approach from the previous sections (see equations 5 and 3).

The figure shows that the child penalty for women has declined substantially over time. In the 1970s and 1980s fathers not only didn't experience a child penalty, but actually received an increase in income as a result of having their first child. However, over time this child bonus for fathers has decreased, and currently fathers largely experience no change in income following the birth of their first child. Combining the two groups, while the reduction in the relative child penalty has been substantial from the 1970s to the current day, 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 different policies on the relative child penalty: paid paternity leave and improved access to formal child care. Paid paternity leave targets the father's participation in child rearing, while formal child care offers the state as a "third partner" to share the burden of child rearing with mothers.

5.1 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 1992,

Table 3: Impact on the children: Test scores at age 10

	Math	Reading	English
Female child	-0.106***	0.154***	-0.0252***
	(0.00342)	(0.00344)	(0.00353)
Same sex parents	0.105	0.280***	0.209***
	(0.0834)	(0.0761)	(0.0808)
Controls			
Age (mother × father)	✓	✓	✓
Education level (mother × father)	✓	✓	✓
Pre-birth income (mother, father, interacted)	✓	✓	✓
Year	✓	✓	✓
Observations	303,490	302,468	302,670
Children of lesbian couples	134	133	133
Children of gay couples	4	4	4

Note: Cross sectional regressions of test scores on couple types, controlling as indicated. Sample consist of all children born 2001-2007 in the main sample described in Section 3, before conditioning on the first child or the age of the parents at first birth. Standard errors in parentheses are clustered at mother and father using two-way clustering. Test scores are normalized within course and year to have mean zero and standard deviation 1.

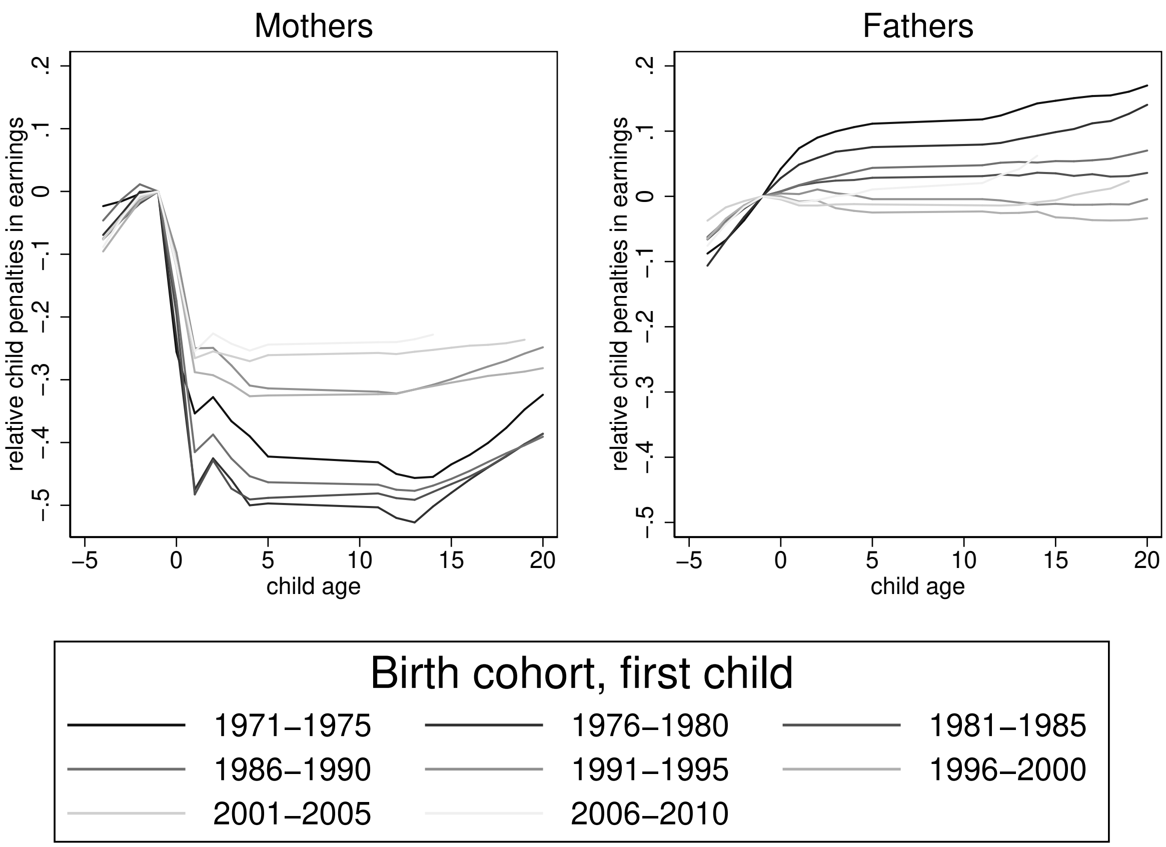


Figure 8: The relative child penalty in income over time for mothers and fathers in heterosexual couples

Norway mandated a four week period of parental leave for fathers. 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 children (increasing β) and might also decrease the distaste fathers have for mothers working outside the home (reducing α). Paternity leave could also increase the productivity of fathers in home production (increasing k_a). Within the framework of our model, all of these effects would decrease the relative child penalty.

A number of papers have estimated the impact of paternity leave policies on different outcomes.[12] A few particularly relevant ones include (Cools *et al.*, 2015; Kotsadam and Finseraas, 2013) who find positive impacts of the Norwegian paternity leave policies on child outcomes looking at the 1993 reform. Dahl *et al.* (2014) find substantial peer effects of the Norwegian policy in 1993 using a regression discontinuity approach. 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 increases mother's earnings but has 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 (2014) finds a large and persistent change in the division of household labor from a Canadian daddy quota. This selection of papers from a broader literature captures the fact that the existing literature 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.

We add to this literature by looking specifically at the impact of paternity leave on the child penalty in the context of our event study design from the first half. In Table 4 we report every leave reform in Norway from 1992-2013. The maternal and paternal quota columns report the amount of parental leave 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. Paid ma-

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

Table 4: Parental leave reforms in Norway

Reform date	Total leave	Compensation	Maternal quota in weeks	Paternal quota (weeks)	Shared leave (weeks)	Max weeks mother
April 1st, 1992	35 (44.4)	100% (80%)	8 (2 before birth)	0	27	35
April 1st, 1993	42 (52)	100% (80%)	9 (3 before birth)	4	29	38
July 1st, 2005	43 (53)	100% (80%)	9 (3 before birth)	5	29	38
July 1st, 2006	44 (54)	100% (80%)	9 (3 before birth)	6	29	38
July 1st, 2009	46 (56)	100% (80%)	9 (3 before birth)	10	27	36
July 1st, 2011	47 (57)	100% (80%)	9 (3 before birth)	12	26	35
July 1st, 2013	49 (59)	100% (80%)	17 (3 before birth)	14	18	35

ternity leave is conditional on the mother working a sufficient amount and receiving a minimum income in the months prior to giving birth (with some variation in the amounts across reforms). For both parents to receive paid leave, both parents must work a minimum amount and receive a minimum income in the months prior to birth. 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 to manipulate birth dates around the cutoff. In the robustness checks Figure 12 we show that there is no statistically significant change in the density of births just before and just after the reform.

In this paper, we estimate results using the 1993, 2005, 2006, 2009, 2011, and 2013 reforms. Because we are looking at short term impacts of children, we use the same regression discontinuity design as Dahl *et al.* (2014). As in all regression discontinuity designs, identification relies on continuity in the underlying regression functions. 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 and 2013, and supplement this with couples with first children born in 1993. We further restrict the samples to births around the cutoff using the optimal bandwidth, see below. We begin by estimating the impact of each reform separately. We estimate a fuzzy RD separately for mothers and fathers using the following specification:

$$
\begin{aligned}
y_{it} &= \sum_{t \neq -1} \beta_t \mathbb{1}(x_i > 0) + f(x_i) + \epsilon_{it} \\
L_i &= \gamma_b \mathbb{1}(x_i > 0) + g(x_i) + \eta_{it}
\end{aligned}
$$

Where we regress earnings at event time t for a mother or father with a child born close to the reform in year b and leave takeout L_i on a dummy for the child being born after the reform date and separate local linear polynomials f and g of the running variable on either side of the cutoff date. 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 local estimates around the cutoff. Because the reforms differ in the increase of quota implemented,w e scale the first stage and reduced form estimates to represent the impact of an additional week of quota so that the reforms are comparable. We estimate and report robust bias-corrected confidence intervals (see (Calonico *et al.*, 2014) and Calonico *et al.* (2018)) together with the conventional, heteroskedasticity-robust confidence intervals.[13] For additional details see Cattaneo *et al.* (2018a,b)

We report first stage estimates for these specifications in Table 5. We see clear and significant effects of all reforms except for the 2006 reform, whether using robust bias-correcting inference or conventional inference that only accounts for heteroskedasticity.

We next plot the reduced form and first stage estimates together for each of the five reforms in Figure 9 Despite the strong first stages in the top panels discussed above, 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 and couples who are not. These results imply that paternity leave does not cause fathers to parent more

[13]Many models in this section are estimated using the robust RD commands for Stata written by Matias D. Cattaneo 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.* (2018d).

Reform year	2005	2006	2009	2011	2013	Pooled	Stacked
RD estimate	0.908	0.609	0.978	1.106	0.543	1.195	0.899
conventional standard error	0.474	0.421	0.105	0.325	0.234	0.166	0.089
robust standard error	0.544	0.510	0.125	0.375	0.274	0.189	
conventional p-value	0.0554	0.148	0.000	0.001	0.020	0.000	0.000
robust p-value	0.0441	0.279	0.000	0.001	0.076	0.000	
Robust 95% CI	0.0288	-0.448	0.758	0.480	-0.0507	0.891	
	2.160	1.550	1.246	1.949	1.022	1.630	
Optimal bandwidth	43.10	51.34	72.00	33.36	69.97	41.43	
Obsevations	17,173	17,714	18,892	18,641	18,649	91,069	91,069
Efficient observations	4,316	5,355	7,704	3,709	7,588	22,292	28,672
Weights in pooled	0.185	0.195	0.205	0.205	0.211		
Weights in stacked	0.151	0.187	0.269	0.129	0.265		
Quota increase	1	1	4	2	2	1.75	2.20

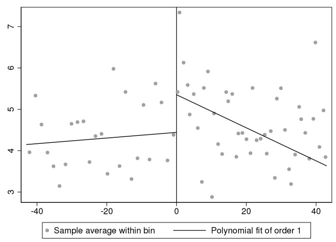
Table 5: RDD first stage estimates

Note: 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, and using the quota in stead of the treatment indicator to secure scaling. Conventional standard errors are heteroskedasticity-robust, but not bias-corrected.

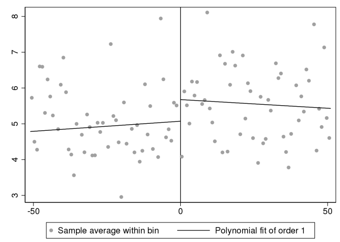
equally with mothers, at least not in such a way that mothers experience less severe child penalties. Estimates are, however, very imprecisely estimated.

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 recenter the running variable to be zero at the relevant cutoff for all individuals and run semiparametric RD estimates in the pooled sample. We call this the pooled estimate, and report the first stage specification for this procedure in Table 5 above. 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 the precision of our estimates. An alternative and more straigthforward 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. 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 relatively small, indicating that the variance coming from the approximation error is relatively small. Second, the approximation error should be smaller for the stacked than the pooled specification because we allow the local polynomials to differ between cutoffs. Nonetheless, inference from this specification is only correct if the model is well specified, so that approximation error vanishes asymptotically. We test the robustness to the functional form in the next section.

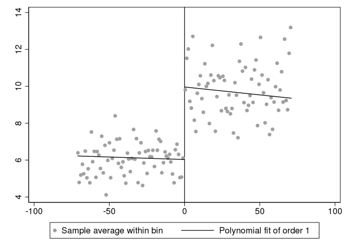
The last two rows of Table 5 reports the results from these two specifications. The pooled estimate is - somewhat surprisingly - larger than any of the cutoff-specific estimates, indicating more than a one week increase in leave use per week increase in the paternity leave quota. Second, although the estimate is highly significant, notice that the standard errors of the pooled estimate are still larger than the most precisely estimated individual cutoff. In contrast, the



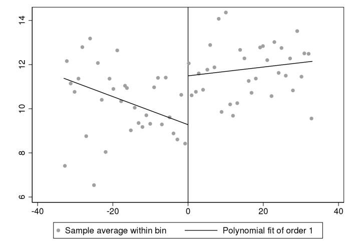
(a) First stage 2005



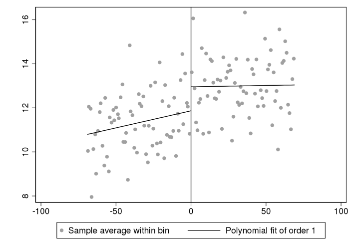
(b) First stage 2006



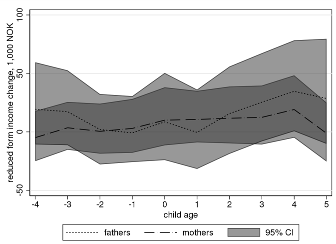
(c) First stage 2009



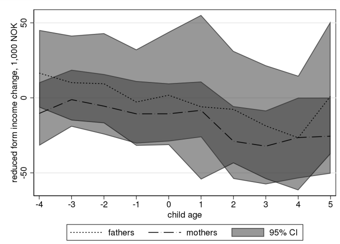
(d) First stage 2011



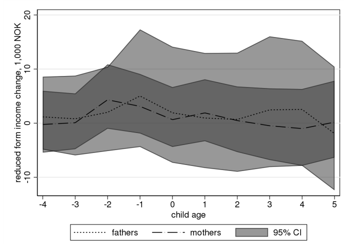
(e) First stage 2013



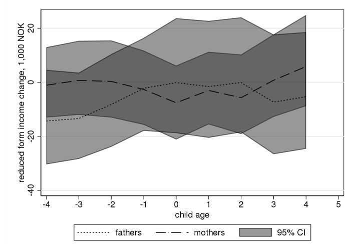
(f) Reduced form, 2005



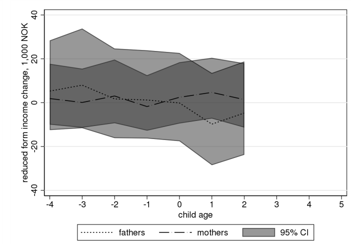
(g) Reduced form, 2006



(h) Reduced form, 2009



(i) Reduced form, 2011



(j) Reduced form, 2013

Figure 9: Robust RDD estimates, paternity leave reforms

Note: Top panels 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. Bottom panels shows sharp RD estimates of the impact of an additional week of the reforms on maternal and paternal income by year, scaled to represent the effect of one week of increased paternity leave quota. Optimal MSE-reducing bandwidths, triangular kernel and local linear polynomials on either side of cutoff. Confidence intervals are robust and bias-corrected.

stacked specification delivers improved precision over any of the individual estimates, finding a more reasonable .9 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' subsequent labor supply using the pooled and stacked estimates from above. For the stacked estimates, we revert to the cutoff-specific treatment indicators as instruments because the fuzzy RD takes care of the scaling across reforms for us. This specification exactly reproduces the cutoff-specific first stage estimates and so is a natural way to stack the reforms. The results from the stacked and pooled fuzzy RD estimates for mothers and fathers are presented together with the first stage in Figure 10. The top panel illustrates how the various reforms affected paternity leave takeout, mirroring the estimates from table 5. The bottom two figures presents the impacts on mothers and fathers yearly incomes over time. The estimates are flat and centered at zero, confirming the findings from before of little impact of paternity leave use. The stacked estimates do, however, provide more precise estimates than the results using separate reforms, particularly for mothers using the stacked specification, ruling out positive impacts larger than around 5,000 NOK per week of paternity leave use for all years post-birth.

Last, to provide a sense of the potential percentage change in the child penalty, we re-scale the figures so that the$y$-axis to represent the percentage of the child penalty, as calculated in the first half of the paper. We present these results in Figure **??**. While point estimates are as before close to zero, the lower bound of the effect is still informative. If we take the maximum reduction in the child penalty after age 1 we find that one week of paternity leave at most decrease the mother's child penalty by around 5%, and estimates are similar for all years after age 1. Similar estimates for fathers are too imprecise to draw firm conclusions, in part because the initial child penalty is very close to zero.

The results in this subsection cover a variety of different paternity leave expansions. Despite the number of reforms we study and the strength of the first stage for many of the reforms, we never find a statistically significant impact of paternity leave on the child penalty. Based on these results, we conclude that paternity leave does not appear to reduce the relative child penalty. Given that this frequently suggested program targeting fathers does not appear to affect the child penalty, we now turn to the alternative approach available to governments: directly support mothers.

5.1.1 Robustness Checks

The critical assumption for the validity of our RD tests is that the population of couples around the discontinuity are identical. In Table 6 we report estimates that show that on observables, individuals around the cutoff are statistically indistinguishable from each other with only a few exceptions.

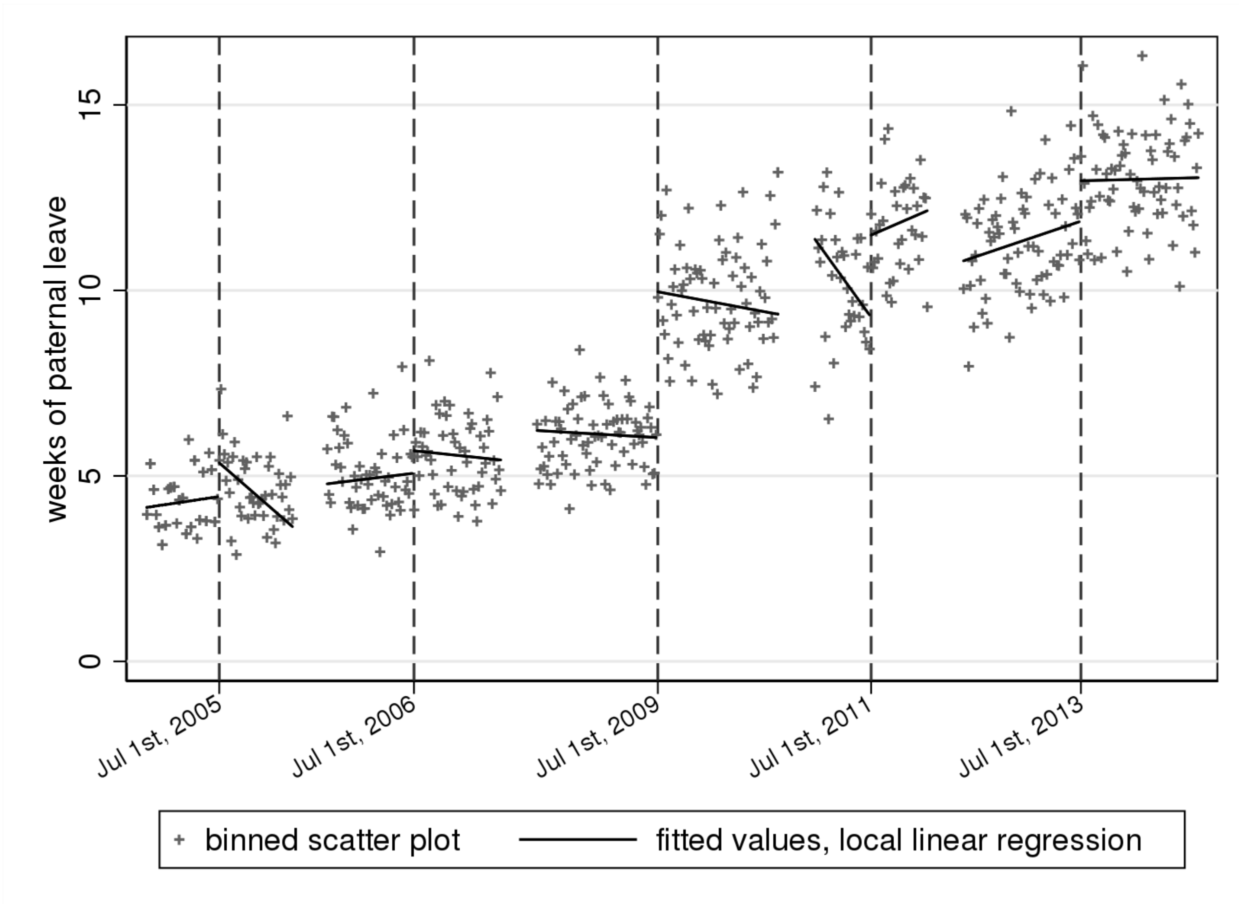
Additionally, we can test for manipulation around the cutoff. If parents were able to manipulate either conception or birth at the cutoff in order to qualify for reforms, then we would expect a statistically significant change in the density of births around the cutoff. In Figure 12, we show this is not the case.

5.2 Improved access to formal child care

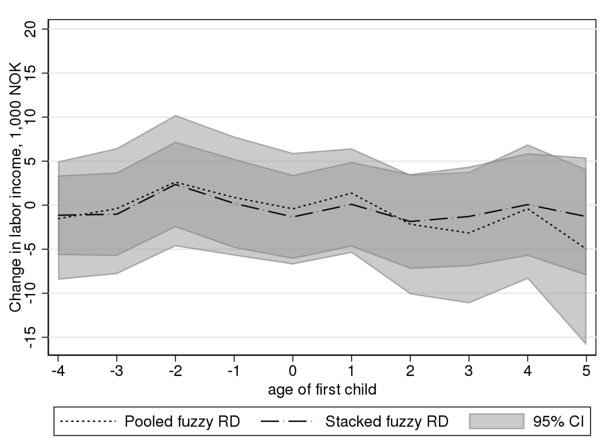
[IN PROGRESS]

6 Conclusion

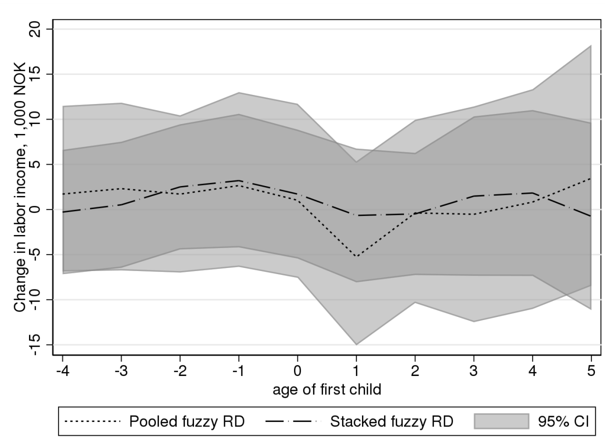
[IN PROGRESS]



(a) First stage estimates



(b) Mothers



(c) Fathers

Figure 10: Pooled and stacked fuzzy RD estimates

Notes:

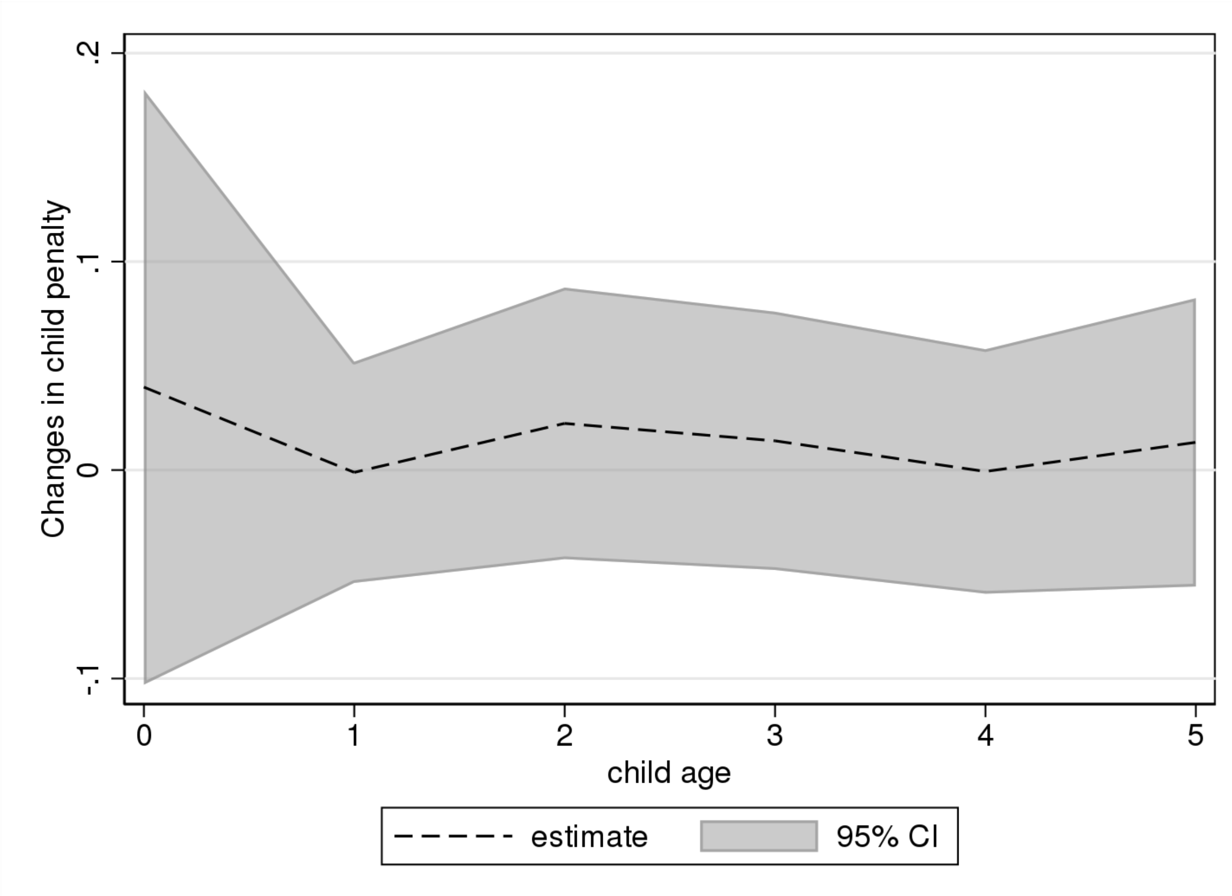
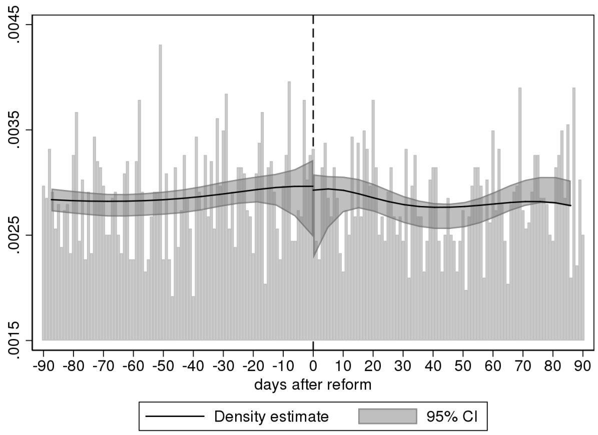


Figure 11: Scaled stacked fuzzy RD estimates, mothers

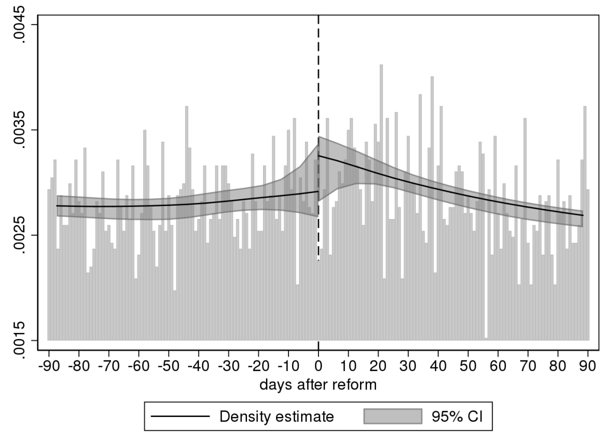
Note: Figure shows the stacked fuzzy RD estimates for mothers, scaled by the estimated child penalties from the baseline so that the estimates can be interpteted as the relative increase or decrease in the child penalty per week increase in paternity leave takeout.

Table 6: Sharp RD balancing tests

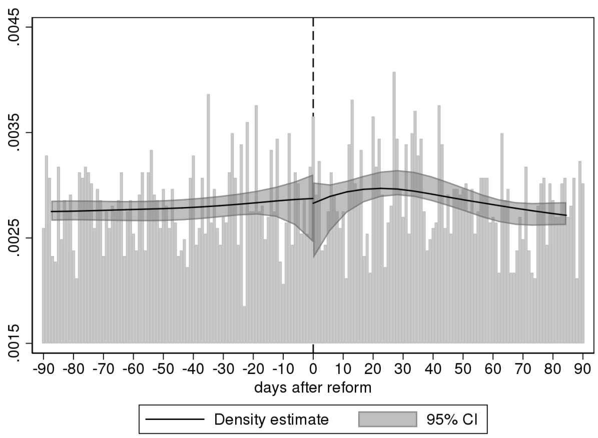
Variable	Reform year	2005	2006	2009	2011	2013	Pooled	Stacked
Father's	RD estimate	0.0893	-0.247	0.0229	-0.108	0.170	-0.00571	0.0256
age	conventional s.e.	0.259	0.253	0.0786	0.158	0.159	0.0918	0.0436
	robust s.e.	0.259	0.253	0.0786	0.158	0.159	0.0918	
	conventional p	0.687	0.247	0.727	0.412	0.204	0.941	0.556
	robust p	0.584	0.277	0.850	0.402	0.199	0.821	
Mother's	RD estimate	0.173	-0.0174	0.0442	-0.202	0.294	0.0816	0.0731
age	conventional s.e.	0.342	0.297	0.0936	0.189	0.187	0.111	(0.0515)
	robust s.e.	0.342	0.297	0.0936	0.189	0.187	0.111	
	conventional p	0.547	0.944	0.571	0.197	0.0600	0.378	0.156
	robust p	0.506	0.973	0.671	0.294	0.103	0.441	
Maternity	RD estimate	2.050	-1.692	0.0702	0.00876	1.187	0.117	-0.0464
leave	conventional s.e.	1.614	1.401	0.364	0.607	0.540	0.337	0.193
takeout	robust s.e.	1.614	1.401	0.364	0.607	0.540	0.337	
	conventional p	0.135	0.157	0.825	0.986	0.0145	0.691	0.810
	robust p	0.141	0.142	0.605	0.957	0.00979	0.506	
Father's	RD estimate	0.178	-0.477	0.0242	-0.0137	0.00835	-0.0255	0.0137
years of	conventional s.e.	0.214	0.215	0.0488	0.0963	0.104	0.0653	(0.0306
education	robust s.e.	0.214	0.215	0.0488	0.0963	0.104	0.0653	
	conventional p	0.322	0.0101	0.552	0.865	0.923	0.645	0.655
	robust p	0.490	0.0123	0.680	0.819	0.964	0.553	
Mother's	RD estimate	0.00107	-0.516	0.0519	-0.0263	0.0797	-0.0140	0.0299
years of	conventional s.e.	0.219	0.221	0.0532	0.105	0.0934	0.0626	0.0319
education	robust s.e.	0.219	0.221	0.0532	0.105	0.0934	0.0626	
	conventional p	0.995	0.00772	0.241	0.762	0.314	0.789	0.349
	robust p	0.898	0.00785	0.297	0.765	0.295	0.748	



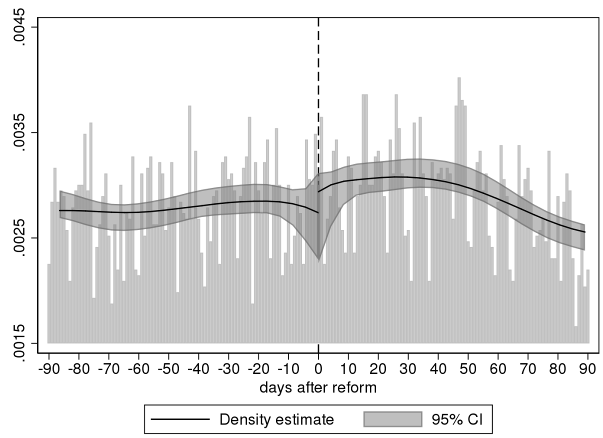
(a) 2005, $p = 0.640$



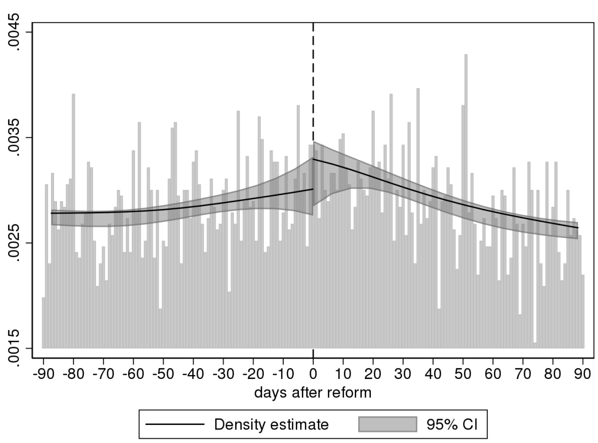
(b) 2006, $p = 0.750$



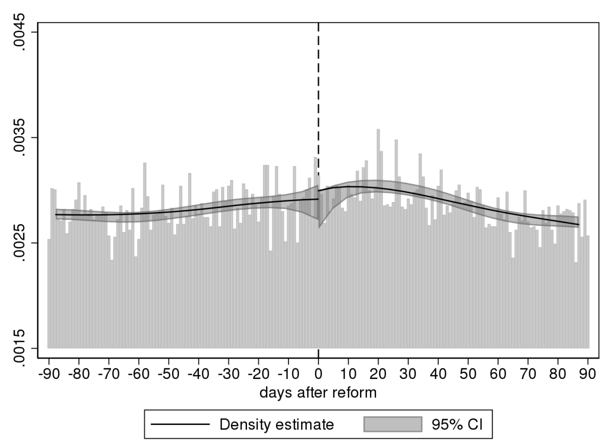
(c) 2009, $p = 0.662$



(d) 2011, $p = 0.783$



(e) 2013, $p = 0.635$



(f) All reforms, $p = 0.536$

Figure 12: Density plots below and above cutoffs

Notes: Graphs show density estimates above and below the cutoff using methods described in Matias D. Cattaneo and Ma (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.

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.

Andresen, M. E. and Havnes, T. (2018). Child care, parental labor supply and tax revenue.

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. (2016). *Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?* Tech. rep., IZA Discussion Papers.

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

Bergsvik, J., Kitterod, R. H. and Wiik, K. A. (2018). Parenthood and couples' relative earnings in norway 2005-2014. *Working Paper*.

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

—, RocioTitiunik and Vazquez-Bare, G. (2018d). Analysis of regression discontinuity designs with multiple cutoffs or multiple score.

Chung, Y., Downs, B., Sandler, D. H., Sienkiewicz, R. *et al.* (2017). *The Parental Gender Earnings Gap in the United States*. Tech. rep.

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

Dahl, G. B., LÃžken, 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.

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.

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.

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

Kotsadam, A. and Finseraas, H. (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 ZweimÃŒller, 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 ZweimÃŒller, 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.

Matias D. Cattaneo, M. J. and Ma, X. (2017). Simple local polynomial density estimators.

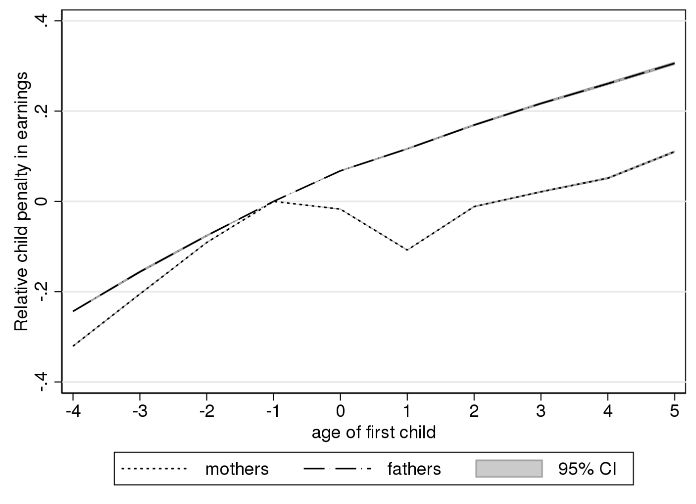
OECD (2014). OECD family database: PF2.1 - key character of parental leave systems. available at `http://www.oecd.org/els/soc/PF2_1_Parental_leave_systems_1May2014.pdf`.

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.

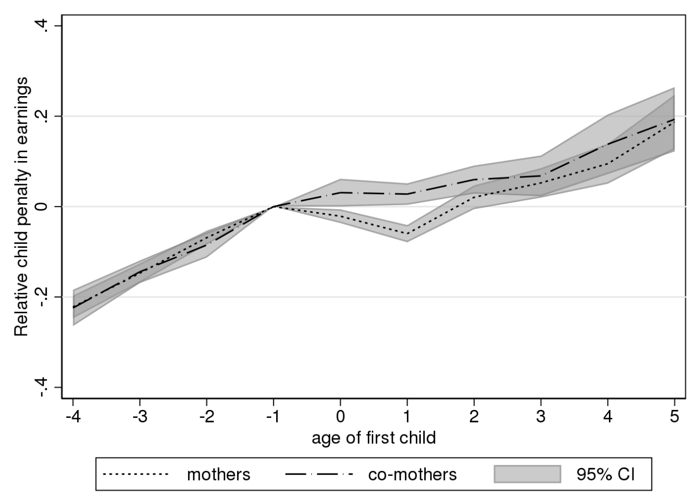
Patnaik, A. (2014). Reserving time for daddy: The short and long-run consequences of fathers' quotas, unpublished. Available at `https://dx.doi.org/10.2139/ssrn.2475970`.

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

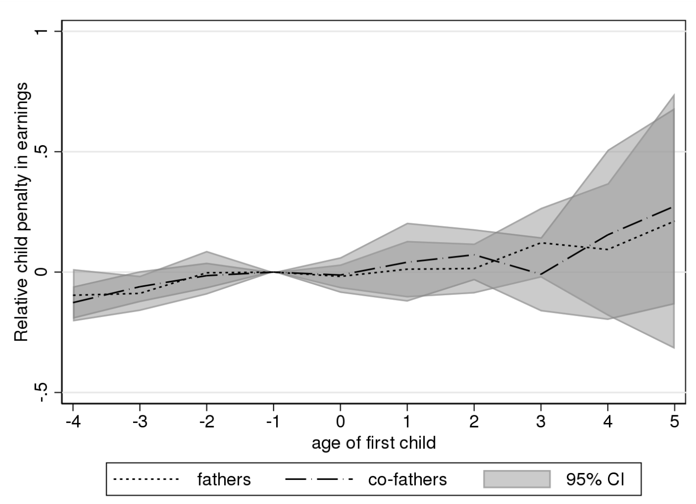
A Additional results



(a) Heterosexual couples



(b) Lesbian couples



(c) Heterosexual couples

Figure 13: Event studies with no age and year controls