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

Martin Eckhoff Andresen∗ and Emily Nix†

PRELIMINARY DRAFT - PLEASE DO NOT CITE OR CIRCULATE‡

Abstract

While the gender gap in income has narrowed over the past 50 years, women in heterosexual couples continue to experience significant reductions in labor market income following the birth of children, while their partners experience no such income drops. This “relative child penalty” has been well documented and accounts for the majority of the remaining gender gap in income in numerous 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 biological cost of giving birth. In this paper, we document child penalties for Norwegian heterosexual and same sex couples. We find that same sex couples experience significantly different relative child penalties compared to heterosexual couples. We develop a simple economic model that incorporates the main explanations for the relative child penalty and generates testable predictions. The model, combined with the empirical results, suggests that the child penalty experienced by women in heterosexual couples is primarily explained by preferences and gender norms, with a smaller contribution due to biological costs of birth. In the second half of the paper we provide causal estimates on the impact of two family policies aimed partly at reducing the child penalty: paternity leave and subsidized early child care. Our precise and robust regression discontinuity results based on six paternity leave reforms show no significant impact of paternity leave use on the relative child penalty, nor on the use of paternity leave for subsequent kids. Early subsidized care seems to have more promise as a policy tool for affecting child penalties, as we find imprecise positive impacts from a large Norwegian reform that expanded access to high quality child care.

∗Statistics Norway, mrt@ssb.no
†University of Southern California, enix@usc.edu
‡We thank seminar participants at University of Rochester, Claremont McKenna University, Statistics Norway, RAND, Arizona State University, and Louisiana State University. We also thank Kenneth Aarskaug Wiik, Edwin Leuven, Matias D. Cattaneo, Sebastian Calonico, Heather Antecol, Adam Sheridan and Nina Drange for helpful comments and suggestions.
**JEL-codes:** J13, J22

**Keywords:** Gender wage gap, labor supply, child penalty, paternity leave, same sex couples, event study, regression discontinuity
1 Introduction

The gender gap has narrowed significantly over the past 50 years, as documented by Blau and Kahn (2000). However, one component of the gender 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” and has 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. (2018), Kuziemko 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.

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

---

1 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 (see, for example, Olivetti and Petrongolo (2016)).

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, 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.
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, 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 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 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. 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, a prediction of the model is that if women have greater preferences for child care, the child penalties for lesbian mothers will be smaller than the child penalty 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 a 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 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. While the initial household income penalty experienced by lesbian couples on the birth of the first child is identical to 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 to control for comparative advantage differences across couple types. The results are unchanged after implementing the first approach, and the second approach is in progress.

These patterns suggest that while biology plays a small role in the relative child penalty, 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 rule out the possibility that all of the differences are in fact driven by greater preferences for child quality among heterosexual couples. We show that while this change to the model is 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.

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. Policymakers 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. In the second half of this paper, we estimate the impact of two proposed family policies aimed partly at reducing the relative child penalty: paternity leave quotas 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 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. Scaling the results, we can rule out reductions of an extra week of paternity leave on mothers’ earnings larger than around 5 to 7 per cent of the child penalty. Our results indicate that paternity leave does not change father preferences or norms for child care in a way that affects the child penalty experienced by fathers or mothers in the short run. However, it may be that the impact of paternity leave policies take some time to materialize, or that preferences of treated fathers have changed and they are spending more leisure time with children, which does not show up in income. 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 bearing in a way that does not show up in either spouse’s earnings, we could expect to see paternity leave take up change for subsequent children as a result of the use of leave for the first child.
Again, however, 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 around 0.1 additional week of leave 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 incomes over time. The market for care for toddlers was severely rationed before this reform, which increased subsidies to the investment in and running of child care institutions. 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 the individual child care use with the rationed, municipality-level availability of slots as in Andresen and Havnes (2018). Results indicates positive effects on mothers’ labor income at ages 2 and 3 that scales to reduce the child penalty by around 40% for each additional full year of early child care use.

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. (2018), and Angelov et al. (2016) to identify child penalties. 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 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 understand 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 the women report more negative opinions toward female employment after giving birth relative to before birth.

4Lundborg 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.
Our paper 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. Schneebbaum (2013) also finds that lesbian women earn more than heterosexual women, but focuses on the differences between primary and secondary earnings, 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 childcare 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) and time use (Martell and Roncolato, 2016) for same sex couples.

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\(^5\), 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. 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

\(^5\)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.
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 flexibly control for income, age, education,
municipality of residence and other factors that could affect child outcomes. 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. Pure
differences without controls, most relevant for the oral arguments raise in Obergefell v. Hodges,
are even larger, strongly significant and in the range of .4 to .6 standard deviations in all three
subjects.

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.\textsuperscript{7} A
few particularly relevant ones include Cools et al. (2015) and 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, which is related to how we investigate impacts of (own)
leave use for subsequent kids is affected by initial exposure to the reforms. 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 (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, pointing at least to the possibility that it may
decrease child penalties. We add to this literature by exploiting six consecutive paternity leave

\textsuperscript{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).
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. (2016) 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. (2016) 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 (2018) on which we build, while Havnes and Mogstad (2011) find no effects of a similar expansion of care for older kids in Norway in the mid 70’s. Other related papers in this field are nicely summarized in e.g. Blau and Currie, 2006; Akgunduz and Plantenga, 2018; Bauernschuster and Schlotter, 2015; Morrissey, 2016. In our paper we focus on the impact of access to child care on the individual and relative child penalties experienced by men and women within heterosexual couples. Consistent with a number of other papers in this literature finding positive impacts of child care access on female labor market outcomes, we find that increased access to child care reduces the relative child penalty by reducing the individual child penalties experienced by heterosexual women.

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 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, 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). 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:

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

where $c$ is consumption and $\theta$ 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. $\beta$ represents the value of child quality and $\eta$ is the additional utility women get from being at home with children, capturing potential differences in gender preferences over time with children. $\alpha$ is the disutility men get from each hour their wife works when they have children, capturing gender norms around child care.9

---

9We 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.

9Survey 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

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

where $\bar{S}$ is an indicator equal to 1 if the spouse is a woman who gave birth, and $\delta$ 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

$$\theta = k_a h (1 - t_a) + k_b h (1 - t_b)$$

where $k_i \geq 0$ are productivity parameters, $h' > 0$, $h'' \leq 0$, and $h(0) = 0$.

The household maximizes utility by choosing each spouse’s division of labor in periods 2 and 3, where household utility is given by

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

and $\lambda_i$ 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. 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 $\lambda_a$ 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,

---

10 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:

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

(1)

The first order conditions for the second and third periods are given below.

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

The wage equations given by the first order conditions yield the following predictions:

1. **Preferences**: The income penalty is increasing for all women as $\eta$ 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 $\eta$ is smaller than the increase in the income penalty for heterosexual women. The relative child penalty for heterosexual couples is increasing in $\eta$ 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, $\eta$ 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 $\delta$ increases. The relative child penalty for lesbian and heterosexual couples is increasing in $\delta$ at an increasing rate if $h'' < 0$ and at a constant rate otherwise. $\delta$ 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 $\alpha$ increases (and the income penalty for heterosexual men is decreasing). The relative child penalty for heterosexual couples is increasing in $\alpha$ at an increasing rate if $h'' < 0$ and at a constant rate otherwise. By construction, $\alpha$ 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 Table 1 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 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. 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. We also report simulations demonstrating the impact of each of the other mechanisms in Figure 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 1 as we vary each parameter ($\eta$, $\delta$, $\alpha$, and $w_a$) individually.
Table 1: Summary of the predictions of the model

<table>
<thead>
<tr>
<th>Individual Child Penalty</th>
<th>Heterosexual</th>
<th>Lesbian</th>
<th>Gay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferences ($\eta$)</td>
<td>Female spouse</td>
<td>Both spouses (&lt;hetero)</td>
<td>Neither spouse</td>
</tr>
<tr>
<td>Biology ($\alpha$)</td>
<td>Female spouse</td>
<td>One spouse</td>
<td>Neither spouse</td>
</tr>
<tr>
<td>Gender norms ($\delta$)</td>
<td>Female spouse</td>
<td>Neither spouse</td>
<td>Neither spouse</td>
</tr>
<tr>
<td>Specialization ($\frac{w_a}{k_a} - \frac{w_b}{k_b}$)</td>
<td>Female spouse</td>
<td>One spouse</td>
<td>One spouse</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relative Child Penalty</th>
<th>Heterosexual</th>
<th>Lesbian</th>
<th>Gay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferences ($\eta$)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Biology ($\alpha$)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Gender norms ($\delta$)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Specialization ($\frac{w_a}{k_a} - \frac{w_b}{k_b}$)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Note: Left panel shows 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.

Figure 1: Model Predictions: Simulations for Preferences and Biology
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 ?? 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 has 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). Borusyak and Jaravel (2016) revisits the identification problem in event study designs, pointing to the problem of aggregating post-event dummies and the impossibility of identifying cohort- or individual fixed effects together with age- and event time dummies, none of which should be problems in our setting.

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:

\[
y_{it} = \sum_{j \neq -1} \sum_{k} \alpha_{jk} \mathbb{I}[t = j, K_i = k] + \sum_{l} \sum_{m} \beta_{lm} \mathbb{I}[age_{it} = l, X_i = m] + \sum_{n} \sum_{o} \gamma_{no} \mathbb{I}[T_{it} = n, X_i = o] + \sum_{p} \eta_{p} \mathbb{I}[K_i = p] + \epsilon_{it}
\]

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, father or co-father in a gay couple or even single mothers. \( \mathbb{I}[A] \) 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. This means that the effect of age and year on income is the same for all women, be they in heterosexual or lesbian couples. Equation 2 is equivalent to running the regressions separately for mothers and father if we only estimate the equation for heterosexual couples.\(^\text{11}\)

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.

\(^{11}\)While it is possible to estimate equation 2 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.
Our objects of interest are $\alpha_{jk}$, the change in the outcome for a parent of type $k$ at child age $j$ 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 2 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.

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

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 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

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 other factors that determine changes in labor income corresponding to the arrival of children are identical across couple types. In particular, our model predicts that the relative productivities in labor market and home production of the two spouses, $\frac{w_i}{k_i} - \frac{w_{i-1}}{k_{i-1}}$, 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 2.

Unfortunately, we observe neither wages nor home productivity. We observe pre-child incomes, $I_{it}$ and $I_{-it}$. This is sufficient if $k_i = k_{-i}$, 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 $I_{it} - I_{-it}$ 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 (for example, women do more household chores as girls then men), then controlling for $I_i - I_{-i}$ should also be sufficient.

To control for specialization, we flexibly control for the differences in own and spouse’s earnings four years prior to birth interacted with event dummies, by adding $\sum_j \theta_j 1[t = j](I_i - I_{-i})$ to equation 2, 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 $\alpha_{jk}$ 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 2, 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.

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 2 using the weighted sample of matching heterosexual couples, and then use the approach from Altonji, Elder and Taber (2005) to bound the differences in
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. 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 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 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. This change to the guidelines allowed

\[Aarskaug Wiik et al. (2014)\] investigates the stability of these same sex marriages and partnerships, showing that they are less stable than heterosexual marriages.
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 5 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 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 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 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.\textsuperscript{13} Domestic adoption at birth is very rare in Norway,\textsuperscript{14} 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 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 this was first introduced in 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 80’s, reforms that we exploit 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

\textsuperscript{13}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 should not be a relevant option.

\textsuperscript{14}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.
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 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 she’s not eligible for parental leave. Following the parental leave, Norway has a well developed, regulated and highly subsidized child care sector with high coverage. Following the large scale reform from 2002 that we exploit in section 6.2, child care attendance was between 95 and 100% for 3 - 5 year old children in 2017. For 2-year olds, coverage was around 93%, while for 1-year olds the share was around 72%, but the coverage rates have risen sharply for the younger kids since the early 2000’s. 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 to young children who do not use the subsidized formal care system. Because of the heavy subsidies to the formal care, the market

\[15\] 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.
for paid child care outside this system is very rare, but subsidies is available for both private and public suppliers of care.

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 mother and father comes from the official population register, and is provided on January 1st every year from 2000 onward. We also have access to a permanent file of links between children and parents, including all residents ever registered in Norway. 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 pensionable incomes that include taxable benefits such as sickness and parental leave and benefits. 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-2002 data and include only the employment spell with the most contracted hours for the post-2003 data.\footnote{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.} 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 works 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 parental leave spells we measure how many weeks of leave was taken for a particular child, which we need to infer from the start- and stop dates of the leave spells and birth dates of the children because links directly to the child is not available.

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 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, we know exactly which children attended how much care for each month. From these precise data, we construct precise measures of full-time equivalent months of child care use from ages 13 - 36 months.

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, 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.

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 additionally identify a sample of single mothers as the mothers of children who are never registered with a father, and provide results for this sample in the appendix.

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 lesbian couples who gets 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 some time before and after birth to observe earnings.

This leaves us with a main sample of 250,296 heterosexual couples, 634 lesbian couples, 32 gay couples and 4,998 single mothers, 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 \( t + 5 \) or \( t + 15 \), 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 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 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, as it takes some time for the co-mother to be legally registered.

5 Main results

In Figure 4 we present the main results. The graphs report estimates of $C_{jk}$ (see equation 3) generated by the simple event study in equation 2. 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.\textsuperscript{19} 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.

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 12\% 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.

\textsuperscript{19}In Appendix Figure B1 we also report the raw mean earnings by event time for each couple type, without imposing any of the structure from equation 2, and the results are quantitatively similar.
Table 2: Summary statistics by couple type

<table>
<thead>
<tr>
<th>Birth year (first child)</th>
<th>Heterosexual couples</th>
<th>Lesbian couples</th>
<th>Gay couples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Long sample</td>
<td>Main sample</td>
<td>Main sample</td>
</tr>
</tbody>
</table>

A: Child characteristics

<table>
<thead>
<tr>
<th></th>
<th>Heterosexual couples</th>
<th>Lesbian couples</th>
<th>Gay couples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth year</td>
<td>1992.0</td>
<td>2007.7</td>
<td>2010.7</td>
</tr>
<tr>
<td></td>
<td>(11.5)</td>
<td>(4.00)</td>
<td>(2.87)</td>
</tr>
<tr>
<td>Multiple birth</td>
<td>0.015</td>
<td>0.020</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.14)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>Female child</td>
<td>0.49</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(0.49)</td>
</tr>
<tr>
<td>Age at adoption</td>
<td>0.022</td>
<td>0.48</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.81)</td>
<td>(0.94)</td>
</tr>
</tbody>
</table>

B: Parent characteristics, year before birth

<table>
<thead>
<tr>
<th>Parent type (K)</th>
<th>Mother 1</th>
<th>Father 2</th>
<th>Mother 1</th>
<th>Father 2</th>
<th>Mother 3</th>
<th>Co-mother 4</th>
<th>Father 5</th>
<th>Co-father 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at first birth</td>
<td>26.3</td>
<td>28.7</td>
<td>27.8</td>
<td>30.3</td>
<td>32.2</td>
<td>32.8</td>
<td>38.4</td>
<td>38.2</td>
</tr>
<tr>
<td>Pensionable income (1,000s of 2017 NOK)</td>
<td>250.3</td>
<td>346.3</td>
<td>362.7</td>
<td>487.3</td>
<td>488.9</td>
<td>480.0</td>
<td>737.4</td>
<td>813.2</td>
</tr>
<tr>
<td>Years of education</td>
<td>14.2</td>
<td>14.0</td>
<td>15.2</td>
<td>14.6</td>
<td>16.4</td>
<td>16.0</td>
<td>17.2</td>
<td>17.1</td>
</tr>
</tbody>
</table>

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.
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 do have a preference for time with children over career. Note that an alternative formulation of the model might assume that \( \eta \) 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 catchup 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 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 types and we observe quite different distributions of earnings and education before birth between couple types. Next, in Figure 5 we report estimates controlling for household specialization as measured by income differences in \( t - 4 \) interacted with event dummies, as discussed in Section 3.1 above. Note that this figure presents the remaining child penalty after removing the portion of the penalty explained by the income differences. Except for some differences in the impact before birth that is likely caused by autocorrelation of incomes over time, the figures are remarkably similar to the baseline estimates. 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.

The child penalty experienced by women in heterosexual couples is so large, it would seem to imply an overall household income penalty. In Figure 6 we show this is the case by using total income of the two spouses as the outcome. 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
Figure 4: Estimated child penalties across couples types

Note: Figures show the estimated child penalties from equation 2, scaled as described in eq. 3. 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.
5.1 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 amount 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 3. Therefore, the estimates are interpretable as the effects of children at age (in months) $j$, relative to the effect 12 months before birth. Note that unfortunately we cannot look at wages directly due to lack of data (we observe only annual income).

Results are presented in figures 7 and 8. We begin in figure 7 by repeating the baseline estimates, but unlike 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 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 em-
Figure 6: Child penalty, total household income

employment relation. Unlike the baseline outcome of pensionable 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 maternity leave benefits that replace earnings and are included in our income measure. Following, this, employment returns, but stays below -.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 measuring holding 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 20 percentage points reduction, indicates that there is response both on the extensive and intensive margins: Some mothers drop out of the labor force entirely, 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
similar 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 parts of the differences 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 8, 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 recover after 4-5 years. Lesbian co-mothers behave much like their partners after the year of leave, while heterosexual fathers increase total contracted hours somewhat. 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: the leave out share of mothers of children below 15 years of age among the workers who have their primary employment relation with the firm. Both of these measures, however, are defined only for employed people, and since we’ve 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 8. 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.

Finally, our measure of days of sickness absence allows us to address one component of the biological explanation for the child penalties more directly, specifically that 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 should to a limited extent capture short term illness such as seasonal cold or flu. It does, however, 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. 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, but remember that some of this could be causes by subsequent pregnancies. The pattern is relatively similar for both partners in lesbian couples, while heterosexual fathers take slightly more sickness absence after the birth of children than before.

5.2 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, however, if this 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 results for the test scores at age 10 for the children of heterosexual and same sex couples. The leftmost column, corresponding to no controls, indicate that children of same sex couples do much better than children of heterosexual couples, in the range of .4 to .6 standard deviations in the three subjects. Moving left, 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 do around .2

\footnote{Despite this, we occasionally see non-employed individuals in these data. We exclude the few non-employed individuals who are registered with absence spells.}
(a) Total pensionable 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 7: 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 8: Decomposition II: Child penalties for heterosexual (left) and lesbian (right) couples
standard deviations better in both reading and English even when controlling for our large range of observable characteristics. 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 indicate that the stronger positive selection into child bearing and the prevalence of planned children could be driving these differences.

We find that the children of same sex couples perform better on school tests, and the estimates are large and significant for all subjects in most specifications. 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.

6 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. Despite the persistence of the child penalty within heterosexual couples, history suggests that decreases in the child penalty are possible. In Figure 9 we graph the child penalty of women and men in heterosexual couples from 1971-2010. 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 sections (see equations 3 and 2).

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, sometimes termed a child premium. 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 1970s 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 one important policy tool aimed partly at decreasing this gap and increasing the involvement of fathers in child rearing: Paid paternity
Table 3: Impact on children: Test scores at age 10

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.395***</td>
<td>0.410***</td>
<td>0.565***</td>
</tr>
<tr>
<td></td>
<td>(0.0858)</td>
<td>(0.0832)</td>
<td>(0.0800)</td>
</tr>
<tr>
<td></td>
<td>0.363***</td>
<td>0.352***</td>
<td>0.529***</td>
</tr>
<tr>
<td></td>
<td>(0.0853)</td>
<td>(0.0833)</td>
<td>(0.0794)</td>
</tr>
<tr>
<td></td>
<td>0.283***</td>
<td>0.263***</td>
<td>0.433***</td>
</tr>
<tr>
<td></td>
<td>(0.0853)</td>
<td>(0.0836)</td>
<td>(0.0803)</td>
</tr>
<tr>
<td></td>
<td>0.0893</td>
<td>0.146*</td>
<td>0.248***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0835)</td>
<td>(0.0773)</td>
</tr>
<tr>
<td></td>
<td>0.0766</td>
<td>0.170**</td>
<td>0.235***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0838)</td>
<td>(0.0777)</td>
</tr>
</tbody>
</table>

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 per course of test scores on couple type, controlling as indicated. Sample consist 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 mother and father using two-way clustering. Test scores are normalized within course and year to have mean zero and standard deviation 1.
Figure 9: 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 3 and 2. Separate plots by birth cohort, including confidence intervals, is found in Figure figure B6 in the appendix.

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 children.
(increasing $\beta$) and might also decrease the distaste fathers have for mothers working outside the home (reducing $\alpha$). 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 could decrease the relative child penalty.

In Table 4 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 to manipulate birth dates around the cutoff. In Figure C1, 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. We further restrict the samples to births in a window around the reform date 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 and each year using the following specification:

\[
\begin{align*}
    y_{it} &= \beta_t \bar{L}_i + f_t(x_i) + \epsilon_{it} \\
    L_i &= \gamma \mathbb{1}(x_i > 0) + g(x_i) + \eta_{it}
\end{align*}
\]

Where $x_i$, the running variable, is the number of days after the reform date that the child was
### Table 4: Parental leave reforms in Norway

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


born. For the 2014 reform, which decreased the leave quota, the running variable is instead coded as days before the reform. $f_i(x_i)$ 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 local estimates around 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 an 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.\(^{21}\) For details, see Cattaneo et al. (2018a,b) 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 C1 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. Additionally, 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

\(^{21}\)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. (2018c)
Table 5: RDD first stage estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RD estimate</td>
<td>0.973</td>
<td>0.652</td>
<td>0.952</td>
<td>1.116</td>
<td>0.880</td>
<td>0.680</td>
<td>1.105</td>
<td>0.838</td>
</tr>
<tr>
<td>conventional standard error</td>
<td>(0.482)</td>
<td>(0.406)</td>
<td>(0.0988)</td>
<td>(0.317)</td>
<td>(0.300)</td>
<td>(0.179)</td>
<td>(0.147)</td>
<td>(0.0898)</td>
</tr>
<tr>
<td>robust standard error</td>
<td>0.552</td>
<td>0.489</td>
<td>0.117</td>
<td>0.367</td>
<td>0.355</td>
<td>0.216</td>
<td>0.168</td>
<td></td>
</tr>
<tr>
<td>conventional p-value</td>
<td>0.0435</td>
<td>0.108</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>robust p-value</td>
<td>0.0350</td>
<td>0.215</td>
<td>0.000</td>
<td>0.001</td>
<td>0.008</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>17,146</td>
<td>17,684</td>
<td>18,867</td>
<td>18,624</td>
<td>18,656</td>
<td>18,972</td>
<td>109,949</td>
<td>109,949</td>
</tr>
<tr>
<td>Optimal bandwidth</td>
<td>41.48</td>
<td>50.92</td>
<td>75.36</td>
<td>33.95</td>
<td>51.51</td>
<td>68.66</td>
<td>41.78</td>
<td></td>
</tr>
<tr>
<td>Efficient observations</td>
<td>4,125</td>
<td>5,260</td>
<td>8,090</td>
<td>3,703</td>
<td>5,779</td>
<td>7,507</td>
<td>26,874</td>
<td>34,464</td>
</tr>
<tr>
<td>Weights in pooled</td>
<td>0.153</td>
<td>0.162</td>
<td>0.169</td>
<td>0.170</td>
<td>0.175</td>
<td>0.171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weights in stacked</td>
<td>0.120</td>
<td>0.153</td>
<td>0.235</td>
<td>0.107</td>
<td>0.168</td>
<td>0.218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quota increase</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>-4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
the cutoff. In Figure C1 in the appendix, we show graphically that this is not the case, while the \( p \)-values reported in each panel are for a test of equal densities on either side of the cutoff is identical, using methods from Cattaneo et al. (2017, 2018c). None of the tests can reject that the densities are the same.

We report first stage estimates for these specifications in Table 5, separately for each reform. We see clear and significant effects of all reforms except perhaps 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 10, scaling by the quota increase to get estimates that reflect one additional week of quota. 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 equally with mothers, at least not in such a way that mothers experience less severe child penalties. Estimates are, however, 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 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 cutoffs. 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
Figure 10: 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 third 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.
and the robust standard error estimate for the pooled specification is relatively 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 5 reports the first stage results from these two 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. 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 stacked specification delivers improved precision over any of the individual estimates, finding a more reasonable .84 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 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 5 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: 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 particular 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 paternity leave quotas.

The results from the stacked and pooled fuzzy RD estimates for mothers and fathers are pre-
sented 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 5 and showing clear discontinuities at the cutoffs. The bottom two figures presents the impacts on mothers and fathers yearly incomes over time. Notice first, reassuringly, that there is no effect of paternity leave use on pre-birth outcomes, which can be interpreted as a balancing exercise or placebo test. Following birth, the estimates are flat and centered at zero, confirming the findings from before of little impact of paternity leave use. The 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 NOK 5,000 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 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 ??.

While point estimates are as before close to zero, the lower bound of the effect is still informative. We can rule out reductions larger than around 5 - 7% of the 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 is very close to zero.

Results so far provide clear evidence that paternity leave use does have a causal effect on the distribution of market work within the couple for the couples induced to use it by the quota. It might, however, influence gender norms or preferences around the distribution of home work in ways that does not influence labor supply. One possible measure of such norms is the use of paternity leave itself. To investigate whether paternity leave use has a direct effect on fathers choice of spending time with children, we exploit the fact that many of the fathers that have their first child around the time of the reforms subsequently goes on to have more children. We therefore estimate our fuzzy RD model using the father’s leave takeout for subsequent children as the outcome, instrumenting the initial takeout for the first child with the exposure to the reform as before.\footnote{We have estimated these models using a dummy variable for any leave takeout for both the endogenous variable and the outcome, and find zero effects for this specification as well. Results are available upon request.} We cluster standard errors on the father to account for the fact that each father may get multiple kids following the first child that determines treatment. 

47
Figure 11: Main RD estimates of paternity leave use

Note: Top panel shows first stage around each reform date, using local linear polynomials and triangular weights. Bottom panels show fuzzy RD estimates of the impact of paternity leave use on earnings over time, using all 5 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.
Table 6 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 6, we see little evidence of any permanent impact on norms as measured by takeout of paternity leave: Not a single one of our reforms provide any statistically significant results. Focusing on our preferred stacked estimates, the results indicate non-significant effects of .07 more 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 .35. An important caveat for these results is that many fathers are likely constrained to a corner solution for both the first and subsequent kids due to the paternity leave quota. If paternity leave use affects preferences for future paternity leave in a way that would make fathers prefer to take more leave, but not exceed 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
Table 6: Paternity leave norms: Fuzzy RD of paternity leave use on leave for subsequent kids

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RD estimate</td>
<td>0.791</td>
<td>-1.207</td>
<td>0.119</td>
<td>-0.671</td>
<td>-0.690</td>
<td>0.562</td>
<td>-0.0435</td>
<td>0.0692</td>
</tr>
<tr>
<td>conventional std error</td>
<td>(0.615)</td>
<td>(2.733)</td>
<td>(0.166)</td>
<td>(0.606)</td>
<td>(1.294)</td>
<td>(0.574)</td>
<td>(0.214)</td>
<td>(0.147)</td>
</tr>
<tr>
<td>robust std error</td>
<td>0.727</td>
<td>3.292</td>
<td>0.196</td>
<td>0.684</td>
<td>1.517</td>
<td>0.706</td>
<td>0.248</td>
<td></td>
</tr>
<tr>
<td>conventional p-value</td>
<td>0.198</td>
<td>0.659</td>
<td>0.471</td>
<td>0.268</td>
<td>0.594</td>
<td>0.327</td>
<td>0.839</td>
<td>0.638</td>
</tr>
<tr>
<td>robust p-value</td>
<td>0.324</td>
<td>0.562</td>
<td>0.479</td>
<td>0.305</td>
<td>0.476</td>
<td>0.247</td>
<td>0.833</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>20,027</td>
<td>19,994</td>
<td>17,896</td>
<td>13,843</td>
<td>5,690</td>
<td>971</td>
<td>78,421</td>
<td>78,421</td>
</tr>
<tr>
<td>Optimal bandwidth</td>
<td>49.42</td>
<td>45.60</td>
<td>56.34</td>
<td>36.22</td>
<td>63.74</td>
<td>56.53</td>
<td>39.70</td>
<td></td>
</tr>
<tr>
<td>Efficient observations</td>
<td>5,690</td>
<td>5,367</td>
<td>5,801</td>
<td>3,012</td>
<td>2,128</td>
<td>229</td>
<td>18,110</td>
<td>22,227</td>
</tr>
</tbody>
</table>

Notes: Fuzzy RD estimates of the impact of one more week of paternity leave for the first child on the weeks of paternity leave use for subsequent children. Standard errors are clustered by father.

paternity leave expansions. Despite the number of reforms we study and the strength of the first stage for most of the reforms, 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

Figure 13a shows the child care coverage rates over time in Norway, separately by age of the children. As is evident, in the early 2000’s the formal care sector for preschoolers was well developed in Norway, with more than 80% of Norwegian 3 - 5 year olds attending care, following a reform and gradual expansion of care for these children in the 1970’s (Havnes and Mogstad, 2011). For toddlers, however, coverage was much lower at less than 40%, and the market was strongly rationed, as documented in Andresen and Havnes (2018) and evident from surveys on the actual and preferred modes of child care for these children. This was the background for the Childcare Concord from 2002, a broad bipartisan agreement to increase the availability of care for toddlers. Following this, coverage increased rapidly for 1-2 year olds over the next 6 years as shown in Figure 13a.
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 (2018) and to estimate the impact on the child penalty in this section. As shown, the exact timing of expansions was unpredictable from the municipalities’ side, and subject to a range of constraints that were hard to predict. Furthermore, the exact timing of expansion does not seem to be predictable by pre-reform characteristics of the municipalities, 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 long sample born in the years 2000 to 2006, thus being two years around the time of the reform-induced expansions of care. We drop parents with multiple kids, and lock the municipality of exposure to be the one where the child resides at age 2. We do not restrict attention to children with no younger siblings, although this is customary in the literature, because this is likely endogenous. We restrict attention to effects at child ages 0 through 5, using the years before birth as placebo outcomes. This leaves us with a sample of 123,628 couples.\textsuperscript{23}

For this sample, we separately regress earnings for each parent in each year on municipality- and year fixed effects, municipality specific linear trends and our measure of child care use from

---

\textsuperscript{23}Notice that this is less than half the size of the samples of cohabiting mothers and fathers in Andresen and Havnes (2018), giving us less precision due to the focus on first-born children.
the cash for care data. Because child care is endogenous to labor supply, we instrument the care use decision 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:

\[
y_{it} = \pi_k + \tau_{T_it} + \beta \bar{m}_i + \epsilon_{it}
\]

\[
m_i = \bar{\pi}_k + \bar{\tau}_{T_{it}} + \gamma_1 CC_{1k} + \gamma_2 CC_{2k} + \bar{\epsilon}_{it}
\]  

(4)

Where \( \tau_{T_{it}} \) are calendar year fixed effects, \( \pi_k \) are municipality fixed effects and \( m_i \) is our measure of child care use from ages 13 - 36 months from the cash for care data. The instruments are \( CC_{1k} \), the share of slots for 1-year olds in the municipality at age 1 to the population of one year olds, and \( CC_{2k} \), the share of slots for 2-year olds to the population at the same age in the municipality of residence. To increase precision, we also add municipality-specific linear trends to the specification in 4. Notice that this is a slightly simpler specification than Andresen and Havnes (2018), which measure child care use at age 2, does not include municipality linear trends and include a set of controls. We measure the full period of child care use because we are not primarily interested in labor supply at age 2 only, and include municipality-specific linear trends to increase precision. Andresen and Havnes (2018) show that their results are very similar when using linear trends, and also when excluding controls. Standard errors are clustered at the municipality level and robust to heteroskedasticity.

The variation we exploit thus comes from correlation in the deviations of municipality-specific linear time trends in labor supply and expansions of child care as measures by slots. As long as the exact timing of expansions of care are uncorrelated with other drivers of deviations from trends in labor supply, our approach recovers the causal effect of additional use 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 as the mothers of children who wanted child care before the reform, but was restricted by the low supply. Andresen and Havnes (2018) provide a range of robustness- and specification checks
Table 7: First stage estimates, formal care use

<table>
<thead>
<tr>
<th>Months of child care use at ages 13 - 36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage rate at age 1</td>
</tr>
<tr>
<td>Coverage rate at age 2</td>
</tr>
<tr>
<td>Municipality fixed effects</td>
</tr>
<tr>
<td>Year fixed effects</td>
</tr>
<tr>
<td>Municipality-specific trends</td>
</tr>
<tr>
<td>N / mean dep. var.</td>
</tr>
<tr>
<td>$F$</td>
</tr>
</tbody>
</table>

*Note:* First stage estimates of eq. 4. Standard errors in parentheses, clustered at municipality.

to investigate the robustness of the exclusion restriction. First stage estimates from this specification is presented in Table 7, where we see that the availability of slots in care has a strong influence on the months of early care use. In particular, and not surprising, expansions of care in the calendar year the child turns 1 has strong effects on care use, with an effect of 5.2 months increased care use over the period from our preferred specification in column 2. The additional impact of expansions of care at age 2 is smaller, and nonsignificant for our preferred specification. The $F$-statistic is above 75, indicating a strong first stage.

Moving on, we present the second stage estimates in Figure 13a and 13b, where we have scaled the estimates by 12 to represent a full year of early child care use. For mothers, we reassuringly find no significant effects in the years leading up to birth, although there are some signs of a positive trend in labor supply that makes is interpret the post-birth impacts with some caution. Nonetheless, we see positive and significant impacts on earnings in the years 2-3, which are the years of treatment. For years 4-5, estimates drop and lose significance. For fathers, estimates are imprecise and never reach statistical significance, showing little evidence of any impact on paternal labor supply of child care use. These results are largely in line with Andresen and Havnes (2018).
The peak impacts on maternal labor supply are in the range of 35,000 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 13c, 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 approximately 40% at ages 2-3. We conclude that early child care shows more promise as a policy tool for reducing child penalties than paternity leave.

7 Conclusion

In the first half of this paper we show that same sex couples experience the child penalty very differently than 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 explains the vast majority of the relative child penalty experienced by heterosexual couples. Moreover, the large child penalty experienced by heterosexual mothers translates to 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 similar a number of children as heterosexual couples and, if anything, children of lesbian parents outperform the 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, policy might aim to address the behavior of fathers, and the most commonly proposed such policy is paternity leave. Second, policy might aim to provide a viable child care substitute for households to utilize in place of one of the parent’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
Figure 13: Impact of a year of child care use at ages 13-36 months on income
the relative child penalty. Next, we show that expansions of high quality child care, using a difference in difference approach, do reduce the child penalty experienced by heterosexual women by around 40%. 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.
References


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


A Additional results and robustness checks

This appendix contains various robustness checks and additional results for each section:

A: Theoretical model

- Additional simulations of model parameters in Figure A1

B: Child penalties

- Raw earnings around child birth and simple event studies without age- and year controls in Figure B1
- Number of children per parent type in figure B3, main results conditioning on no additional children until \( t + 5 \) in Figure B4
- Allowing age profiles to depend flexibly on gender and education level (9 dummies) in Figure B2
- Controlling for absolute advantage as measured by education levels in Figure B5 and education level and gender specific age profiles
- Long run estimates separately by 5-year birth cohort in Figure B6

C: Paternity leave

- Balancing checks of covariates around thresholds in Figure C1
- Density tests in Figure C1

D: Formal care application

- Without municipality specific linear trends in Figure D1
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.
Figure B1: Mean earnings by event time (left) and raw child penalties (right)

Note: Left panels 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 B2: Controlling for education- and gender-specific age profiles

Figure B3: Number of children over time by parent type

Figure B4: Only couples where no partner has additional kids until $t + 5$
Figure B5: Controlling for absolute advantage by interacting education level pre-birth with event dummies

Note: Also includes education level and gender-specific age profiles.

Table C1: Sharp RD balancing tests

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s age</td>
<td>RD estimate</td>
<td>0.081</td>
<td>-0.252</td>
<td>0.021</td>
<td>-0.110</td>
<td>0.169</td>
<td>-0.033</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(0.253)</td>
<td>(0.256)</td>
<td>(0.079)</td>
<td>(0.157)</td>
<td>(0.159)</td>
<td>(0.063)</td>
<td>(0.068)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.607</td>
<td>0.265</td>
<td>0.875</td>
<td>0.399</td>
<td>0.201</td>
<td>0.671</td>
<td>0.866</td>
</tr>
<tr>
<td>Mother’s age</td>
<td>RD estimate</td>
<td>0.183</td>
<td>-0.018</td>
<td>0.047</td>
<td>-0.208</td>
<td>0.288</td>
<td>-0.036</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(0.341)</td>
<td>(0.296)</td>
<td>(0.093)</td>
<td>(0.189)</td>
<td>(0.187)</td>
<td>(0.084)</td>
<td>(0.084)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.483</td>
<td>0.968</td>
<td>0.623</td>
<td>0.287</td>
<td>0.109</td>
<td>0.792</td>
<td>0.497</td>
</tr>
<tr>
<td>Maternity leave</td>
<td>RD estimate</td>
<td>1.353</td>
<td>-1.143</td>
<td>-0.104</td>
<td>-0.122</td>
<td>0.118</td>
<td>-0.321</td>
<td>-0.202</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(1.320)</td>
<td>(1.312)</td>
<td>(0.289)</td>
<td>(0.497)</td>
<td>(0.424)</td>
<td>(0.210)</td>
<td>(0.213)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.211</td>
<td>0.286</td>
<td>0.961</td>
<td>0.701</td>
<td>0.520</td>
<td>0.132</td>
<td>0.532</td>
</tr>
<tr>
<td>Father’s years of ed.</td>
<td>RD estimate</td>
<td>0.178</td>
<td>-0.472**</td>
<td>0.030</td>
<td>-0.016</td>
<td>0.003</td>
<td>0.056</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(0.213)</td>
<td>(0.215)</td>
<td>(0.048)</td>
<td>(0.096)</td>
<td>(0.104)</td>
<td>(0.061)</td>
<td>(0.042)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.489</td>
<td>0.013</td>
<td>0.586</td>
<td>0.799</td>
<td>0.983</td>
<td>0.243</td>
<td>0.945</td>
</tr>
<tr>
<td>Mother’s years of ed.</td>
<td>RD estimate</td>
<td>0.009</td>
<td>-0.519***</td>
<td>0.053</td>
<td>-0.031</td>
<td>0.080</td>
<td>0.044</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(0.218)</td>
<td>(0.222)</td>
<td>(0.054)</td>
<td>(0.104)</td>
<td>(0.094)</td>
<td>(0.059)</td>
<td>(0.041)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.942</td>
<td>0.008</td>
<td>0.301</td>
<td>0.723</td>
<td>0.297</td>
<td>0.391</td>
<td>0.834</td>
</tr>
<tr>
<td>Father’s ed. missing</td>
<td>RD estimate</td>
<td>-0.012</td>
<td>0.014</td>
<td>-0.003</td>
<td>0.002</td>
<td>-0.005</td>
<td>-0.004</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(0.014)</td>
<td>(0.012)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.338</td>
<td>0.252</td>
<td>0.191</td>
<td>0.784</td>
<td>0.386</td>
<td>0.680</td>
<td>0.211</td>
</tr>
<tr>
<td>Mother’s ed. missing</td>
<td>RD estimate</td>
<td>-0.010</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.007</td>
<td>-0.010</td>
<td>-0.004</td>
<td>-0.006**</td>
</tr>
<tr>
<td></td>
<td>robust s.e.</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>robust p</td>
<td>0.367</td>
<td>0.605</td>
<td>0.550</td>
<td>0.118</td>
<td>0.126</td>
<td>0.453</td>
<td>0.011</td>
</tr>
</tbody>
</table>

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 centered robust RD estimates across all five cutoffs. Robust bias-corrected inference. * p < 0.1 ** p < 0.05 *** p < 0.01
Figure B6: The child penalty over time, heterosexual couples

*Note:* Child penalties estimated from equation 2 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.
Figure C1: 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.
Figure D1: Formal care estimates, no municipality specific linear trends