What Causes the Child

Penalty and How Can It Be Reduced? Evidence from Same-Sex Couples and Policy Reforms*

Martin Eckhoff Andresen[†]

Emily Nix[‡]

Abstract

New parenthood causes large decreases in labor market incomes for mothers but not fathers, a stylized fact known as the "child penalty." We use a simple household model combined with a comparison of child penalties in heterosexual non-adopting, heterosexual adopting, and same-sex couples to better understand what causes the child penalty. Our results largely rule out giving birth and comparative advantage within the household as mechanisms, leaving preferences and gender norms as the main explanations, although we cannot disentangle these last two mechanisms. Building on these results we also provide causal estimates of two policies aimed at reducing the child penalty. We find small and insignificant impacts of paternity leave use on the child penalty, but find a 25% reduction in the child penalty from a large Norwegian reform that expanded access to child care.

JEL-codes: I21, J13, J22, J71

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

^{*}We thank seminar participants at ASU, Claremont McKenna, CSU Fullerton, Duke, EALE, Erasmus, LSU, Purdue, RAND, SOLE, SOFI Stockholm University, Stanford, Statistics Norway, UC Riverside, University of Oslo, University of Rochester, VATT Helsinki, and Warwick University. We also thank Heather Antecol, Manuel Bagues, Sebastian Calonico, Matias D. Cattaneo, Nina Drange, James Fenske, Yana Gallen, Trude Gunnes, Andrea Ichino, Edwin Leuven, Petra Persson, Adam Sheridan, Thor Olav Thoresen, Kenneth Aarskaug Wiik, Natalia Zinovyeva, and Antonio Dalla Zuanna for helpful comments and suggestions. All errors remain our own. Andresen gratefully acknowledges financial support from the Norwegian Research Council (grant no. 236947). This version: May 21, 2020. Latest version here.

[†]Statistics Norway, Oslo, Norway, mrt@ssb.no

[‡]Corresponding Author: Emily Nix, Marshall School of Business, University of Southern California, enix@usc.edu; USC FBE Dept. HOH Hall - 231, MC-1422, 701 Exposition Boulevard, Ste. 231 Los Angeles, CA 90089-1422 USA

1 Introduction

A growing number of papers demonstrate that parenthood causes a substantial drop in labor market income for mothers, but little or no labor market income drop for fathers.¹,² This stylized fact, commonly referred to as the "child penalty," is strikingly consistent across countries and for both well- and poorly educated mothers.³ As other determinants of the gender income gap have declined in importance, the proportion of the gap that can be explained by the child penalty has increased. For example, in Denmark the child penalty accounted for 80% of the gender income gap in 2013, up from 40% in 1980 (Kleven *et al.*, 2019b).⁴ Not only is the child penalty pivotal to the gender income gap, it also has important macroeconomic consequences, including implications for fertility, as recently demonstrated in Doepke and Kindermann (2019).⁵ In this paper we provide a better understanding of what causes the child penalty and how it might be reduced.

Four explanations are commonly given for the child penalty. First, biologically only women can give birth and in general only the woman who gives birth breastfeeds. Giving birth is a major health shock which can have long term consequences for pro-

¹For an overview see Kleven *et al.* (2019a). Additionally, earlier papers documenting the child penalty include Chung *et al.* (2017) and (Lundberg and Rose, 2000) in the United States, Angelov *et al.* (2016) in Sweden, and Kleven *et al.* (2019b) in Denmark.

 $^{^{2}}$ In fact, (Lundberg and Rose, 2002) find that men's labor supply and wage rates increase following birth of children, and the increase is larger in response to sons relative to daughters. See also (Choi *et al.*, 2008).

³We show this result in the Online Appendix, and it was also demonstrated in the earlier NBER version of Kleven *et al.* (2019b).

⁴Of 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.

⁵The authors model household bargaining over children and show that "the distribution of the burden of child care between mothers and fathers is a key determinant of fertility", which additionally has implications for future GDP and growth. While the authors take the unequal distribution of childcare as a given, in this paper we shed light on why the burden is unequal. These papers build on a rich literature in labor and development showing these effects at the micro level, see, e.g., Feyrer *et al.* (2008).

ductivity and earnings. Breastfeeding and time at home with the child while recovering from giving birth may cause the mother who gives birth to spend more time (and grow more attached) to the child, which could also have long term earnings consequences. Second, as women often make less than their husbands, men may have a comparative advantage in market work relative to household work compared with women, and households may efficiently specialize after birth. Third, women may have higher preferences for spending time with children than do men. Fourth, couples may default to traditional gender norms when deciding who should bear the costs of child-rearing. To formalize the implications of these four possible explanations for the child penalty, in the first part of the paper we develop and solve a simple household model.

The model and its solution yield two important conclusions that inform the empirical analysis. First, while it is not possible to disentangle mechanisms by looking at heterosexual couples alone, under reasonable assumptions a comparison with same-sex and adopting couples can be used to confirm or reject potential mechanisms. This motivates our first set of empirical results that compares child penalties across heterosexual, adopting, and same-sex couples. Note that we combine the preferences and gender norms mechanisms in the model, since these two mechanisms are fundamentally difficult, if not impossible, to disentangle.⁶ Second, based on the mechanisms identified by the model (and the data), we estimate the causal impact of two policies commonly proposed to mitigate them: paternity leave and subsidized early child care.

We take these insights to the data in two parts. First, using the event study approach from Kleven *et al.* (2019b) and administrative data from Norway we estimate child penalties for heterosexual non-adopting (hereafter "heterosexual") couples, heterosexual adopting (hereafter "adopting") couples and same-sex female couples⁷. We find that

⁶We discuss this in more detail in Section 2.

⁷There are too few same-sex male couples with kids in Norway to get precise estimates, so we focus

while there is no impact of parenthood on fathers' earnings, women in heterosexual couples experience a drop in income of approximately 20 percent following the birth of the first child, and this drop persists for at least five years after birth. This large drop in female income translates to an overall household income drop of 6 to 8 percent for heterosexual households that persists over time. We find almost identical patterns for adopting couples, with similarly large drops in income for mothers and no drops in income for fathers following the adoption of the first child. For same-sex female couples we find a dramatically different pattern: both women experience an income penalty after birth, but the birth mother experiences a larger drop of 13 percent, while her partner (hereafter "the co-mother") experiences a drop in income of 5 percent. Despite her larger immediate drop in income, the birth mother catches up with the co-mother two years after birth. From then on, both mothers experience similarly sized decreases in their income which decrease over time; by four years after birth both same-sex mothers' incomes have fully recovered. While the initial household income penalty experienced by same-sex female couples is approximately the same size as that of heterosexual couples (although shared more evenly between partners), by five years after birth same-sex female couples no longer experience a household income penalty. Using the same event study framework, we find that the main difference between heterosexual and same-sex mothers is at the intensive margin of labor supply, with little differences in the extensive margin, sickness absence or occupational sorting.

These empirical patterns, combined with the predictions from the model, allow us to largely reject two of the above four explanations for the child penalty in heterosexual couples. Our results demonstrate that the first explanation, that usually the woman gives birth to a child in heterosexual couples, can only explain part of the child penalty,

only on same-sex female couples.

and only in the first two years after birth. We can also reject the second explanation, that the difference in the child penalty between heterosexual and same-sex couples is explained by comparative advantages in market versus home production by controlling for several measures of relative productivity. We conclude that the majority of the child penalty in heterosexual couples is due to the third and fourth explanations, gendered differences in preferences for child care and gendered norms, although it is not possible to fully disentangle these two explanations.⁸

Our results that gender norms and preferences are largely driving the child penalty suggest that two commonly proposed policies could be effective at reducing the child penalty: paternity leave and subsidized early childcare. Paternity leave could increase the utility fathers get from spending time with children relative to mothers and also change gender norms around child care. Subsidized early child care, by reducing the cost of market care for children in the household budget constraint, could impact the trade-off between the gain to mothers from spending time with children versus the gains from increased consumption. Whether these policies work in practice, though, is an empirical question, so in the second part of the paper we estimate the causal impact of each policy on the child penalty.

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 to 2014. We estimate a strong first stage: the reforms significantly increase paternity leave takeup. However, despite fathers taking additional leave, we find no significant impact on the child penalty

⁸While comparing the outcomes of children born to same-sex and heterosexual couples is not this paper's focus, we also present descriptive evidence that same-sex partners sharing the parenting load more equally and experiencing smaller income penalties does not lead to worse outcomes for their children. We find that children of same-sex 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, even when controlling for a large range of observable differences between heterosexual and same-sex couples.

and we show that this zero result is quite precise. Moreover, paternity leave does not impact fathers' takeup of available shared leave for subsequent children, a measure that could potentially capture changes in the norms around child care within the couple. Both results indicate that paternity leave has limited potential to reduce the child penalty.

For subsidized early child care, we use a large-scale Norwegian reform from 2002 that expanded child care availability for 1- and 2-year-olds. The reform increased subsidies to child care institutions, leading to a rapid expansion of previously rationed care slots. To identify the impact of increased access to high quality child care on the child penalty, we exploit the variation across municipalities and over time in construction of new slots and centers, instrumenting individual child care use with the rationed, municipality-level availability of slots in a variation of the setup in Andresen and Havnes (2019). Our results indicate positive effects on mothers' labor income at ages two and three that scales to reduce the child penalty by around 25 percent for each additional full year of early child care use, although the impacts are not persistent in the long run.

Our paper is most closely related to the literature on child penalties. We use the simple event study approach from Angelov *et al.* (2016), Chung *et al.* (2017), Kleven *et al.* (2019b), and Bergsvik *et al.* (2019) to identify child penalties. Lundborg *et al.* (2017) also show that the child penalty occurs among heterosexual couples who use IVF to get pregnant, using quasi random variation in fertility after IVF treatment.⁹ 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 classes, that does not experience large child penalties. As these papers show, the child penalty is an important phenomenon in most of the developed world

⁹A recent working paper (Bensnes *et al.*, 2019) finds short-lived child penalties when accounting for the impact of additional kids using multiple IVF treatments.

whose role in explaining remaining gender income gaps cannot be understated. We make three additional contributions to this literature. First, we use a household model combined with a comparison with estimated child penalties for female same-sex couples and adopting couples to better understand the mechanisms behind the child penalty within heterosexual couples.¹⁰ Related to our work in this regard, in a very recent working paper, Kleven *et al.* (2020) also use adopting couples and find similar results as in this paper, namely that adopting couples experience a large and sustained child penalty, just like heterosexual couples. Their result complements our conclusion in this paper that giving birth does not explain the child penalty. Second, we are among the first to estimate child penalties in same-sex couples, which are important on their own. Most closely related in this regard are Moberg (2016) and Rosenbaum (2019) who also estimate the response to childbirth for same-sex couples in Sweden and Denmark, respectively.¹¹¹²

Third, motivated by what we find in terms of mechanisms, we estimate the causal impacts of paternity leave and access to high quality early child care on the child penalty. There is a large literature on the impact of both of these policies on a range of outcomes. We contribute to the literature by isolating the impact of these policies on the child penalty. Our results on paternity leave are related to and consistent with Antecol *et al.*

¹⁰Our findings regarding mechanisms are consistent with Kleven *et al.* (2019a) who show that the magnitude of the child penalties experienced by women are correlated with elicited gender norms across countries.

¹¹Additional related papers in this literature are Black *et al.* (2007), Baumle (2009), Schneebaum (2013), Antecol and Steinberger (2013), Carpenter (2005; 2007; 2008; 2009; 2017) and Aksoy *et al.* (2018) who compare earnings between heterosexual and same-sex individuals. Regarding parenting, Goldberg *et al.* (2012) look at a sample of 55 female same-sex couples and find they report sharing household chores and child care more evenly than a comparison group of 65 heterosexual parents. Others have investigated parental leave use (Evertsson and Boye, 2018; Rudlende and Lima, 2018) and time use (Martell and Roncolato, 2016) for same-sex couples, as well as the impact of legal recognition on employment (Alden *et al.*, 2015; Sansone, 2019). For a more general introduction to the history of research on lesbians and gay men in economics, see Badgett and Hyman (1998).

¹²Although we do not consider workplace discrimination against mothers, there is a large literature that finds mixed results (e.g. see Gallen (2018) and Bagues *et al.* (2017)).

(2018) who find that moving toward more gender neutral benefits in response to children does not help women in academia, and may even hurt their careers relative to men. We show that the results from Antecol *et al.* (2018) are not unique to academia. These results also tie in to a larger literature that examines the impacts of paternity leave on parents' earnings and labor supply and finds mixed results.¹³ 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 are not substantially affected by exposure to paternity leave. This result is similar to the finding in Bana et al. (2018), that men take much less paid family leave than women in California. However, while we find no impact on the child penalty or future leave taking of fathers, this does not rule out other positive impacts of paternity leave. Patnaik (2019) finds a large change in the division of household labor from a Canadian paternity leave expansion and Persson and Rossin-Slater (2019) find that when fathers have more flexibility to stay home, there are positive impacts on the mother's health. Regarding child care, a large literature (summarized in Blau and Currie, 2006; Akgunduz and Plantenga, 2018; Morrissey, 2016) contains a range of estimates on the elasticity of female labor supply to child care availability. Of most relevance here are Havnes and Mogstad (2011) who find small effects from a child care reform for preschoolers, and Andresen and Havnes (2019) who find considerably larger effects from a child care reform for toddlers. In this paper we focus on the impact of these policies on one particular outcome of interest, the child penalty.

¹³For a good overview of this literature, see Rossin-Slater (2017). Most closely related to this paper, Rege and Solli (2013) find a decrease in fathers' earnings long term in Norway from a 1993 reform using a difference in difference approach, Druedahl *et al.* (2019) find that a Danish increase in the the daddy quota from 2 to 4 weeks increased mothers' share of household earnings; 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; Cools *et al.* (2015) estimate the effect of paternity leave extensions in Norway and, like us, find no effect on traditional labor supply allocations in the family although they do find improvements in children's test scores; and Andersen (2018) find that father's leave reduces the within household gender gap in Denmark.

The remainder of the paper is organized as follows. In Section 2 we present the model and its solution. In Section 3 we describe how we identify and estimate the child penalty, our main outcome of interest throughout the paper. In Section 4 we summarize the data and present summary statistics. In Sections 5 and 6 we present the main results. Section 7 concludes.

2 A model of household labor supply with children

In this section we write and solve a simple model of household labor supply in the presence of children. While we do not estimate the model, the solution is useful insofar as it allows us to more formally discuss how comparisons between couple types can help us disentangle mechanisms. The model incorporates four of the most commonly suggested mechanisms for the child penalty: costs of giving birth, specialization within households, larger female preferences for child care, and gender norms¹⁴ around child care.¹⁵ The model is deliberately stylized to bring out the implications of the four mechanisms, and what we may be able to learn by comparing heterosexual, adopting, and female same-sex couples. We also use the simple model to formalize how the policies we analyze in Section 6 might impact the child penalty. Our model is loosely adapted from Fernández *et al.* (2004) and Olivetti (2006).¹⁶

In each household in our model, there are two adults, either a man and a woman,

¹⁴Survey evidence shows large differences in the norms towards working women with young children compared with working women without children. As an example, 80 percent of the respondents in the ISSP in 2002 think that married women without children should work full time in the United States, while only around 15 percent think the same about women with children below school age. Similar differences appear for other countries, including Sweden and Denmark (see International Social Survey Program (ISSP) from 2002, see also Kleven *et al.* (2019a)).

¹⁵While we do not explicitly discuss discrimination in the labor market as a separate mechanism, this channel can be folded into the gender norms mechanism.

¹⁶Another example in this literature where the model is directly estimated is Adda et al. (2017).

or two women. ¹⁷ There is 1 period before birth and N periods after birth, and in each period each adult is endowed with 1 unit of time. In the first period, the two adults both inelastically supply their labor to the market. At the start of the second period a woman in the household gives birth or, for adopting parents, the child is adopted.¹⁸ Thereafter, the household consists of the two adults and the child, and the households must choose the amount of labor each adult allocates between home and labor market production, and the amount of childcare purchased on the market. The quasi linear utility function of each spouse *i* in each period *t* is given by:

$$U_{i,t}(c_t,\theta_t,h_{-i,t}) = c_t + \beta \ln \theta_t + \eta_{i,t} \ln(1-h_{i,t})$$

$$\tag{1}$$

where c_t is consumption, θ_t is child quality, and β represents the value of child quality, which is identical for all individuals and constant across time (in period 1, $ln\theta_t=0$). $h_{i,t}$ and $h_{-i,t}$ represent the fraction of time spent working away from home of individual *i* and his or her spouse, respectively, and $(1-h_{i,t})$ represents the time individual *i* spends with the child. We capture both gender norms and gendered preferences for childcare through the term $\eta_{i,t}\ln(1-h_{i,t})$, which indicates that women may get greater utility from time with children than men, under the assumption that if individual *i* is female, $\eta_{i,t} = \overline{\eta}_t$, which in every period is larger than the equivalent for men, $\underline{\eta}_t$, so that $\overline{\eta}_t > \eta_t$ $\forall t$.¹⁹ Note that we combine these two mechanisms into a single term in

¹⁷While couples with two men are also of interest, we have too few with children in our data to get precise empirical estimates.

¹⁸We do not model the fertility decision or allow parents to make labor market decisions in anticipation of children. While these are important issues (see, e.g., Bursztyn *et al.* (2017) and Doepke and Kindermann (2019)), they are beyond the scope of this paper. We do allow for an income gap before children, which could capture some of these points.

¹⁹One might suspect that the preferences for children (and time spent with them) is stronger among same-sex couples, because getting pregnant for most of them will involve greater costs, although artificial insemination is a much easier procedure than IVF. If that were true, we would expect the income costs from preferences for spending time with children to be even larger for female same-sex couples.

the model because we view the two as impossible to disentangle. For one example, if gender norms shape preferences of young children (girls play with dolls, boys play with trucks), then preferences will also capture gender norms.

Child quality is produced by the following production function that takes as inputs each parent's time and the child care purchased on the market, denoted $h_{m,t}$.

$$\theta_t = k_i \psi(1 - h_{i,t}) + k_{-i} \psi(1 - h_{-i,t}) + k_m \psi(h_{m,t}) \tag{2}$$

where we assume that $\psi' > 0$, $\psi'' \le 0$, and $\psi(0) = 0$.

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

$$c_t = w_i (1 - \delta_t \bar{S}_i) h_{i,t} + w_{-i} (1 - \delta_t \bar{S}_{-i}) h_{-i,t} - p h_{m,t}$$
(3)

where \bar{S}_i is an indicator equal to 1 if individual *i* is a woman who gave birth. δ_t is the time varying productivity shock of giving birth. Note that δ_t is not isolated to the year of birth. While breastfeeding and the actual act of birth are short run events, there are a number of reasons why giving birth might affect productivity long term. First, there is substantial evidence that health shocks have long term consequences for earnings, and giving birth is certainly a major health event for women. Second, breastfeeding and spending time at home with the child while recovering from giving birth might promote longer term attachment to the child, which might have long term earnings costs. p is the cost of purchasing childcare on the market. The combination of wages and productivity at home capture comparative advantage differences. For example, if

 $\frac{w_i}{k_i} > \frac{w_{-i}}{k_{-i}}$, then partner *i* has a comparative advantage in market production and partner -i has a comparative advantage in producing child quality.

In the context of the model, if we define each individual's income as $y_{i,t} = w_i h_{i,t}$, then the percentage change in income in each period t for individual i relative to his or her income the year before birth can be written as $\Delta Y_{i,t} = \frac{y_{i,t} - y_{i,1}}{y_{i,1}}$. The child penalty is the difference in this percentage change in income between the two spouses.

The household maximizes utility by choosing each spouse's division of labor in each period and the amount of childcare purchased on the market, where household utility is given by

$$\sum_{i} \lambda_i U_{i,t}(c_t, \theta_t, h_{-i,t}, h_m)$$

and λ_i is the weight of each spouse in household decisions. This assumes Pareto efficiency in household decisions and is consistent with a number of household bargaining problems.²⁰

There are no dynamics to the problem. This means we can solve the problem separately for each period t, maximizing h_i and h_{-i} in each period. For each period, the couples solve the following equation, taking the home production process in equation 2 as given, where for simplicity we suppress the time subscripts:

$$\max_{h_i,h_{-i},h_m} (\lambda_i + \lambda_{-i}) \left(w_i h_i + w_{-i} h_{-i} - p h_m - \delta w_i h_i \bar{S}_i - \delta w_{-i} h_{-i} \bar{S}_{-i} + \beta \ln \theta \right)$$

$$+ \lambda_i \eta_i \ln(1 - h_i) + \lambda_{-i} \eta_{-i} \ln(1 - h_{-i})$$

$$(4)$$

²⁰This is a very simple model by design. It assumes Pareto efficiency, but this has some important drawbacks. See Del Boca and Flinn (2012) for a discussion of alternative approaches. Notice that we assume that the bargaining weights do not vary by couple type. An alternative approach to capture gender norms could be to assume that in same-sex couples $\lambda_a = \lambda_b$ and in heterosexual couples $\lambda_a > \lambda_b$, where λ_a represents the Pareto weight of the man.

The solution can be characterized by the following first order conditions, where for simplicity we normalize $\lambda_i + \lambda_{-i} = 1$:

$$\frac{w_{father}}{k_{father}} = \frac{\beta \psi'(1 - h_{father})}{\theta} + \frac{\lambda_{father} \underline{\eta}}{k_{father}(1 - h_{father})}$$
(5)

$$\frac{(1-\delta)w_{mother}}{k_{mother}} = \frac{\beta\psi'(1-h_{mother})}{\theta} + \frac{\lambda_{mother}\overline{\eta}}{k_{mother}(1-h_{mother})}$$
(6)

$$\frac{(1-\delta)w_{mother,s}}{k_{mother,s}} = \frac{\beta\psi'(1-h_{mother,s})}{\theta} + \frac{\lambda_{mother,s}\overline{\eta}}{k_{mother,s}(1-h_{mother,s})}$$
(7)

$$\frac{w_{co-mother}}{k_{co-mother}} = \frac{\beta \psi'(1 - h_{co-mother})}{\theta} + \frac{\lambda_{co-mother}\overline{\eta}}{k_{co-mother}(1 - h_{mother,s})}$$
(8)

$$p = k_m \frac{\beta}{\theta} \psi'(h_{market}) \tag{9}$$

These wage equations specify how the labor supply of heterosexual fathers, heterosexual mothers, and same-sex birth mothers and co-mothers will be impacted by children under each mechanism and yield the following insights:

- Costs of giving birth: For each period where δ > 0, there will be a child penalty for both heterosexual and female same-sex couples, but not for adopting couples. If the costs of giving birth are the only reason for the child penalty then the child penalty will be identical for heterosexual and same-sex female couples and negligible for adopting couples. Note that because we allow δ to vary over time, while the costs of giving birth might be large initially, the model allows for these costs to decrease over time.
- 2. Specialization based on comparative advantage: Fathers have a comparative advantage in market versus home production if $\frac{w_{father}}{k_{father}} > \frac{w_{mother}}{k_{mother}}$. If this is true,

then heterosexual couples will optimally specialize with mothers reducing their labor supply to the market and fathers increasing their labor supply to the market in response to children. If we compare heterosexual and female same-sex couples with the same comparative advantage differentials and comparative advantage explains the child penalty for heterosexual couples, then the child penalty will be identical for heterosexual and female same-sex couples.

3. Gender norms/Gendered differences in preferences for child care: As the gendered differences in preferences and/or gender norms around child care grow larger (η
 – η increases), mothers in heterosexual and adopting couples will decrease their labor supply while fathers will increase their labor supply. However, all else equal, if η increases then the labor supply of heterosexual mothers will decrease by a larger amount than the labor supply of same-sex birth mothers and co-mothers.

Notice that every mechanism leads to a child penalty for heterosexual couples, which is why it is impossible to disentangle mechanisms when looking only at heterosexual couples. Adding same-sex and adopting couples allows us to possibly distinguish between mechanisms using these predictions from the model.

The model also formalizes how paternity leave and subsidized early childcare might reduce the child penalty. Specifically, equation 9 shows that the introduction of subsidized early childcare, by decreasing the price of market care, p, could cause the family to increase the amount of formal child care they purchase, h_m . If this occurs, it will also cause an increase in the labor supplied to the market by the mother and/or father. Paternity leave (in the form of "use it or lose it" quotas for fathers) might impact the child penalty for heterosexual couples by changing gendered preferences and/or gender norms through changes in η .

3 Identifying Child Penalties

To identify the child penalty, our main object of interest throughout the paper, we adopt an event study framework as in Kleven *et al.* (2019b). 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 B1 in the Appendix and are almost identical, but Kleven *et al.* (2019b) show that including age and time dummies performs better. Event study frameworks such as this have been used to investigate, among other things, the economic impacts of inheritances (Druedahl and Martinello, 2016), hospital admissions (Dobkin *et al.*, 2018) and family health shocks (Fadlon and Nielsen, 2017)²¹

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

²¹See, e.g., Jacobson *et al.* (1993) and McCrary (2007) for earlier examples.

penalties

$$y_{it} = \underbrace{\sum_{j \neq -1}^{\text{Parent-type event time dummies}}}_{\text{Gender-specific age profiles}} + \underbrace{\sum_{j \neq -1}^{\text{Gender-specific age profiles}}}_{m} \beta_{lm} \mathbb{1}[age_{it} = l, X_i = m]$$
(10)
$$+ \underbrace{\sum_{n=0}^{n} \gamma_{no} \mathbb{1}[T_{it} = n, X_i = o]}_{\text{Gender-specific year shocks}} + \underbrace{\sum_{p} \eta_p \mathbb{1}[K_i = p]}_{\text{Parent-type fixed effects}} + \epsilon_{it}$$

Where X_i is the gender (male, female) of parent *i*, age_{it} is the age of parent *i* at event time *t*, T_{it} is the calendar year for individual *i* at event time *t*, and K_i is the parent type: mother or father in a heterosexual couple, and mother or co-mother in a same-sex couple. $\mathbb{1}[A]$ is the indicator function for event *A*. Standard errors are clustered by couple and robust to heteroskedasticity. The event time dummy the year before birth is omitted, which implies that all estimates of event dummies are relative to the year before birth for that specific parent type.²² As all parents in our sample eventually have children, 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. Kleven *et al.* (2019b) show that the event study approach we use here performs well in identifying both short- and long-run child penalties compared with alternative approaches such as using instruments for the timing of birth.

A recent literature (Abraham and Sun, 2020; Borusyak and Jaravel, 2018; Goodman-Bacon, 2018; Novgorodsky and Setzler, 2019; Schmidheiny and Siegloch, 2019) formally

²²Note 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 same-sex female couples. While it is possible to estimate equation (10) separately for heterosexual mothers and fathers and same-sex birth mothers and co-mothers, estimating the equation jointly allows us to exploit the large number of heterosexual couples to help identify these control variables for same-sex couples as well as heterosexual couples. In the Online Appendix, we present a number of robustness checks that suggest our results are not driven by this restriction. Also note that the raw event studies presented in the Online Appendix give the same qualitative results as in our main outcomes.

discuss identifying assumptions in event study frameworks and their relation to differences in differences, generally assuming parallel trends and no anticipation. Parallel trends in our setting requires that couples who get their first child in e.g. 2005 and 2010 would have had the same change in mean labor supply between two particular years had they had no children. No anticipation requires that for all years before birth, observed outcomes must equal outcomes in the counterfactual case where they had never had children. The plausibility of the latter is often investigated using estimates of trends in outcomes before birth. Note that significant pre trends is the rule, not the exception, in the literature on child penalties estimated using event study frameworks. Kleven et al. (2019a) document upward sloping pre-trends at about the magnitude as we find for both women and men in Denmark, Sweden, Germany and Austria, Sieppi and Pehkonen (2019) the same for Finland, while pre-trends for the UK and the US are downward sloping for women in Kleven et al. (2019a). Notice, however, that this does not necessarily constitute a violation of no anticipation, as long as these trends in outcomes would have been observed also in the counterfactual case where the couple would never have had children in our sample window. This could happen if couples tend to have children around the time when they have steeper wage growth than the average for their age due to the natural sequencing of finishing education and entering the (full time) labor market for the first time. We further discuss the interpretation of the common trends in Section 5 and provide evidence from a sample of adopting parents, where no pre-trends are present and post-birth estimates are equivalent, that support this interpretation and our identification strategy.

Our objects of interest are α_{jk} , the change in the outcome for a parent of type k at child age j relative to the earnings the year before birth.²³ Ideally, we would use

²³Notice that these child penalties include the impact of subsequent children that may appear in later years.

a log-linear specification of equation (10) to interpret the coefficients as percentage changes in earnings, but the presence of zeros in the outcome complicates matters. Thus, to convert these absolute estimates to percentage child penalties, we follow Kleven *et al.* (2019b) and construct the following measure of the child penalty.

$$C_{jk} = \frac{\hat{\alpha}_{jk}}{\mathbb{E}(\hat{y} \mid t = j, K_i = k)} \tag{11}$$

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

The simple event study identifies the causal effect of having children on labor market outcomes of mothers and fathers in heterosexual and adopting couples and birth mothers and co-mothers in same-sex couples. These results are interesting on their own, but we will also discuss how these results, combined with the predictions from our model, shed light on the mechanisms determining the child penalty in heterosexual couples when we present the first set of results in Section 5.

However, one assumption we require is worth discussing in more detail here. In particular, our model predicts that we can rule out specialization based on comparative advantage as explaining the child penalty in heterosexual couples if we compare heterosexual and same-sex female couples with similar comparative advantage differentials, $\frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}}$, and find that the child penalties are not identical. This suggests that we should add interactions of $\frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}}$ and the event time dummies to the specification in equation (10) if we wish to compare couples with similar relative productivities. Thus, to control for specialization, we flexibly control for the differences in own and partner's education

the year before birth interacted with event dummies, by adding $\sum_j\!\theta_j\,\mathbbm{1}[t\!=\!j](e_i\!-\!e_{-i})$ to equation (10), where e_i is years of education, measured the year before birth. θ_j captures the part of the child penalty that is explained by the relative differences in education. To the extent that comparative advantage is captured by relative education of the two spouses, these flexible event dummy controls will pick it up and we can attribute the remaining child penalties from α_{jk} to the other possible mechanisms highlighted by the model. As an alternative control for comparative advantage, we control for differences in earnings at the start of our panel, an alternative measure likely to capture relative differences in productivity that may cause different couples to respond differently to children. These results are very similar to the specification using educaditon differences, and are presented in the Online Appendix. When presenting these results, we scale by the predicted earnings from the baseline estimates in equation (10), and bootstrap confidence intervals for the scaled results clustering on couple. We interpret any remaining child penalties in earnings as coming from sources other than specialization. As an alternative specification, we restrict to same-sex and heterosexual couples who are similar on a large set of observables in a nearest-neighbor matching exercise.

4 Data and institutional setting

Our data comes from Norwegian administrative registers covering the entire resident population. We use unique identifiers to link individuals across registers, over time, and to family members. Our main outcome of interest throughout the paper is the child penalty in annual labor market earnings, obtained from the tax records. Importantly, these are wage incomes that include taxable benefits such as sickness and parental leave benefits.²⁴ In subsection 5.2, using data from the FD-Trygd, the register of the Norwegian Public Insurance system, we also discuss additional labor market outcomes related to employment spells.²⁵ From these spells, we construct the following measures of *monthly* labor supply: dummies for employment spell exceeding 4, 20, and 30 contracted hours per week; whether or not the primary employment is in the public sector (2003 - 2014 only) and a proxy measure of the family friendliness of the firm.²⁶ In addition, we measure the total working hours of all employment spells for the years 2003 - 2014.

For the analysis comparing heterosexual couples and same-sex female couples in Section 5, we benefit from the fact that through the 1993 Partnership Act, Norway became the second country in the world to legally recognize same-sex partnerships, so we have a longer panel of same-sex couples compared with most countries. However, there were restrictions regarding children²⁷ until 2002, when same-sex couples became legally eligible to adopt the children of their partners. Thus, we take extra caution to identify children of same-sex couples, adopting couples, and heterosexual couples, and restrict the analysis to children born in or after 2001. We describe these steps in more detail in Online Appendix Section A.

Consistent with previous papers, we keep only first-born children of both parents.

²⁴For sickness absence and parental leave spells we pull data from FD Trygd. 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 were taken for a particular child. Details on these measures are provided in Online Appendix A.

²⁵The database covers most important employment spells from 1992 to 2003 and all employment spells (excluding self-employment) from 2003 to 2014. To create comparable measures across most of the sample period, we exclude spells of self-employment from the pre-2003 data and include only the employment spell with the most contracted hours for the post-2003 data. In more than 95 percent of cases, the spell considered most important in the pre-2003 data is the one with the longest contracted hours.

²⁶Family friendliness is the leave-out-mean of mothers with children below 15 years who work in the firm.

²⁷Same-sex couples were not eligible for domestic adoptions or publicly subsidized assisted fertility treatment, and the registered partner of a woman giving birth was not automatically registered as the second parent (as with the *pater est* principle established for married heterosexual couples).

In the case of multiple births, we keep the couple in the sample only once. We keep only couples where both spouses reside in Norway the year before birth. Lastly, we keep only couples where the first child is born when both parents are aged between 22 and 60 years, giving us some time before and after birth to observe earnings. This leaves us with a sample of 250,296 heterosexual couples and 634 same-sex female couples.²⁸

To investigate labor market responses to the child's birth, we match these mothers and fathers to their labor market earnings in all years from t-4 to t+5, centered around the birth of the first child. Note that for children born after 2012, we will not see a full 5 years of income after birth because our data ends in 2017. We report summary statistics for this part of the paper in Online Appendix Table A5 columns 2 and 3. Same-sex female couples are slightly older than heterosexual couples at first birth, and are also slightly more educated.²⁹ We find that both partners in female same-sex couples have higher pre-birth labor earnings relative to heterosexual mothers, and the gap between mothers is smaller than the gap within heterosexual couples. While some of this is likely driven by older age at first birth, it also suggests the importance of controlling not only for income, but also for pre-child income gaps in order to understand the role of comparative advantage in determining the child penalty (see Section 2).

In Section 6 we estimate the impact of paternity leave and child care availability on the child penalty. Following a birth, Norwegian parents have been entitled to a generous paid parental leave since 1977. Total parental leave is currently 49 weeks at 100 percent replacement or 59 weeks at 80 percent replacement rate, but the length of leave has

²⁸The number of same-sex male couples (32) with children is unfortunately too small to yield precise estimates, so we focus only on heterosexual and same-sex female couples. The lack of same-sex male couples with children is consistent with the difficulty same-sex male couples face when trying to have children. See Section A.

²⁹Reflecting the rules on establishing legal co-parent status (see online Appendix Section A), the age at adoption is slightly delayed for same-sex female couples compared with heterosexual couples, as it takes some time for the co-mother to be legally registered.

steadily increased since the mid 1980s, reforms that we exploit and describe in more detail in Section 6.1. Benefits are capped at around 600,000 Norwegian kroner (NOK), roughly 70,000 USD, with many employers topping up. The leave is split in three with a quota for the mother, one for the father (since 1993), and the rest to be distributed between the parents.³⁰

We additionally exploit data on child care use and availability. Following parental leave, Norway has a well-developed, well-regulated, and highly subsidized child care sector (see Online Appendix Figure D5a). Because of the heavy subsidies for formal care, the market for paid child care outside this system is very small, but subsidies are available for both private and public suppliers of formal care. For the measure of slots, we use administrative data from the child care centers on the number of slots for children of different ages on December 15 each year. At the individual level, however, we can also measure the exact use of child care for those aged 13 to 35 or 36 months, depending on the cohort. For these ages, a cash-for-care benefit was given to parents whose children did not attend formal care in a given month. If we assume that all parents who did not use formal child care applied for the benefit, which is relatively generous,³¹ we know exactly which children attended how much care for each month. From these data, we construct precise measures of full-time equivalent years of formal individual child care use from ages 13 to 36 months.

³⁰In 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. For fathers and co-mothers, both parents must qualify. 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.

³¹Throughout 2001-2009, which is the period we exploit, the benefit was around 3,500 NOK or 420 USD per month.

5 Heterosexual, adopting, and same-sex child penalties

In Figure 1 we present the main results.³² The graphs report estimates of C_{jk} (see equation (11)) generated by the simple event study in equation (10). Before turning to the post-birth estimates of child penalties, observe that the trends in income is positive before birth for all parents, indicating that there is income growth in the years preceding birth that exceeds the income growth given by the overall age profiles. This is in line with more or less every estimate of child penalties in the literature, where exactly the sort of patterns we observe before birth in these graphs are the rule, not the exception.³³,³⁴ In addition to being consistent with the prior literature, in Figure (2) we show that the pre-trends for a sample of adopting parents, who cannot time children to coincide with, for example, the events described above, are completely flat, while at the same time the child penalties following birth are very similar to our main results, providing support that the no anticipation assumption is met in our context (see Section 3 for more discussion). Finally, notice that the pre-trends are very similar across couple types could be unaffected even if (parts) of the pre-trends are caused by anticipation effects.

The results for heterosexual couples are shown on the left and for same-sex female couples on the right. As has been shown in many other papers, we find that upon

³²In Online 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 (10), and the patterns are the same.

³³See Section 3 for a more detailed discussion of the relevant papers and their findings, but note that both Kleven *et al.* (2019b) and Kleven *et al.* (2019a), as well as many others find the same patterns pre-birth as we document here.

³⁴In event study papers, significant pre-trends are often interpreted as evidence against the no anticipation assumption as discussed in Section 3, but as long as this income growth would have happened even if the couple had not gotten a child at all (or at another time), it does not necessarily violate this assumption. This could happen, for example, if the natural sequence of events is to finish education and enter the (full time) labor market for the first time before considering children. As long as these events are not caused by the anticipation of children and would have happened even in their absence, the assumption is not violated.

the birth of their first child, mothers in heterosexual couples experience large income penalties, in the range of 20 percent of their counterfactual earnings, whereas fathers experience no income penalty.

The graph for female same-sex couples is strikingly different. We find that both mothers experience a drop in income the year after the child is born, but that initially the birth mother has a larger drop in income. These drops in income, however, are much smaller than that of heterosexual mothers, at around 13 percent and 5 percent of counterfactual earnings for mothers and co-mothers, respectively. Moreover, two years after birth the birth mother catches up to the co-mother and her penalty is no longer statistically significantly different from her partner's. By five years after birth, the income penalty for both women has largely disappeared.



(a) Heterosexual (left) and same-sex female (right) couples



(b) Controlling for comparative advantage using years of education differences in t-1



(c) Heterosexual couples, nearest neighbor

Figure 1: Estimated child penalties across couples types

Note: Figures in the top panel show the estimated child penalties from equation (10), scaled as described in equation (11). Sample construction and data as defined in section 4. Figures in the middle panel show the estimated child penalties where we control for initial differences in productivity using differences in labor market earnings at the beginning of our sample period. The graphs show the remaining child penalty after removing the event dummies interacted with the pre-birth within couple income gaps. Figures in the bottom panel show child penalties estimated from the baseline model in a the heterosexual couple sample matched to the same-sex female couples on pre-birth are model in a nearest neighbor matching exercise. Bootstrapped 95 percent confidence intervals in gray using 200 replications and clustering by couple.

In Panel B of Figure 1 we re-estimate the child penalties for each couple, and include additional controls based on the model to remove the impact of comparative advantage and isolate the portion of the child penalty due to other mechanisms (see Section 3). The results remain virtually identical, suggesting that comparative advantage cannot explain the child penalties within heterosexual couples, or the different patterns in same-sex female couples compared with heterosexual couples.^{35,36} One possible concern with these results is that there may be other differences between couple types that explain the differences in child penalties, which are not captured by the model driven controls for comparative advantage in Panel B. To address this concern we present child penalties for heterosexual and same-sex female parents using a nearest neighbor matching exercise in Panel C. We match on a variety of pre-birth characteristics such as municipality of residency, both parents' age and education and their interaction, and number of kids to account for twins and the rare triplets. We then re-run our baseline model in the matched heterosexual and same-sex female samples. Although precision is lower in the sample of same-sex female couples, the results are similar to the baseline estimates for both exercises.

Next, in Figure 2 we estimate child penalties for a sample of approximately 1,800 adopting parents over 2001-2014, who adopt a child who is 3 years or younger.³⁷ We

³⁵Instead of using the pre-birth education gaps, we have also estimated the child penalties using the gap in income at the start of our panel, another measure of labor market productivity (see Online Appendix Figure B3). The results are almost identical.

 $^{^{36}}$ In the Online Appendix we also graph the child penalties for heterosexual couples where the mother makes more than the father pre-birth. In these cases, comparative advantage should cause the father to specialize in child care and have a larger penalty. Instead, we find that the mother continues to experience a large and similarly sized drop in income (see Online Appendix Figure B4). For men, there appears to be a premium, perhaps driven by the fact that some of the men had temporarily low earnings in year t-1. When further restricting the sample to couples where both spouses earned more than 300,000 NOK in the year before birth, this large premium for fathers disappears, whereas the child penalty for mothers remains.

³⁷We have constructed the sample of the adopting parents in the same way as for the baseline sample but concentrated on couples where at least one of the spouses take up at least some special parental leave available for adopting parents. We restrict the sample to couples where the child arrives in Norway at the latest in the calendar year the child turns 3, but notice that event time is still relative to the year of birth of the child, not the year the child arrived in Norway. This is the reason why the onset of the

find that the child penalty is similar to the baseline estimates for heterosexual couples at approximately 17% from age 2 onward (and the later onset is likely due to to on average later arrival of the child), with no sign of catch-up and no child penalty for the father. These results show that even among heterosexual couples who desired a child but where the woman did not give birth, there is still a large child penalty. Additionally, note that the pre-trends are very flat in this sample of parents whose children are adopted. We take this, as well as the similarity in the post birth patterns for heterosexual biological and adoptive parents, as suggestive evidence that anticipation effects preceding birth is not driving our estimates.

child penalty is slightly delayed compared to the baseline estimates.



Figure 2: Child penalties for adopting parents

Note: This figure shows child penalties for a sample of around 1,800 adopting parents from 2001-2014 from equation 10 on the sample of adopting parents only. Notice that event times are normalized to the birth year of the child, not the year in which the adoption actually happened, which will often be a year or two later.

The child penalty experienced by women in heterosexual couples is so large, it would seem to imply an overall household income penalty. In Figure 3 we show this is the case by using the total income of the two spouses as the outcome. What is particularly interesting is that both heterosexual and same-sex female couples experience statistically indistinguishable initial total income declines on the birth of the first child. However, this drop in income persists for heterosexual couples, whereas it decreases over time for female same-sex couples.



Figure 3: Child penalty, total household income

Notes: Figure shows estimates of the child penalty on the sum of the two partners' labor market incomes. We control for gender specific calendar year and age dummies for both spouses as in equation (10), and scale the estimates as in equation (11). Standard errors are bootstrapped, clustering by couple.

Based on all of these results a number of conclusions can be drawn. First, and perhaps most obvious, the striking difference in response to childbirth we find among heterosexual couples is not a necessary outcome of child-rearing. Second, giving birth in and of itself does not cause large and persistent labor market penalties. This result is supported not only through the comparison between same-sex and heterosexual couples, but also by looking at the child penalty for adopting parents. Thus, the fact that only the woman can give birth in a heterosexual couple cannot, on its own, explain the child penalty, since first, adoptive mothers also experience a large and sustained income penalty and second, the birth mother in female same-sex couples experiences a smaller income penalty after birth that is only significantly different than the co-mother's penalty in the year of and the year after birth, and moreover, by five years after birth, no longer experiences an income penalty at all. With additional assumptions based on the model we can go a step further. When we use the approach suggested by the model to correct for comparative advantage differentials by interacting either the education or income differences between couples with event dummies and removing that effect from the estimates, as we do in the middle panel of Figure 1 (or in the Online Appendix for the income differences), the strikingly different patterns remain.³⁸³⁹ The different patterns also remain when we estimate the model using a nearest neighbor exercise in the bottom panel of Figure 1. This suggests that the comparative advantage mechanism cannot explain the child penalty in heterosexual couples. Thus, we are able to reject two out of the four most commonly suggested explanations for it. This leaves us with gender norms and gendered differences in preferences for child care as the main mechanisms behind the child penalty.

5.1 Robustness checks

In the Online Appendix, we report results from a number of robustness checks. First, one might be concerned that the differences between heterosexual and female samesex couples are explained by differential fertility. In Online Appendix Figure B7, we show that fertility patterns for same-sex female couples are almost identical to those of heterosexual couples. Second, female same-sex couples may switch who gives birth over time, and this could explain the catch up experienced by the woman who initially gives birth. In online Appendix Figure B8 we estimate child penalties for heterosexual and same-sex couples who do not have a second child in the five years following the birth of the first child. While this is an endogenous sample selection that reduces the sample size and precision for female same-sex female couples and so we should be careful in

 $^{^{38}}$ The differences before birth are likely caused by autocorrelation of incomes over time, as income differences are measured in year $t\!-\!4$.

³⁹We get the same results when we control for comparative advantage using education differences within the couple pre-birth (see Online Appendix Figure B3).

interpreting these estimates, the patterns remain the same. Third, same-sex female couples may choose the healthier woman to give birth, which reduces the penalty from giving birth for these couples. While the results for adopting parents are also not consistent with giving birth as the main mechanism for the child penalty (see Figure 2), a few additional facts suggest this is not likely to be a concern for the same-sex couple analysis. In the summary statistics (Online Appendix Table A5), we show that both same-sex mothers take more time for sick leave before birth compared to heterosexual mothers (conditional on employment), which suggests that same-sex mothers are not healthier than heterosexual mothers before birth. In addition, we find that the average number sick days taken for the mother who gives birth is higher than for the mother who does not give birth in same-sex female couples, which is not consistent with a theory in which same-sex couples routinely choose the healthier spouse to give birth to the child.

Last, we have shown that same-sex couples share the burden of child rearing more evenly, and experience less severe household income penalties compared to heterosexual couples. It is natural to ask if this reduction in the child penalty comes at the cost of worse outcomes for children. In Online Appendix Section B.3 we show this is not the case. We find the opposite in Table B1: children of same sex couples outperform children of heterosexual couples, as measured by age 10 test scores. This result overcomes previous data shortcomings to contribute to a charged debate regarding the impact of same-sex parents on child outcomes.⁴⁰

⁴⁰For example, see oral arguments for the landmark 2015 Supreme Court case *Obergefell v. Hodges*, which legalized same-sex marriage in the United States.

5.2 Decomposing the child penalties

To further understand the anatomy of the child penalty and what female same-sex couples do differently than heterosexual couples, we estimate the child penalty separately for the following determinants of income: extensive margin labor market participation, intensive margin participation (weekly contracted hours of work), family friendliness or public sector status of the firm, and days of sick leave. We present these results in Appendix Section B.4. We find that same-sex mothers are equally likely to switch to flexible careers compared with heterosexual mothers, but do not have long term income penalties from having had children. This suggests that occupational flexibility alone cannot explain the large and sustained income penalties from having had children experienced by heterosexual women. The most important difference in terms of choices made by heterosexual and same-sex mothers is that heterosexual mothers experience more sustained drops in labor market hours. To summarize the main takeaway from these results, the differences between the child penalties of heterosexual and same-sex mothers seem to be largely driven by differences in the response on the intensive margin as opposed to responses on the extensive margin (both in terms of exiting the labor market or switching occupations).41

6 The impact of family friendly policies

Despite the persistence of the child penalty within heterosexual couples, Online Appendix Figure B11 suggests that decreases in the child penalty are possible, as we find large decreases in the penalty from 1971 to 2010. In this last part of the paper, we explore

⁴¹It would be interesting to see if these decomposition results replicate in other countries. Unfortunately, we lack the data to do a full cross-country comparison of the decomposition of the child penalty across countries.

the causal impacts of two policies that occurred during this period, paternity leave and early subsidized childcare, that could theoretically be effective at reducing the child penalty, given the mechanisms identified in the previous section.

6.1 Paternity leave

As a means of increasing fathers' involvement in raising children, the so called daddy quotas (leave that can only be taken by fathers) of the Scandinavian countries have attracted considerable interest.⁴² "Use it or lose it" paternity leave, by strongly encouraging fathers to spend more time with their children, might increase the value fathers place on time with children and might also decrease the importance of gender norms.⁴³ Within the framework of our model and based on the mechanisms identified in the previous section, both of these effects could decrease the child penalty. Thus, while paternity leave policy changes cannot be used to isolate individual mechanisms, such leave could theoretically reduce the child penalty. In this section we test that hypothesis.

⁴²In addition to Scandinavian countries, 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) (OECD (2014)). A number of firms in the United States also offer paternity leave.

⁴³Paternity leave could also increase the productivity of fathers in home production (increasing k_i in equation (2)), although the previous section suggests this is not an important mechanism.

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

Table 1: Parental leave reforms in Norway, in weeks

Note: Parental leave in weeks. Numbers in parenthesis (except maternal quota) indicate weeks of leave if taken at 80 percent compensation, otherwise at 100 percent. *Source:* NOU 2017:6 (2017).

In Table 1 we report every leave reform in Norway from 1992 to 2014. The maternal and paternal quota columns report the amount of parental leave in weeks that is reserved exclusively for the mother and father. The remaining leave can be shared among parents however they choose and is reported in column 6. The reforms were generally announced in October the year before implementation as part of the budgeting process, making it nearly impossible to plan conception in response to the announcement of the quota change in order to manipulate birth dates around the cutoff in April or July. In Online Appendix 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.⁴⁴ 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, whereas parents of children born right after each reform were subject to the

⁴⁴We exclude the 1992 reform because it requires a donut-RD framework to identify the effects, which is not necessary for the other results.

changes. For this exercise, we draw on heterosexual couples with first children born in each reform year.⁴⁵ We set leave to zero for fathers where we observe no leave take-up.⁴⁶ We rely on the following fuzzy regression discontinuity (RD) setup separately for both mothers' and fathers' earnings measured at each event time t relative to child birth:

$$y_{it} = \beta_t L_i + f_t(x_i) + \epsilon_{it}$$

$$L_i = \gamma \mathbb{1}(x_i \ge 0) + g(x_i) + \eta_{it}$$
(12)

where x_i , the running variable, is the number of days after the reform date that the child was born. $f_t(x_i)$ and $g(x_i)$ are local linear polynomials that are separate on either side of the cutoff. We use the optimal bandwidth that minimizes the mean squared error of the RD estimate to define the sample of births we use, and a triangular weighting function in order to obtain estimates local to the cutoff. We estimate and report robust bias-corrected confidence intervals (Calonico *et al.*, 2014) together with the conventional, heteroskedasticity-robust confidence intervals. We then scale the effects on earnings to reflect the percentage changes in the child penalty.⁴⁷ 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. We provide empirical support for this assumption using balancing tests in Online Appendix Table C1.⁴⁸

⁴⁵Because we want to capture mothers and fathers exposed to the leave reforms, we include in the sample only couples where the mother took some leave, indicating that she is eligible, because users of the alternative one-time benefit would not be affected.

⁴⁶Online Appendix A provides additional details on the construction of our parental leave measure.

⁴⁷Many models in this section are estimated using the robust RD commands for Stata written by Matias D. Cattaneo et al., to whom we owe thanks. These include rdrobust, rddensity, rdbwselect and others. These packages are documented in Calonico *et al.* (2018) and Cattaneo *et al.* (2018).

⁴⁸An important imbalance revealed in this table is maternity leave take-up, as some of the reforms



Figure 4: Fuzzy RD first stage estimate

Note: First stage estimates for each reform, using local linear polynomials, triangular weights and optimal bandwidths. Top numbers are weeks of paternity leave quota.

we exploit increase paternity leave quota at the expense of the shared leave most often taken by the mother. We do not believe these relatively small changes in maternity leave take-up from already high levels to be driving our results. In Online Appendix C.3, we exploit the fact that some of these reforms expanded paternity leave use at the expense of maternity leave, while others expanded the total leave length. This allows us to instrument for both the maternity and paternity leave use, confirming the baseline results of the effects of paternity leave on the child penalty.
Reform year	2005	2006	2009	2011	2013	2014	Pooled	Stacked
RD estimate per week	0.79**	1.05***	0.98***	0.82***	0.69***	0.81***	0.86***	0.88***
conv. standard error	(0.33)	(0.33)	(0.095)	(0.28)	(0.23)	(0.11)	(0.10)	(0.067)
robust standard error	0.40	0.40	0.11	0.33	0.28	0.14	0.12	
conventional p-value	0.016	0.002	0.000	0.003	0.003	0.000	0.000	0.000
robust p-value	0.031	0.007	0.000	0.006	0.019	0.000	0.000	
Observations	14,598	15,111	16,501	16,500	16,173	14,240	93,123	93,123
Optimal bandwidth	64.7	54.1	73.2	47.8	65.3	42.5	60.6^{+}	60.6^{\dagger}
Efficient observations	5,302	4,797	6,877	4,659	6,264	3,685	31,584	31,584

Table 2: RD first stage estimates

Notes: Robust semiparametric RD estimates of the effect of paternity leave reforms on paternity leave take-up 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 refer to the weighted average of reform-specific estimates. Stacked estimates stacks models for each reform, allowing polynomials to vary over cutoffs and using the cutoff-specific bandwidths and weights. Conventional standard errors are heteroskedasticity-robust, but not bias-corrected. ***p < 0.01, **p < 0.05, *p < 0.1, using conventional, heteroskedasticity-robust standard errors. [†]Average bandwidth.

We see clear effects of all reforms on the take up of paternity leave in Figure 4. The first stage estimates are always significant, whether using robust bias-correcting inference or conventional inference that only accounts for heteroskedasticity.⁴⁹ In order to increase precision, we next combine the six reforms. The common way of stacking multiple reforms in RD studies is to re-center the running variable to be zero at the relevant cutoff for all individuals and run semiparametric RD estimates in the pooled sample, which restricts the functional form of the polynomials and the optimal bandwidth to be the same for each reform. In addition to this restriction, naive re-centering is problematic in our case because the treatment scaling varies across reforms, from a decrease of four weeks to an increase of four weeks and various changes in between. 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 ker-

⁴⁹Note that all results have been scaled to reflect one week of quota expansion.

nel weights from the individual specifications. The results are scaled to reflect one week of quota expansion by using an indicator of the number of weeks of quota increase rather than a dummy at the cutoff. Unfortunately, we cannot calculate bias-corrected standard errors for this specification, but we argue that the problem should be relatively minor.⁵⁰ Our preferred first stage estimate from the stacked specification indicates that granting fathers another week of paternity leave quota increases leave take-up by .88 weeks.

For the stacked fuzzy RD, we revert to the cutoff-specific treatment indicators as instruments because the fuzzy RD takes care of scaling. This specification reproduces the cutoff-specific first stage estimates for each reform reported in Figure 4 for a given bandwidth and so is a natural way to stack the reforms. When interpreting these fuzzy RD estimates, it is important to keep in mind that these estimates are local average treatment effects: they capture the effects of additional leave use on earnings for couples induced to use more leave because they were exposed to the reforms. 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 (see column 5 of Table 1). In case of heterogeneous treatment effects, the average effect for the compliers need not be the same as that for the population. Despite this, we argue that the LATE is a particularly policy-relevant treatment effect in our case because it reflects the effects of paternity leave use for fathers induced to take more leave by the policy instrument, which is arguably the population of interest to policy makers.

⁵⁰First, notice that the difference between the conventional and the robust standard error estimate for the reform-specific cutoffs in Table 2 are 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 alternative naive pooled estimator 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.



Figure 5: Fuzzy RD estimates of paternity leave use on maternal child penalty

Notes: The figure shows fuzzy RD estimates of the impact of an additional week of paternity leave use on the mother's child penalty, using all six reforms. The pooled estimate refers to the weighted average of the reform-specific estimates, while the stacked estimate stacks the cutoff-specific specifications for precision. Robust bias-correcting inference reported for the pooled estimate and conventional, heteroskedasticity-robust inference for the stacked estimate.

Figure 5 reports the impacts on the child penalty using the the stacked and pooled fuzzy RD estimates. The *y*-axis in this figure represents the percentage change in the child penalty, as estimated from the event studies from the first half of the paper. ⁵¹ Point estimates are close to zero, suggesting no impact of paternity leave on the child penalty. This zero is relatively precise, as the lower bound of the confidence intervals rules out reductions larger than around 5 to 7 percent of the maternal child penalty per week of paternity leave use for children ages 1 through 5.⁵²

⁵¹Effects on mothers' and fathers' annual incomes are reported in Appendix Figure C2.

⁵²One might believe that paternity leave could have long run effects on norms. However, the first paternity leave reform occurred in 1992 in Norway, which begs the question, how long should one wait to see long run effects?

While these results suggest paternity leave does not substantially reduce the child penalty, such leave might influence gender norms or preferences around the distribution of home work in ways that do not influence labor market earnings. One possible measure of such norms is an increased use of shared leave by fathers for future children. To investigate whether paternity leave use has a direct effect on the father's choice to spend time with his children, we exploit the fact that many of the fathers that have a child around the time of the reforms subsequently go on to have more children. We therefore estimate our fuzzy RD model using as an outcome the father's leave take-up for the next child for all children born up to and including 2014 in a setup similar to the peer effects estimates from Dahl *et al.* (2014).⁵³ In order not to use the outcome variable for one child as the treatment variable for another, we restrict attention to the first child each father has in one of the reform years and look at outcomes for the next child. Notice that we cannot use the 2014 reform for this exercise, as we cannot reliably measure paternity leave use for kids born after 2014.

⁵³Notice that if fertility was endogenous to the parental leave reforms, this might constitute an endogenous sample selection criteria. Hart *et al.* (2019) investigate fertility response to the 2009 reform and find no evidence of such effects, but Farré and González (2019) find negative impacts of paternity leave on fertility in Spain.

Reform year	2005	2006	2009	2011	2013	Pooled	Stacked
RD estimate per week	-0.313	-0.501	-0.0624	-0.256	0.79**	-0.268	-0.092
conv. standard error	(1.08)	(0.499)	(0.123)	(0.314)	(0.359)	(0.359)	(0.11)
robust standard error	1.32	0.592	0.147	0.366	0.404	0.412	
conventional p-value	0.772	0.315	0.612	0.416	0.003	0.428	0.40
robust p-value	0.873	0.284	0.605	0.403	0.024	0.491	
Observations	14,201	14,761	14,086	9,704	704	53,456	53,456
Optimal bandwidth	60.9	48.4	60.4	46.8	46.6		
Efficient observations	4,821	4,245	4,893	2,778	159	16,896	16,896

Table 3: Paternity norms: Fuzzy RD of paternity leave on leave for next child

Notes: Fuzzy RD estimates of the impact of one more week of paternity leave for a child on the weeks of paternity leave use for the next. Conventional standard errors are heteroskedasticity-robust but not bias corrected. ***p < 0.01, **p < 0.05, *p < 0.1 based conventional standard errors.

Table 3 provides the results of this exercise, for each reform separately and the pooled and stacked estimates for all reforms. Across the rows, we see little evidence of any permanent impact on norms as measured by take-up of paternity leave for later kids: except for the 2013 reform, where the efficient sample size is only 159 children and we find a marginally significant effect, none of the reforms provide statistically significant results, and point estimates are negative. Focusing on our preferred stacked estimates, the results indicate a non-significant effect of .1 *less* weeks of leave for subsequent children for each week of paternal leave taken for the first child, where the top of the 95 percent confidence interval rules out effects larger than around 0.12 week extra leave for subsequent kids per week of leave for the first child.

6.2 Improved access to early child care

An alternative approach to reduce the child penalty is for the government to reduce the price (reduce p in equation (3) of the model) of a high-quality substitute for mother's time In this section, we estimate the impact on the child penalty of providing high-quality sub-

sidized child care to mothers in Norway. Online Appendix Figure D5a shows the child care coverage rates over time in Norway, separately by age of the children. These figures show that the formal care sector for preschoolers was well developed in Norway by the early 2000's, with more than 80 percent of Norwegian 4 and 5 year olds attending care.⁵⁴ For toddlers (aged 1 to 3 years), however, coverage was much lower (between 30 to 50 percent), and the market was strongly rationed. These facts are documented in greater detail in Andresen and Havnes (2019), including additional evidence from surveys on the actual and preferred modes of child care for children at these ages. The underrepresentation of children aged 1 to 3 in formal care was the impetus for the Child Care Concord in 2002, a broad, bipartisan agreement to increase the availability of care for toddlers. Following this reform, coverage increased rapidly for 1- and 2-year-olds over the next years as shown in Online Appendix Figure D5a. However, the expansion varied considerably between municipalities and over time (see online Appendix Figure D5b), making the expansion of care availability a potential instrument for the endogenous choice of how much child care to use. This is the variation exploited to estimate the effects of formal care use in Andresen and Havnes (2019). In this section we use the same variation to estimate the impact on the child penalty of increasing access to high-quality formal child care.

For this application, we start with all children born in the years 2000-2006, who will be subject to the reform-induced expansions of care in 2002-2008.⁵⁵ We assign children to their municipality of residence at the age of 1 and look at couples where both parents reside in that municipality when the child is 1. While much of the literature restricts the sample to children without younger siblings, we view future fertility as a potentially

⁵⁴The prevalence of care is the result of a reform and gradual expansion of formal care for these children in the 1970's (Havnes and Mogstad, 2011).

⁵⁵This includes a few thousand twins. Clustering at the municipality level accounts for within-family clustering.

endogenous outcome of the reform, and therefore do not restrict the sample to youngest children. To be consistent with the results we have presented thus far, we look at the effects on maternal and paternal income from the child's birth to 5 years of age, and use the years before birth as placebo outcomes. This leaves us with a sample of around 103,000 couples.⁵⁶

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

$$y_{it} = \pi_k + \gamma_{T_{it}} + \beta_{a_{it}} + \phi_t m_i + \epsilon_{it}$$

$$m_i = \tilde{\pi}_k + \tilde{\gamma}_{T_{it}} + \tilde{\beta}_{a_{it}} + \gamma_1 C C_k^1 + \gamma_2 C C_k^2 + \tilde{\epsilon}_{it}$$
(13)

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

The variation we exploit thus comes from the variation in expansion of care across municipalities and over time. As long as the exact timing of expansion of care is uncorre-

⁵⁶Notice that because we restrict the sample in this paper to first born children, it is a little less than half the size of the samples of cohabiting mothers and fathers in Andresen and Havnes (2019). This gives us less precision but is consistent with the rest of the paper. Because of the inherent focus on labor supply over time, we also measure child care use throughout the full 13 - 36 months period we can measure, in contrast to the preceding paper that is mostly concerned with child care use and labor supply during the calendar year the child turns two.

lated with other drivers of parents' child penalty, our approach recovers the causal effect of an extra year of early child care on the child penalty 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 were restricted by the low supply. Andresen and Havnes (2019) shows that the exact timing of expansion was subject to a range of constraints that were hard to predict, and the timing of expansion was not necessarily easy to predict even for the municipalities themselves. Online Appendix Figure D6 provides some support for the idea that expansions did not systematically vary across municipalities with different pre-reform characteristics (except, of course, the initial coverage rate), while Andresen and Havnes (2019) provide a range of specification checks that demonstrate the robustness of the instrument.

Years of child care use at ages 13 - 35 or 36 months 0.787*** (0.0543)0.764*** (0.0552)Coverage rate at age 1 0.630*** 0.632*** Coverage rate at age 2 (0.0629)(0.0660)Municipality fixed effects ~ ~ Year fixed effects √ Age profiles Education-specific age profiles N103,172 103,157 mean dep. var 1.031 1.031 F167.9 138.5

Table 4: First stage estimates, formal care use

Note: First stage estimates of equation (13) for mothers. Point estimates for fathers (not shown) are very similar. Standard errors in parentheses, clustered at municipality.

First stage estimates from this specification are presented in Table 4, column 1, where we see that expansions of care both at age 1 and at age 2 have a strong impact on early child care use, with an additional slot in care at age 1 increasing care use by around 0.8 years and at age 2 by about 0.6 years. Because our endogenous variable

captures the intensity of use throughout the full period, these coefficients are not 1; as additional slots are generally opened in August, children may not have the chance to exploit them to capacity the whole year. The IV strategy thus scales the reduced form estimates to reflect a full year of early child care use. The F-statistic is above 150, indicating a very strong first stage.



Figure 6: Impact of a year of early child care use on mother's child penalty *Note:* IV results from equation (13) reflecting the impact on mother's labor earnings in 1,000 NOK across child age for an extra year of early child care use at ages 13-36 months, scaled with the estimated child penalties from the first part of the paper to represent the change in the child penalty for mothers.

In Figure 6, we report estimates from the second stage, scaled with the estimated baseline child penalties to present the relative effect of a full year of child care use on the child penalty. The baseline model discussed so far is indicated with diamonds. Results show that the child penalty is reduced by around 25 to 30 percent for mothers when their children are between the ages 2 and 3, but the impact appears only in the years of treatment (note that age 3 is included in treatment since some children will receive slots

just before turning 3, so treatment will also occur at age 3 for these kids). In Online Appendix Figure D7 we present results separately for mothers' and fathers' earnings. These results show that the main impact is on mothers, who see significant increases in their labor market earnings. As a robustness check, we include the education level-specific age profiles in equation 13. The first stage from this specification is hardly affected by this, as documented by column 2 in table 4. The second stage results are also very similar. We conclude that early child care shows more promise as a policy tool for reducing child penalties in heterosexual couples than paternity leave, although it does not appear to have a permanent impact.

7 Conclusion

In the first half of this paper we show that female same-sex couples experience a very different child penalty than that of heterosexual and adopting couples. Based on our results we are able to largely rule out two of the most common explanations for the child penalty in heterosexual couples, the costs of giving birth and comparative advantage, although the costs of giving birth may play a small role in the first two years after birth. This leaves gender norms and preferences over child care as the most likely mechanisms behind the child penalty. With these mechanisms in mind, we then turn to two policies that might be effective at reducing the child penalty for heterosexual couples: paternity leave and subsidized early child care. We find that while fathers take more paternity leave when exposed to a non-transferable quota, paternity leave has no impact on the child penalty. In addition, paternity leave has no impact on whether the father takes additional leave for future children, pointing to limited impact on gender norms. In contrast, we show that early child care use reduces the child penalty for mothers by

around 25 percent per year of use in the years of treatment. These results suggest that if policy makers wish to decrease the child penalty, they should focus on providing better child care to families, not on offering paternity leave to fathers.

Our paper sheds light on both why the child penalty occurs and how policy might impact the child penalty. While we have focused on two of the most commonly proposed policies to reduce the child penalty, there are a number of additional policy changes that could impact the child care penalty differently. Better understanding the impacts of different policies, as well as further disentangling the relative importance of gender norms and preferences, are both productive avenues for future research.

References

- ABRAHAM, S. and SUN, L. (2020). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. Papers 1804.05785, arXiv.org.
- ADDA, J., DUSTMANN, C. and STEVENS, K. (2017). The career costs of children. *Journal of Political Economy*, **125** (2), 293 337.
- AKGUNDUZ, Y. E. and PLANTENGA, J. (2018). Child care prices and maternal employment: A meta-analysis. Journal of Economic Surveys, 32 (1), 118–133.
- AKSOY, C. G., CARPENTER, C. S. and FRANK, J. (2018). Sexual orientation and earnings: New evidence from the united kingdom. *ILR Review*, **71** (1), 242–272.
- ALDEN, L., EDLUND, L., HAMMARSTEDT, M. and MUELLER-SMITH, M. (2015). Effect of registered partnership on labor earnings and fertility for same-sex couples: Evidence from Swedish register data. *Demography*, **52** (4), 1243–1268.
- ANDERSEN, S. H. (2018). Paternity leave and the motherhood penalty: New causal evidence. Journal of Marriage and Family, 80 (5), 1125–1143.
- ANDRESEN, M. E. and HAVNES, T. (2019). Child care, parental labor supply and tax revenue. *Labour Economics*, p. 101762.
- ANGELOV, N., JOHANSSON, P. and LINDAHL, E. (2016). Parenthood and the gender gap in pay. *Journal* of Labor Economics, **34** (3), 545–579.
- ANTECOL, H., BEDARD, K. and STEARNS, J. (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies? *American Economic Review*, **108** (9), 2420–41.

- and STEINBERGER, M. D. (2013). Labor Supply Differences Between Married Heterosexual Women and Partnered Lesbians: A Semi-Parametric Decomposition Approach. *Economic Inquiry*, **51** (1), 783–805.
- BADGETT, M. L. and HYMAN, P. (1998). Explorations introduction: Towards lesbian, gay, and bisexual perspectives in economics: Why and how they may make a difference. *Feminist Economics*, 4 (2), 49–54.
- BAGUES, M., SYLOS-LABINI, M. and ZINOVYEVA, N. (2017). Does the gender composition of scientific committees matter? *American Economic Review*, **107** (4), 1207–38.
- BANA, S., BEDARD, K. and ROSSIN-SLATER, M. (2018). Trends and disparities in leave use under california's paid family leave program: New evidence from administrative data. In AEA Papers and Proceedings, vol. 108, pp. 388–91.
- BAUMLE, A. K. (2009). The cost of parenthood: Unraveling the effects of sexual orientation and gender on income. *Social Science Quarterly*, **90** (4), 983–1002.
- BENSNES, S., LEUVEN, E. and HUITFELT, I. (2019). Dynamic Compliance to IVF-Treatments How Big Is the Child Penalty.
- BERGSVIK, J., KITTERØD, R. H. and WIIK, K. A. (2019). Parenthood and couples' relative earnings in Norway 2005-2014. Working Paper, no. 873.
- BLACK, D. A., SANDERS, S. G. and TAYLOR, L. J. (2007). The economics of lesbian and gay families. *Journal of Economic Perspectives*, 21 (2), 53–70.
- BLAU, D. and CURRIE, J. (2006). Pre-School, Day Care, and After-School Care: Who's Minding the Kids?, Elsevier, Handbook of the Economics of Education, vol. 2, chap. 20, pp. 1163–1278.
- BORUSYAK, K. and JARAVEL, X. (2018). Revisiting event study designs. Available at SSRN: https://ssrn.com/abstract=2826228 or http://dx.doi.org/10.2139/ssrn.2826228, sSRN.
- BURSZTYN, L., FUJIWARA, T. and PALLAIS, A. (2017). 'Acting Wife': Marriage market incentives and labor market investments. American Economic Review, 107 (11), 3288–3319.
- CALONICO, S., CATTANEO, M. D., FARRELL, M. H. and TITIUNIK, R. (2018). RdrobUSt: Stata module to provide robust data-driven inference in the regression-discontinuity design.
- –, and TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82 (6), 2295–2326.
- CARPENTER, C. S. (2005). Self-reported sexual orientation and earnings: Evidence from california. *ILR Review*, **58** (2), 258–273.
- (2007). Revisiting the income penalty for behaviorally gay men: Evidence from nhanes iii. Labour Economics, 14 (1), 25 34.
- (2008). Sexual orientation, work, and income in canada. Canadian Journal of Economics/Revue canadienne d'économique, 41 (4), 1239–1261.
- (2009). Sexual orientation and outcomes in college. Economics of Education Review, 28 (6), 693 703.

- and Еррілк, S. T. (2017). Does it get better? recent estimates of sexual orientation and earnings in the united states. Southern Economic Journal, 84 (2), 426–441.
- CATTANEO, M. D., JANSSON, M. and MA, X. (2017). Simple local polynomial density estimators.
- -, and (2018). Manipulation Testing Based on Density Discontinuity. Stata Journal, 18 (1), 234-261.
- CHOI, H.-J., JOESCH, J. M. and LUNDBERG, S. (2008). Sons, daughters, wives, and the labour market outcomes of west german men. *Labour economics*, **15** (5), 795–811.
- CHUNG, Y., DOWNS, B., SANDLER, D. H. and SIENKIEWICZ, R. (2017). The Parental Gender Earnings Gap in the United States. (17-68).
- COOLS, S., FIVA, J. H. and KIRKEBØEN, L. J. (2015). Causal effects of paternity leave on children and parents. *The Scandinavian Journal of Economics*, **117** (3), 801–828.
- DAHL, G. B., LØKEN, K. V. and MOGSTAD, M. (2014). Peer effects in program participation. *American Economic Review*, **104** (7), 2049–74.
- DEL BOCA, D. and FLINN, C. (2012). Endogenous household interaction. *Journal of Econometrics*, **166** (1), 49–65.
- DOBKIN, C., FINKELSTEIN, A., KLUENDER, R. and NOTOWIDIGDO, M. J. (2018). The Economic Consequences of Hospital Admissions. *American Economic Review*, **108** (2), 308–352.
- DOEPKE, M. and KINDERMANN, F. (2019). Bargaining over babies: Theory, evidence, and policy implications. *American Economic Review*, **109** (9), 3264–3306.
- DRUEDAHL, J., EJRNÆS, M. and JØRGENSEN, T. H. (2019). Earmarked paternity leave and the relative income within couples. *Economics Letters*, **180**, 85 88.
- and MARTINELLO, A. (2016). Long-Run Saving Dynamics: Evidence from Unexpected Inheritances. (2016:7).
- EKBERG, J., ERIKSSON, R. and FRIEBEL, G. (2013). Parental leave: A policy evaluation of the Swedish daddy month reform. *Journal of Public Economics*, **97**, 131 143.
- EVERTSSON, M. and BOYE, K. (2018). The transition to parenthood and the division of parental leave in different-sex and female same-sex couples in Sweden. *European Sociological Review*, **34** (5), 471–485.
- FADLON, I. and NIELSEN, T. H. (2017). Family health behaviors. (24042).
- FARRÉ, L. and GONZÁLEZ, L. (2019). Does paternity leave reduce fertility? *Journal of Public Economics*, 172, 52–66.
- FERNÁNDEZ, R., FOGLI, A. and OLIVETTI, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, **119** (4), 1249–1299.
- FEYRER, J., SACERDOTE, B. and STERN, A. D. (2008). Will the stork return to Europe and Japan? understanding fertility within developed nations. *Journal of Economic Perspectives*, **22** (3), 3–22.

GALLEN, Y. (2018). Motherhood and the gender productivity gap.

- GOLDBERG, A. E., SMITH, J. Z. and PERRY-JENKINS, M. (2012). The division of labor in lesbian, gay, and heterosexual new adoptive parents. *Journal of Marriage and Family*, **74** (4), 812–828.
- GOLDIN, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, **104** (4), 1091–1119.
- GOODMAN-BACON, A. (2018). Difference-in-Differences with Variation in Treatment Timing. Working Paper 25018, National Bureau of Economic Research.
- HART, R. K., ANDERSEN, S. N. and DRANGE, N. (2019). Effects of extended paternity leave on union stability and fertility. No. 899.
- HAVNES, T. and MOGSTAD, M. (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, **95** (11–12), 1455 – 1465, Special Issue: International Seminar for Public Economics on Normative Tax Theory.
- JACOBSON, L. S., LALONDE, R. J. and SULLIVAN, D. G. (1993). Earnings losses of displaced workers. *The American Economic Review*, **83** (4), 685–709.
- JOHANSSON, E.-A. (2010). The effect of own and spousal parental leave on earnings. (4).
- KLEVEN, H., LANDAIS, C., POSCH, J., STEINHAUER, A. and ZWEIMULLER, J. (2019a). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*.
- -, and SøGAARD, J. E. (2019b). Children and gender inequality: Evidence from Denmark. American Economic Journal: Applied Economics.
- -, and SOGAARD, J. E. (2020). Does biology drive child penalties? evidence from biological and adoptive families. Working Paper.
- LUNDBERG, S. and Rose, E. (2000). Parenthood and the earnings of married men and women. *Labour Economics*, **7** (6), 689–710.
- and (2002). The effects of sons and daughters on men's labor supply and wages. *Review of Economics and Statistics*, 84 (2), 251–268.
- LUNDBORG, P., PLUG, E. and RASMUSSEN, A. W. (2017). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review*, **107** (6), 1611–37.
- MARTELL, M. E. and RONCOLATO, L. (2016). The homosexual lifestyle: Time use in same-sex households. *Journal of Demographic Economics*, **82** (4), 365–398.
- McCrary, J. (2007). The effect of court-ordered hiring quotas on the composition and quality of police. *American Economic Review*, **97** (1), 318–353.
- MOBERG, Y. (2016). Does the gender composition in couples matter for the division of labor after childbirth?
- MORRISSEY, T. W. (2016). Child care and parent labor force participation: A review of the research literature. *Review of Economics of the Household*, pp. 1–24.

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

NOVGORODSKY, D. and SETZLER, B. (2019). Practical guide to event studies.

- OECD (2014). OECD family database: PF2.1 key character of parental leave systems. available at http://www.oecd.org/els/soc/PF2_1_Parental_leave_systems_ 1May2014.pdf.
- OLIVETTI, C. (2006). Changes in women's hours of market work: The role of returns to experience. *Review of Economic Dynamics*, **9** (4), 557–587.
- PATNAIK, A. (2019). Reserving time for daddy: The consequences of fathers' quotas. *Journal of Labor Economics*, **0**, null, forthcoming.
- PERSSON, P. and ROSSIN-SLATER, M. (2019). When dad can stay home: Fathers' workplace flexibility and maternal health. NBER Working Paper No. 25902.
- REGE, M. and SOLLI, I. F. (2013). The impact of paternity leave on fathers' future earnings. *Demography*, **50** (6), 2255–2277.

ROSENBAUM, P. (2019). The family earnings gap revisited: A household or a labor market problem?

ROSSIN-SLATER, M. (2017). Maternity and family leave policy.

RUDLENDE, L. and LIMA, I. (2018). Medmødre tilpasser seg også fedrekvoten. Arbeid og velferd, (3).

- SANSONE, D. (2019). Pink work: Same-sex marriage, employment and discrimination. *Journal of Public Economics*, **180**, 104086.
- SCHMIDHEINY, K. and SIEGLOCH, S. (2019). On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications. CEPR Discussion Papers 13477, C.E.P.R. Discussion Papers.
- SCHNEEBAUM, A. (2013). Motherhood and the lesbian wage premium. Number 2013-4.
- SIEPPI, A. and PEHKONEN, J. (2019). Parenthood and gender inequality: Population-based evidence on the child penalty in finland. *Economics Lettersabra*, **182**, 5 9.

Online Appendix

Appendix A provides details on how we identify heterosexual and female same-sex couples and their children in the data, and reports summary statistics. Appendix B provides additional results and robustness checks when estimating and comparing child penalties across couple types. Appendix C contains robustness checks and additional results for the paternity leave application in Section 6.1 in the main paper, while Appendix D contains the same for the child care application from Section 6.2 in the main paper.

A Details on sample selection and summary statistics

To construct our sample, we rely on registrations of legal parent status in the population registers. In practice, we therefore observe children appearing of female same-sex couples appearing in the data at various times following birth, given the laws described above. 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. 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 the first year until the legal adoption procedure is completed. We restrict attention to children whose parents were both legally registered as their parents in the year the child turns 1 at the latest 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. Note that we drop a handful of female same-sex couples who are registered to multiple kids in the same year and register different parent status for each child. Since more 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. We therefore restrict the window of interest to be between t-4 and t+5 to limit this imbalance.

In Figure A7 we graph the adoption of children by age and year to female same-sex couples. In Table A5 we report summary statistics for the heterosexual and same-sex female couples used in Section 5.



Figure A7: Children registered to same-sex female 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.

	Heterosez	cual couples	Same-sex female couples					
Birth year (first child)	200 : Child cha	1-2014 aracteristics	2001-2014					
Birth year	20	007.7	2010.6					
Multiple birth	(4 0	4.00) .020		(2.89) 0.067				
Female child	(0).14)).49		(0.25) 0.48				
Age at adoption	(0 0 (0	0.50) .028 .179)	(0.49) 0.48 (0.81)					
B: Parent characteristics, year before birth								
Parent type (K)	Mother 1	Father 2	Mother 3	Co-mother 4				
Age at first birth	27.8	30.3	32.2	32.8				
Labor income (1,000s of 2017 NOK) Years of education [†]	(4.23) 339.6 (206.0) 15.1 (2.01)	(5.03) 471.9 (1055.8) 14.6 (2.01)	(4.12) 488.9 (196.7) 16.4 (2.42)	(5.64) 480.0 (308.3) 16.0 (2.65)				
Days of sickness absence vear $t-2$	(2.91) 7.25 (31.7)	(3.01) 6.7 (31.2)	(2.42) 11.7 (37.8)	(2.03) 14.9 (46.5)				
N couples	25	1,490	634					

Table A5: Summary statistics by couple type

Note: Summary statistics

on estimation samples constructed as described in this section. Standard deviations in parentheses.

B Child penalties: Additional results and robustness

Raw mean earnings over time for heterosexual and same-sex female couples are found in Figure B1 together with the simple event study estimates that omit age- and year fixed effects.



Figure B1: Mean earnings by event time (top) and raw child penalties (bottom) *Note:* Top panel show show means of annual labor earnings for the years before and after birth of the first child. Bottom panels show simple event study estimates without year and age fixed effects. Sample construction and data as defined in Section 4.

Figure B2 provides a subsample analysis by (birthing) mother's education, revealing relatively similar effects across groups. Figure B3 shows that the same patterns for comparative advantage still hold when we use pre-birth income differences as a proxy for comparative advantage instead of pre-birth education difference within the couple. Last, Figure B4 shows that the pattern remains similar even when we restrict the sample to heterosexual couples where pre-pregnancy the woman in the couple makes more than her male partner, also in couples where both partners earned more than 300,000 NOK before pregnancy (around 33,000 USD).



Figure B2: Subsample analysis by level of mother's education: high school or below vs. more than high school



Figure B3: Controlling comparative advantage using income differences in t-4



(a) All couples with female breadwinners

(b) Couples with female breadwinners where both spouses earn more than 300,000 NOK

Figure B4: Child penalties in heterosexual couples with female breadwinners *Note:* Female breadwinners measured in year t-1 as couples where the female has higher labor market earnings than the male.

B.1 Robustness to restricted age profiles and yearly shocks

In our baseline model of child penalties (see equation(10)) we specify an event study model that controls for age profiles and yearly shocks. While our sample of female same-sex couples is large relative to that of most other studies of these couples, we have limited precision to estimate child penalties when accounting fully flexibly for age profiles and yearly shocks separately for each parent type, essentially estimating equation (10) for each parent type. In our baseline model, we resolve this by restricting the (fully flexible) age profiles and yearly shocks to be gender but not parent-type specific. This allows us to use the large sample of heterosexual mothers to estimate female-specific age profiles and yearly shocks while retaining power to identify the child penalties for same-sex couples, at the cost of imposing parametric restrictions. In this section, we show that this restriction does not seem to be driving our results.

First, as we know that same-sex female couples are on average better educated than heterosexual couples, we might worry that they have a different age profile,⁵⁷ entering the labor market later but being on a steeper part of the age-earnings profiles than heterosexual mothers at the time of first birth. If this is the case, the restriction that the age profiles be the same for all mothers may force the difference to leak into the estimated child penalties for same-sex mothers. The most straightforward way to test this is to allow the age profiles (and yearly shocks) to be not only be gender-specific, but also education level specific, where education is measured in nine levels the year before birth. The results from this exercise are displayed in Figure B5. While the child penalty is somewhat smaller and there is some limited catch-up among heterosexual couples when accounting for these flexible age profiles, the contrast with female same-sex couples is still remarkable and comparable to the baseline estimates.

⁵⁷We discuss this in the context of age profiles, but similar arguments can be made for yearly shocks, and when implementing alternative models, we relax both.



Figure B5: Controlling for education- and gender-specific age profiles

Second, we can relax the restriction altogether and use the estimated age profiles and yearly shocks to test the restriction we impose. When relaxing this restriction, we struggle with very limited support of same-sex couples in some age groups (below 23 and above 45, approximately) and for some (early) calendar years because the sample is unbalanced over time. This leads to very low precision for these age profiles and yearly shocks and therefore also low precision for the child penalties, which are scaled by these imprecisely estimated coefficients and bootstrapped. When estimating the fully flexible model below, we therefore restrict our sample to couples where both spouses are between 25 and 40 at the age of first birth and who gave birth in 2004 or later, reducing sample size to around 525 same-sex female couples. Figure B6 plots various estimated parameters from this model. The top left panel shows the estimated age profiles, which are similar for all three mother types. Indeed, we cannot reject that they are the same (p = 0.44 for same-sex birth mothers, p = 0.13 for same-sex co-mothers, p = 0.19 for both). Likewise, estimated yearly shocks are relatively similar in the top right panel, and again we cannot reject the restriction we make in the baseline specification that all yearly shocks are the same for all mothers (p = 0.22 for same-sex birth mothers, p = 0.28 for same-sex co-mothers, p = 0.16 for both). When testing the age profiles

and yearly shocks together, again we cannot reject that they are the same (p=0.21 for same-sex birth mothers, p=0.30 for same-sex co-mothers), while there is only marginal significance when testing all four sets of coefficients against the same coefficients for heterosexual mothers (p=0.08). This is some support for the restrictions we make.

In the bottom left panel of Figure B6, we report the estimated raw penalties for the three types of mothers. These are relatively similar to the baseline estimates. We can strongly reject that the raw penalties are the same for same-sex co-mothers and heterosexual mothers (p=0.000), while this is only approaching significance for the difference between the absolute child penalties for heterosexual and same-sex mothers (p=0.15). Remember, however, that baseline earnings are significantly higher for same-sex compared with heterosexual mothers, so that these slightly smaller raw estimates for same-sex mothers translate into much smaller estimates relative to income, as shown in the bottom right panel. Here, the estimated child penalties are relatively similar to the reported baseline estimates, although the catchup is perhaps slightly less pronounced for same-sex mothers over time, and the large difference between the child penalties for same-sex and heterosexual mothers remain. As before, we can strongly reject that the child penalties are the same (p=0.000 for both same-sex mothers). We take this as evidence that the restriction we make in the baseline model is not driving the main results of starkly different responses to the arrival of children among heterosexual and same-sex female couples.





Note: Estimates from a fully flexible model where age profiles and yearly shocks vary across parent types. Sample restricted to first birth between the ages of 25 and 40 for both spouses and children born in 2004 or later to limit imbalance.

B.2 Additional robustness results

Figure B7 shows the mean number of children over our sample window by couple type, revealing that heterosexual and same-sex female couples have very similar (although not perfectly identical) completed fertility. However, given how similar these completed fertility patterns are, we conclude that the differences in child penalties are not driven by the differences in number of additional children following the first child. Moreover, in Figure B8 we show that the estimated child penalties are similar to the baseline and shows the same stark differences across couple types when we restrict to couples that have no additional children until t+5 (but keep in mind that this is an endogenous sample restriction and therefore should be interpreted with care).



Figure B7: Mean number of kids over the observation window, by couple type



Figure B8: Only couples where no partner has additional kids until t+5

B.3 Child test scores

In the main results, we showed that same-sex couples share the burden of child rearing more evenly, and experience less severe household income penalties compared to heterosexual couples. It is natural to ask if this reduction in the child penalty comes at the cost of worse outcomes for children. In Table B1 we present results from a regression of test scores at age 10 for the children of heterosexual and same-sex couples on a dummy for having same-sex parents and an increasing set of control variables across columns. Standard errors are clustered by both parents using two-way clustering. The results in the first column, corresponding to no controls, indicate that at 10 years of age children of same-sex couples do much better than children of heterosexual couples, in the range of 0.4 to 0.6 standard deviations in the three subjects. Moving right, we gradually add more controls for observable pre-birth differences between heterosexual and same-sex couples. Children of same-sex couples still do around 0.2 standard deviations better in both reading and English even when controlling for our large range of observable characteristics. These results suggest that while same-sex parents appear to parent more equally and

experience smaller costs to overall household income, their alternative approach to child rearing does not come at the cost of child outcomes, and may even improve outcomes.⁵⁸

Outcome variable	(1)	(2)	(3)	(4)	(5)			
Math	0.395***	0.363***	0.283***	0.0893	0.0766			
	(0.0858)	(0.0853)	(0.0853)	(0.0835)	(0.0838)			
Reading	0.410^{***}	0.352***	0.263***	0.146^{*}	0.170^{**}			
	(0.0832)	(0.0833)	(0.0836)	(0.0821)	(0.0810)			
English	0.565***	0.529***	0.433***	0.248***	0.235***			
	(0.0800)	(0.0794)	(0.0803)	(0.0773)	(0.0777)			
Pre-birth controls								
Child gender	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark			
Birth year dummies	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark			
Age dummies		\checkmark	\checkmark	\checkmark	\checkmark			
Municipality dummies			\checkmark	\checkmark	\checkmark			
Education level dummies				\checkmark	\checkmark			
Income					\checkmark			
Number of children (min)	316,039	315,880	315,880	315,879	302,468			
- of same-sex female couples	134	134	134	134	133			
- of same-sex male couples	4	4	4	4	4			

Table B1: Impact on children: Test scores at age 10

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

B.4 Decomposing the child penalties

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 as in equation (11). Therefore, the estimates are interpretable as the effects of children

⁵⁸Although a further analysis of the relative performance of children from same-sex and heterosexual couples is beyond the scope of this paper, these results might also indicate stronger positive selection into child bearing among same-sex female couples that is not accounted for by our rich set of controls.

at age (in months) *j*, relative to the effect 12 months before birth.

Results are presented in Figures B9 and B10. We begin in Figure B9 by repeating the baseline estimates, but unlike in Figure 1 these are unscaled. As expected, the child penalties look largely the same as the baseline results with an immediate drop of around 100,000 NOK (approximately 11,600 USD) for mothers in heterosexual couples that persist over the period we investigate and a smaller and decreasing penalty for same-sex mothers. In panel (b) we plot effects on the extensive margin of having any active employment relation. Unlike the baseline outcome of labor earnings, we see a strong dip in employment around the time of child birth for mothers, driven by employment spells not being active when mothers are on leave in contrast to maternity leave benefits that replace earnings and are included in our income measure. Following the initial dip, employment bounces back but stays below -0.1 for the period under study, indicating 10 percentage points lower probability of being employed compared with the baseline employment rate 12 months before birth. In panel (c) we estimate impacts on a dummy indicating a full time job, as defined by contracted weekly hours above 30. The fact that the impact on this measure is larger than on the employment measure, at around a 20 percentage points reduction, indicates that there is response both on the extensive and intensive margins of labor force participation: some mothers drop out of the labor force entirely while others reduce labor supply and work part time following child birth. As before, we find little response among heterosexual fathers for these measures.

For same-sex 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 caused by parental leave directly, co-mothers behave similarly to their partners, reducing labor force participation by around 10 percentage points in response to child birth. For the full time measure, however, the reduction is markedly smaller for same-sex mothers than heterosexual mothers, indicating that part of the differences in income patterns are driven by more mothers working full time in same-sex than heterosexual couples following child birth. This difference is mirrored in the outcome for total hours on top of Figure B10, 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 same-sex mothers is smaller and fully recovers 4-5 years after birth. Same-sex co-mothers behave much like their partners after the first year of leave, while heterosexual fathers increase total contracted hours. Summing up, the differences in the child penalties between heterosexual and same-sex mothers seem to be driven by differences in the response on the intensive, not the extensive margin.

Following Kleven *et al.* (2019b), we also estimate the impact on two measures of workplace flexibility. The first is a dummy for whether the employer is in the public sector, which is known for its flexibility and well regulated working conditions. The second is a measure of family friendliness that we construct at the firm-month level. It represents the share of mothers of children below 15 years of age among the other workers who have their primary employment relation with the firm. Both of these measures, however, are defined only for employed people; since we have shown that employment is endogenous to child bearing, these should be interpreted with care. That caveat aside, the child penalties for these outcomes are plotted in panel (b) and (c) of Figure B10. 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 childbirth, whereas this trend is flat for men. The trend in this outcome is relatively similar for both partners in same-sex female

couples. Our measure of family friendliness suggests that all types of mothers move to more family friendly firms in the period up to and following birth. The fact that same-sex mothers do not experience long term child penalties, but are just as likely as heterosexual mothers to move into family friendly firms, suggests that occupational selection in response to children cannot fully explain the gender income gap post children.

Finally, we use a measure of days of sickness absence to see if childbirth may cause longer term health shocks that impact income. The measure counts the full-time equivalent days of absence due to sickness from physician-certified spells of leave that exceed 16 days, so will generally not include short term illness such as seasonal cold or flu. It also include sickness absence spells for dependents that require the employee to be absent, in particular young children. As with the measures of family friendliness, this measure is conditional on employment.⁵⁹ Results indicate an unsurprising spike in sickness absence for heterosexual and same-sex 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.⁶⁰ The pattern is relatively similar for both partners in same-sex female couples. Heterosexual fathers also take slightly more sickness absence after the birth of children than before.

⁵⁹Despite this, we occasionally see non-employed individuals in these data. We exclude the few non-employed individuals who are registered with absence spells.

⁶⁰Note that some of this could be caused by subsequent pregnancies.



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



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



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



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

Figure B9: Decomposition I: Child penalties for heterosexual (left) and same-sex female (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 B10: Decomposition II: Child penalties for heterosexual (left) and same-sex female (right) couples

B.5 Child penalties over time

Figure B11 shows that the child penalty for women has declined substantially over time. In the 1970's and 1980's, fathers experienced a child premium rather than a penalty. 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 child penalty has been substantial from the 1970's until today, the remaining gap is still large, and largely driven by the penalties experienced by mothers.





Note: Child penalties estimated separately by birth cohort of first child in 5-year intervals. Estimated using the event study framework from equations 10 and 11.

C Paternity leave

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

Unfortunately, the data does not contain direct links to the child or children for which the leave is taken, only to the individual who takes leave. We therefore have to infer the relevant child from the birthdates of the children. To this end, we assign a parent's leave spell to a particular child if it

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

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

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

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

C.1 Balancing tests

Table C1 provides sharp RD balancing tests for a range of covariates in the baseline RD model. Figure C1 provide robust local polynomial estimates of the density of births around the cutoff. Reduced form and first stage estimates separately by reform is plotted in Figure C3.
Variable	Reform year	2005	2006	2009	2011	2013	2014	Pooled	Stacked
Mather's	RD ectimate	0.030	-0.19	0.030	-0.080	0.088	-0.032	-0.07	-0 0043
age	CONV. S.e.	(0.23)	(0.25)	(0.066)	(0.13)	(0.13)	(0.072)	(0.067)	(0.042)
2	robust p	0.80	0.44	0.77	0.41	0.43	0.80	0.75	~
Father's	RD estimate	0.18	0.17	0.020	-0.20	0.19	0.0072	0.075	0.023
age	conv. s.e.	(0.25)	(0.26)	(0.069)	(0.16)	(0.16)	(0.080)	(0.079)	(0.046)
1	robust p	0.51	0.50	0.97	0.31	0.26	0.82	0.36	
Maternity	RD estimate	0.99^{*}	0.10	-0.33***	-0.14	-0.059	-0.67***	-0.028	-0.36***
leave	conv. s.e.	(0.49)	(0.49)	(0.11)	(0.23)	(0.20)	(0.11)	(0.12)	(0.069)
	robust p	0.046	0.96	0.013	0.64	0.98	0.000	0.97	
Father's	RD estimate	-0.0062	-0.32	0.0092	-0.050	0.017	-0.012	-0.055	-0.013
years of ed.	conv. s.e.	(0.20)	(0.20)	(0.047)	(0.074)	(0.091)	(0.054)	(0.046)	(0.029)
	robust p	0.75	0.12	0.92	0.60	0.84	0.94	0.20	
Mother's	RD estimate	0.055	-0.26	0.050	-0.080	0.098	0.040	-0.0084	0.034
years of ed.	conv. s.e.	(0.18)	(0.19)	(0.045)	(0.097)	(0.089)	(0.047)	(0.047)	(0.028)
	robust p	0.84	0.18	0.32	0.46	0.26	0.35	0.85	
Mother's	RD estimate	0.0018	0.0086	-0.0041	0.0024	-0.0086	-0.0040	-0.00042	-0.0034^{*}
ed. missing	conv. s.e.	(0.0080)	(0.0094)	(0.0028)	(0.0053)	(0.0066)	(0.0034)	(0.0027)	(0.0019)
I	robust p	0.82	0.40	0.20	0.61	0.17	0.39	0.91	
Father's	RD estimate	-0.019^{*}	-0.0081	-0.0037	-0.0014	-0.0046	-0.0052	-0.0071^{***}	-0.0046^{**}
ed. missing	conv. s.e.	(0.0094)	(0.0096)	(0.0026)	(0.0043)	(0.0063)	(0.0037)	(0.0027)	(0.0019)
	robust p	0.047	0.36	0.19	0.71	0.44	0.32	0.011	
semi parametric	: sharp RD estin	nates of the	effect of 1	baternity le	ave quotas	s on balanc	ing variable	s using optin	nal bandwidths,
cal linear polyn	omials on either	r side of th	e cutoff. A	ll estimate	s are scale	d to reflect	one week	of quota incr	ease. Pooled est

Table C1: Sharp RD balancing tests

weighted average of reform-specific estimates. Robust, but not bias-corrected standard errors reported, "robust p"-values are bias-corrected. * p < 0.1 ** p < 0.05 *** p < 0.01, based on the robust, but not bias-corrected standard errors (themselves not reported). triangular imates are kernel and loc Note: Robust



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.* (2018). *p*-values reported are for a bias-corrected test of whether the densities at the cutoffs are equal.

C.2 Additional results

Figure C2 reports the impacts on mothers' and

fathers' annual incomes over time using the the stacked and pooled fuzzy RD estimates. There is no effect of paternity leave use on pre-birth outcomes. This is a reassuring, and can be interpreted as an additional placebo test. Following birth, we see no impact of paternity leave use at years 0 and 1 on the labor income of mothers or fathers when most of the leave take-up happens. Nor do we see any impact in the following years; the estimates are flat and centered at zero. Using the stacked specification we can rule out positive impacts larger than around NOK 5,000 on mother's annual earnings in response to each week of paternity leave use for all years post-birth.



Figure C2: Fuzzy and stacked RD estimates of the effects of paternity leave use on mothers' and fathers' earnings.

Note: Left figure shows fuzzy RD estimates of the impact of an additional week of paternity leave use on mother's earnings over time, using all six reforms. Right figure shows fuzzy RD estimates of the impact of paternity leave use on father's earnings over time, where confidence intervals for the pooled specification has been capped at +20,000 NOK to maintain readability of the axis. Pooled estimate refers to the weighted average of reform-specific estimates, 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.





C.3 Accounting for effects of maternal leave

As evident from Table 1, several of the reforms affected not only the paternity leave quota, but also the maternity leave quota and the sum of the maternity leave quota and the shared leave. As documented in Table C1, this resulted in reduced maternity leave take-up roughly for the reforms where the total time a mother could take off work was reduced. Although we argue that this change in maternity leave takeup is relatively minor compared with the change in paternity leave, and at much higher margins, we might worry that it is partly the changed maternity leave that causes any changes in later labor market outcomes, not paternity leave.

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

$$y_{irt} = \beta_t^L L_i + \beta_t^M M_i + \varphi_r^0 x_i \mathbb{1}(x_i < 0) + \varphi_r^1 x_i \mathbb{1}(x_i \ge 0) + \pi_r + \epsilon_{irt}$$

$$L_{ir} = \gamma_{LQ} Q_{ir} + \gamma_{LS} S_{ir} + \varphi_r^{L0} x_i \mathbb{1}(x_i < 0) + \varphi_r^{L1} x_i \mathbb{1}(x_i \ge 0)] + \pi_r^L + \eta_{ir}^L$$

$$M_{ir} = \gamma_{MQ} Q_{ir} + \gamma_{MS} S_{ir} + \varphi_r^{M0} x_i \mathbb{1}(x_i < 0) + \varphi_r^{M1} x_i \mathbb{1}(x_i \ge 0)] + \pi_r^M + \eta_{ir}^M \quad (14)$$

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

When instrumenting for two endogenous variables in an IV-setup, it is not clear how to determine the optimal MSE-reducing bandwidth as before. We therefore use a) the MSE-reducing optimal bandwidth for the first stage of either of the instruments or b) a fixed 50-day bandwidth. As before, we report robust, but not bias-corrected standard errors for the stacked specification.

First stage results for the two endogenous variables are reported in Table C2. Notice that independent variation to identify both effects relies on stacking all reforms, so that we cannot perform these estimates separately by reform. The choice of bandwidth is not of essence: The results are very similar whether we use either the MSE-reducing optimal bandwidths or a fixed 50-day window. Second, note that the reforms work exactly as we would expect: An increase in the daddy quota of 1 week increases paternity leave uptake by almost exactly 1 week when we control for changes to the remaining quota for the mother. Increasing the remaining leave for the mother (comprised of the maternal quota and the weeks of shared leave) increases maternity leave take up by 0.7 to 0.8 weeks. In contrast, the instruments do not work across spouses: Weeks of paternity leave quota does not affect maternity leave use when controlling for the remaining share available to the mother, in contrast to the balancing exercise in Table C1, while the remaining share for the mother does not affect leave uptake for the father when controlling for his own quota. Thus, the stacked specification where we instrument for both parents'

leave take up circumvents the problem of the reforms affecting both margins of leave.

	Weeks of leave		Ba	Bandwidth				
	Mother	Father	reform	bw	N			
A: 50-day bandwidth								
Paternity leave quota	0.066	1.00***	2005	50	4,037			
(Q_{ir})	(0.14)	(0.16)	2006	50	4,303			
			2009	50	4,192			
Remaining leave for mother	0.77***	0.18	2011	50	3,656			
(S_{ir})	(0.21)	(0.21)	2013	50	4,830			
joint F	21.1	76.5	2014	50	3,502			
N			24,520					
B: Maternity leave-optimal bandwidth								
Paternity leave quota	0.069	0.96***	2005	66.9	5,418			
(Q_{ir})	(0.14)	(0.15)	2006	61.3	5,271			
			2009	42.4	4,017			
Remaining leave for mother	0.79***	0.12	2011	43.8	4,252			
(S_{ir})	(0.20)	(0.20)	2013	57.7	5,538			
joint F	24.8	79.7	2014	52.4	4,532			
Ň			29,028		<i>.</i>			
			,					
C: Paternity leave-optimal bandwidth								
Paternity leave quota	0.055	0.98***	2005	58.0	4,770			
(Q_{ir})	(0.14)	(0.15)	2006	68.4	5,844			
< /	· · /	``'	2009	44.6	4,192			
Remaining leave for mother	0.73***	0.12	2011	37.4	3,656			
(S_{ir})	(0.21)	(0.21)	2013	55.3	5,380			
joint F	18.0	74.9	2014	40.8	3,502			
Ň			27.344		,			

Table C2: First stage effects of maternity and paternity leave quotas

Note: First stage results from stacked specification of all six parental leave reforms, instrumenting for weeks of paternity and maternity leave take up as described in eq. 14. Panel A) uses a fixed 50-day bandwidth, panel B) uses the MSE-reducing optimal bandwidth for each reform if instrumenting for maternity leave only, panel C) the same for paternity leave. Heteroskedasticity robust, but not bias-corrected standard errors. '*p < 0.1, **p < 0.05. ***p < 0.01

Because the choice of bandwidth does not seem to matter and because we're primarily interested in the effects of paternity leave, we present fuzzy stacked RD estimates based on this specification using the paternity leave-optimal bandwidth from panel C. As in the base model in the paper we also revert to the reform-specific dummies as instruments when reporting the IV estimates rather than quota measures. Results from the stacked fuzzy RD model where we instrument for both mothers' and fathers' leave take up is presented in Figure C4. The top panel presents effects of paternal leave on mothers' and fathers' earnings by child age, mirroring the estimates from the baseline model. For reference, the coefficients and confidence intervals from the stacked fuzzy RD model where we instrumented for paternity leave use only is added. Except perhaps for the outlier at child age 4, the double IV model provides estimates that are well in line with the baseline model, confirming the precise zero effects of paternity leave on mothers' subsequent labor earnings. Just like in the basic model, it does not seem like paternity leave has a potential for reducing the child penalty.



Figure C4: Effects of maternity and paternity leave use s labor earnings

Note: Top panels shows the impact of a week of paternity leave use on mothers' (left) and fathers' (right) earnings over time, as estimated from a double IV stacked fuzzy RD as detailed in eq. 14. For comparison we also show our stacked fuzzy RD estimates from the baseline model where we only instrument for the weeks of paternity leave. Bottom panels show the impact of an additional week of maternity leave on mothers' (left) and fathers' (right) earnings.

The double IV specification inadvertently also estimates the effects of another week of maternity leave on parents' later earnings. Results are too imprecise to draw strong conclusions, but provide no evidence of any effects. In short, parental leave policies do not seem like a promising tool for reducing child penalties.

D Child care

Figure D7 shows the impact of high equality, subsidized early child care on each parent's earnings. Focusing first on the years of treatment, ages 1-3, we see that the estimates increase in this period up to point estimates of around 27,000 NOK at age 2 and close to 30,000 NOK at age 3, where most of the treatment happens, only to return to zero the last two years of the panel.⁶¹ Estimates are significant at the 5 percent level at age 3 and 10 percent level at age 2, and thus indicate that there is some immediate effects of use on earnings during the years of treatment, perhaps driven by allowing mothers to return to work earlier after child birth. Results for fathers are noisy, but point, if anything, to negative impacts on earnings, which could also reduce child penalties. The pre-birth outcomes, which we can think of as placebo outcomes, indicate small and insignificant impacts of future child care use on past earnings, supporting the estimation strategy.



(a) Child care coverage rates

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

Figure D5: Child Care Coverage *Source:* Statistics Norway Statistikkbanken, tables 09169 and 07459.

⁶¹This estimate is smaller than the baseline estimate in Andresen and Havnes (2019), but a number of differences in the sample and specification may explain this, as well as the lower level of precision in our study due to a sample size than half the size because of the focus on first born children only.



Figure D6: Predicting expansion of slots from pre-reform characteristics

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



(a) Mothers earnings, unscaled

(b) Impact of early child care use on fathers' earnings

Figure D7: Impact of early child care use on parents' earnings

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