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

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 and same-sex couples to better understand what causes the child penalty. 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. In contrast, we 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

†We thank seminar participants at the University of Rochester, Duke, Claremont McKenna University, Statistics Norway, RAND, Arizona State University, LSU, the University of Oslo, VATT Helsinki, Warwick University, Erasmus University, CSU Fullerton, UC Riverside, Purdue, SOLE, and EALE. We also thank Heather Antecol, Manuel Bagues, Sebastian Calonico, Matias D. Cattaneo, Nina Drange, James Fenske, 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: October 28, 2019. Latest version here

‡Corresponding Author: University of Southern California, enix@usc.edu
Keywords: Gender wage gap, labor supply, child penalty, paternity leave, child care, same-sex couples, event study, regression discontinuity, instrumental variables
1 Introduction

A growing number of papers demonstrate that parenthood causes substantial drops in labor market income for mothers, with little or no labor market income drops for fathers.\(^1\) This stylized fact, commonly referred to as the “child penalty”, is strikingly consistent across countries and appears for both high and low educated mothers.\(^2\) 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, Kleven et al. (2019b) show that in Denmark the child penalty accounted for 80% of the gender income gap in 2013, up from 40% in 1980.\(^3\) Not only is the child penalty pivotal to the gender income gap, it also has important macroeconomic consequences, as recently demonstrated in Doepke and Kindermann (2019).\(^4\) In this paper we provide a better understanding of what causes the child penalty and how it might be reduced.

There are a number of possible reasons for the child penalty. We focus on four common explanations. First, only women can give birth and in general only the woman who gives birth breastfeeds. Giving birth is a major health shock which

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

\(^{2}\)We show this result in the Appendix, and it was also demonstrated in the earlier NBER version of Kleven et al. (2019b).

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

\(^{4}\)The authors write a model of 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 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 (for example, see Feyrer et al. (2008)).
could have long term consequences for productivity and earnings. Breastfeeding 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, women often make less than their husbands. As a result, men may have a comparative advantage in market work relative to household work compared to women and households might efficiently specialize after birth. Third, women may get greater utility from spending time with children compared to 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, 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 to same-sex couples can be used to confirm or reject potential mechanisms. This motivates our first set of empirical results, estimating and comparing the child penalties of heterosexual and same-sex female couples. Second, if policy makers wish to reduce the child penalty, what policies might be most effective depends on which mechanisms are at work. This insight underpins our second set of empirical results where we estimate the causal impact of two commonly proposed policies that might be effective based on the mechanisms identified by the first set of empirical results: paternity leave and subsidized early child care.

We take these insights to the data in two parts. First, we estimate child penalties for heterosexual and same-sex female couples using the event study approach.

---

5 There are too few same-sex male couples with kids in Norway to get precise estimates. We discuss this in more detail in Section 4.
from Kleven et al. (2019b) and administrative data from Norway. We find that while there is no impact of parenthood on fathers’ earnings, women in heterosexual couples experience an average drop in income of 20% following the birth of the first child, and this drop persists over time. This large drop in female income translates to an overall household income drop of 6-8% for heterosexual households that persists over time. For same-sex female couples the patterns are dramatically different. We find that both women experience an income penalty after birth, but the woman who gave birth experiences a larger drop of 13% compared to her partner, whose income drops by 5%. Despite a larger immediate drop in income, the mother who gives birth catches up with her partner two years after birth. From that point 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 the household income penalty experienced by 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 four most common explanations for the child penalty in heterosexual couples. Our results demonstrate that the fact that only the woman can give birth to a child in heterosexual couples can only explain part of the child penalty, and only in the first two years after birth. We can also reject
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 gap experienced by heterosexual couples is due to gendered differences in norms and preferences for child care, although it is not possible to fully disentangle these two explanations.

While comparing the outcomes of children born to same-sex and heterosexual couples is not the focus of this paper, we also present descriptive evidence that same-sex partners sharing the burden of parenting 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. This result overcomes previous data shortcomings to contribute to a charged debate regarding the impact of same-sex parents on child outcomes.

Our results on mechanisms suggest that two commonly proposed policies aimed at reducing the child penalty, paternity leave and subsidized early childcare, could be effective. Paternity leave could increase the utility fathers get from spending time with children relative to mothers and could also change gender norms around

---

6Previous studies of children born to same-sex couples have been criticized by both sides of the debate on the basis of three methodological concerns: non-representative samples (studies often used “opportunity samples” where couples volunteer to participate), mislabeling children from heterosexual couples as children of homosexual couples (i.e. children with a parent who divorced and later married a same-sex partner) or vice versa, and small sample size. In this paper, our use of administrative data containing the population of children of same-sex couples in Norway and the ability to identify such children accurately largely overcomes these concerns.

7For example, see oral arguments for the landmark 2015 Supreme Court case Obergefell v. Hodges, which legalized same-sex marriage in the United States.
child care. Subsidized early childcare, 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 last part of the paper we estimate the 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 2005-2014. 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 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 1-2 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)]. Results indicate positive effects on mothers’ labor income at ages 2 and 3 that scales to reduce the child penalty experienced by mothers by around 25% for each additional full year of early child care use,
although the impacts are not persistent in the long run.

Our paper is most closely related to the literature on child penalties. We use the simple event study approach from Chung et al. (2017), Kleven et al. (2019b), Bergsvik et al. (2019), and Angelov et al. (2016) to identify child penalties. Lundborg et al. (2017) also show the child penalty gap occurs among heterosexual couples who use IVF to get pregnant, using quasi random variation in fertility after IVF treatment. 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 whose role in explaining remaining gender income gaps cannot be understated. We make two additional contributions to this literature. First, we use a household model combined with a comparison to estimated child penalties for same-sex female couples to better understand the mechanisms behind the child penalty within heterosexual couples. In this regard our paper contributes to a smaller literature in economics focused on same-sex couples. Most closely related are Moberg (2016) and Rosenbaum (2019) who also estimate the response to child birth for same-sex couples in Sweden and Denmark, respectively. In addition, our findings regarding mechanisms are consistent with Kleven et al. (2019a) who show that the magnitude of the child penalties experi-

---

8 A more recent working paper (Bensnes et al., 2019) finds more short-lived child penalties when accounting for the impact of additional kids using multiple IVF treatments.
9 Additional papers in this literature are Baumle (2009), Schneebaum (2013), and Black et al. (2007) who compare earnings between heterosexual and same-sex couples. Regarding parenting, Goldberg et al. (2012) look at a sample of 55 lesbian couples and find they report sharing household chores and child care more evenly than a comparison group of 65 heterosexual parents. Others have investigated labor supply (Antecol and Steinberger, 2013), parental leave use (Evertsson and Boye, 2018; Rudlende and Lima, 2018) and time use (Martell and Roncolato, 2016) for same-sex couples, as well as the impact of legal recognition (Alden et al., 2015).
enced by women are correlated with elicited gender norms across countries.\footnote{Although we do not consider workplace discrimination against mothers, Gallen (2018) shows that the drop in women’s earnings at work corresponds to a drop in productivity at work, which suggests that workplace discrimination is not an important explanation for the child penalty. While we lack data on hourly wages to do the same comparison, our finding that the largest difference between the couple types is the number of hours worked is consistent with the findings in Gallen (2018).}

Second, 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 \textit{et al.} (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 \textit{et al.} (2018) are not unique to academia. Increasing paternity leave does not reduce the child penalty across the population of professions in Norway. These results also tie in to a larger literature examining the impacts of paternity leave on parent’s earnings and labor supply, finding mixed results.\footnote{For 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 \textit{et al.} (2019) finds that a Danish increase in 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, and Ekberg \textit{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.}

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 \textit{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 posi-
tive impacts of paternity leave. Patnaik (2019) finds a large change in the division of household labor from a Canadian daddy quota and Persson and Rossin-Slater (2019) find that when fathers have the flexibility to stay home, there are positive impacts on the mother’s health. Regarding child care, a growing literature summarized in e.g. Blau and Currie, 2006; Akgunduz and Plantenga, 2018; Morrissey, 2016, contains a large 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.

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 penalties, our main outcome of interest throughout the paper. We summarize the data and present summary statistics in Section 4. In Section 5 and 6, we present the main results. Section 7 concludes.

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

In this section we derive a model of household labor supply in the presence of children, incorporating 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 norms around child care. The model is deliberately stylized to bring out the implications of the four mechanisms, and how
they may differ for heterosexual versus same-sex female couples, in a sharp way. 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).

In each household there are two adults. The two adults may either be a man and a woman, or two women. While couples with two men are also of interest, we have too few with children in our data to get precise empirical estimates. 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 to a child. Thereafter, the household consists of the two adults and the child and 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:

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

where $c_t$ is consumption, $\theta_t$ is child quality, and $\beta$ 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

---

12We also omit single mothers and single fathers, for whom the household model would be very different.

13We do not model the fertility decision or allow parents to make labor market decisions in anticipation of children. While these are important issues (see for example Bursztyn et al. (2017) 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.

14We rule out adoption to simplify the model as it is rare in Norway, although in a robustness exercise in the appendix we look separately at adoptive parents.
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 gender norms through the term \( \alpha h_{-i,t} \bar{Z}_i \), as in Fernández et al. (2004). \( \alpha \) is the disutility men get from each hour their female partner works when they have children, and \( \bar{Z}_{i,t} \) is an indicator equal to 1 if the individual is a male partnered with a female in periods 2 and onward.\(^{15}\) The term \( \eta_{i,t} \ln (1 - h_{i,t}) \) captures the fact that women may get greater utility from time with children than men, under the assumption that if individual \( i \) is female, \( \eta_{i,t} = \bar{\eta}_t \) which in every period is larger than the equivalent for men, \( \bar{\eta}_t \), so that \( \bar{\eta}_t > \eta_t \ \forall t. \)\(^{16}\)

Child quality is produced by the following production function that takes as inputs each parent’s time, as well as child care purchased on the market, denoted \( h_{m,t}. \)

\[
\theta_t = k_i \psi (1 - h_{i,t}) + k_{-i} \psi (1 - h_{-i,t}) + k_m \psi (h_{m,t})
\]

(2)

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

There is no saving or borrowing, and in each period household consumption

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

\(^{16}\)Note, one might suspect that the preferences for children (and time spent with them) is stronger among same-sex couples, because the procedure for most of them will involve greater costs. If this is a concern, a natural comparison group for same-sex female couples getting children is heterosexual couples getting children through IVF. This is not necessarily the case, however, because heterosexual couples doing IVF have fertility problems, while same-sex female couples do not necessarily have any fertility problems. A natural hypothesis is that heterosexual couples doing IVF have even stronger preferences for children than same-sex female couples. Moreover, while we lack data to do such a comparison, as discussed in the introduction existing studies using IVF still find large relative child penalties for heterosexual couples.
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

\[
c_t = w_i \left( 1 - \delta_t \bar{S}_i \right) h_{i,t} + w_{-i} \left( 1 - \delta_t \bar{S}_{-i} \right) h_{-i,t} - ph_{m,t}
\]

where \( \bar{S}_i \) is an indicator equal to 1 if individual \( i \) is a woman who gave birth. \( \delta_t \) is the time varying productivity shock of giving birth. Note that \( \delta_t \) 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 other actions associated with giving birth might promote longer term attachment with 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 $\lambda_i$ is the weight of each spouse in household decisions. This assumes Pareto efficiency in household decisions and is consistent with a number of household bargaining problems.\(^{17}\)

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:

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

$$+ \lambda_i \eta_i \ln (1 - h_i) + \lambda_{-i} \eta_{-i} \ln (1 - h_{-i}) - \lambda_i \alpha h_i \tilde{Z}_i - \lambda_{-i} \alpha h_{-i} \tilde{Z}_{-i}$$

(4)

(5)

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

---

\(^{17}\)This is a very simple model by design. It assumes Pareto efficiency, but this has some important drawbacks. See [Del Boca and Flinn (2012)] for a discussion of alternative approaches. Notice that we assume that the bargaining weights do not vary by couple type. An alternative approach to capture gender norms could be to assume that in same-sex couples $\lambda_a = \lambda_b$ and in heterosexual couples $\lambda_a > \lambda_b$, where $\lambda_a$ represents the Pareto weight of the man.
These wage equations specify how the labor supply of heterosexual fathers, heterosexual mothers, same-sex mothers, and same-sex co-mothers will be impacted by children under each mechanism. For example, suppose the importance of gender norms increases (i.e. $\alpha$ increases). This will increase the right hand side of equation 7. Given that the left hand side $(1 - \delta) w_{mother}$ has not changed, in order for the two sides of the equation to remain equal, $h_{mother}$ must decrease. But if $h_{mother}$ has decreased, then $h_{father}$ must increase, so that equation 6 remains equal. An increase in $\alpha$ has no impact on equations 8 and 9. Thus, an increase in the importance of gender norms will cause the mother’s labor supply to decrease and the father’s labor supply to increase in heterosexual couples, with no impact on same-sex mothers’ labor supplies. More generally, these wage equations yield the following insights:

1. **Costs of giving birth:** For each period where $\delta > 0$, there will be a child penalty for both same-sex female and heterosexual 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. Note that because we allow $\delta$ 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 same-sex female 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 same-sex female couples.

3. **Gendered differences in preferences for child care**: As the gendered differences in preferences for child care grow larger ($\bar{\eta} - \eta$ increases), mothers in heterosexual couples will decrease their labor supply while fathers will increase their labor supply. However, all else equal, if $\bar{\eta}$ increases then the labor supply of heterosexual mothers will decrease by a larger amount than the labor supply of same-sex mothers.

4. **Gender norms**: As described above, gender norms will cause a child penalty for heterosexual couples, but not for same-sex female couples.

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 couples allows us to distinguish between mechanisms using these predictions from the model.
Perhaps the most surprising result that comes out of the model is the fact that
gendered differences in preferences will lead to smaller drops in income after birth
for same-sex mothers compared to heterosexual mothers. The intuition is that in
heterosexual couples, the husband will increase labor supply to the market in order
to compensate for lost income from the mother, while in same-sex female couples
both spouses will have to balance their mutual desires to spend more time at home
with the need to maintain consumption by providing labor to the market.

The model also formalizes how paternity leave and subsidized early childcare
might reduce the child penalty. Specifically, equation 10 shows that the introduc-
tion 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, then it will also cause an increase in the labor supplied to the market
for the mother and/or father. Paternity leave might impact the child penalty for
heterosexual couples by increasing $\eta$ or possibly even decreasing $\alpha$, if the policy
causes fathers to become more comfortable with their partners working while they
are on paternity leave.

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. How-
ever, 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 discontinu-
ity 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).

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 penalties

$$y_{it} = \sum_{j \neq 1} \sum_{k} \alpha_{jk} \mathbb{1}[t = j, K_i = k] + \sum_{l} \sum_{m} \beta_{lm} \mathbb{1}[age_{it} = l, X_i = m]$$

$$+ \sum_{n} \sum_{o} \gamma_{no} \mathbb{1}[T_{it} = n, X_i = o] + \sum_{p} \eta_{p} \mathbb{1}[K_i = p] + \epsilon_{it}$$

Where $X_i$ is the gender (male, female) of parent $i$, $age_{it}$ is the age of parent $i$ at

For earlier examples see e.g. Jacobson et al. (1993), McCrary (2007).
event time $t$, $T_{it}$ is the calendar year for individual $i$ at event time $t$, and $K_i$ is the parent type: mother or father in heterosexual couple, and mother or co-mother in a same-sex couple. $\mathbb{I}[A]$ is the indicator function for event $A$. Standard errors are clustered by couple and robust to heteroskedasticity. The event time dummy the year before birth is omitted, which implies that all estimates of event dummies are relative to the year before birth for that specific parent type. Notice that all parents in our sample eventually have children, so 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 at identifying both short and long run child penalties compared to alternative approaches such as using instruments for the timing of birth.

Our objects of interest are $\alpha_{jk}$, the change in the outcome for a parent of type $k$ at child age $j$ relative to the earnings the year before birth. Ideally, we would use a log-linear specification of equation (11) so that we could interpret the coefficients as percentage changes in earnings, but the presence of zeros in the outcome complicates matters. To convert these absolute estimates to percentage child penalties, we follow Kleven et al. (2019b) and construct the following measure of the child

---

19 Note that while we allow life-cycle and time trends to vary by gender, we do not allow them to differ within gender. This means that the effect of age and year on income is the same for all women, be they in heterosexual or same-sex female couples.

20 While it is possible to estimate equation (11) separately for heterosexual mothers and fathers and same sex 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 the same-sex couples as well as heterosexual couples. In the 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 appendix give the same qualitative results as in our main outcomes.

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

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

(12)

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 couples and mothers and co-mothers in same-sex couples. These results are interesting on their own. Moreover, no additional assumptions are required to identify a potential upper bound on the impact of giving birth on labor market income by comparing mothers who give birth in same-sex and heterosexual couples. However, we require stronger assumptions to rule out or confirm other potential mechanisms behind the child penalty for heterosexual couples. Specifically, our model highlights differences in relative productivities as one such mechanism and predicts that we can rule out specialization based on comparative advantage 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 (11) if we wish to compare couples with similar relative productivities. Thus, to control for specialization, we flexibly control for the differences in own and spouse’s earnings.
prior to birth interacted with event dummies, by adding \( \sum_j \theta_j \mathbb{1}[t = j](y_i - y_{-i}) \) to equation (11), with income differences measured at the start of our panel, 4 years prior to birth\(^{22}\). To the extent that comparative advantage is captured by the relative income levels of the two spouses, these flexible event dummy controls will pick it up and we can attribute the remaining child penalties from \( \alpha_{jk} \) to the other possible mechanisms highlighted by the model\(^{23}\). Notice that these controls capture more than the intended comparative advantage. In particular, they also capture the autocorrelation in earnings over time. When presenting these results, we scale by the predicted earnings from the baseline estimates in equation (11), 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.

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

---

\(^{22}\) As an alternative, we also control for the differences in years of education before birth interacted with event time dummies, another measure related to labor market productivity and report these results in the Appendix.

\(^{23}\) Controlling for income gaps before birth controls for \( \frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}} \) if \( k_i = k_{-i} \), or if the following condition holds. Household production occurs both before and after the child arrives, it is only child quality that appears after birth in the context of our model. In that case, specialization will occur before the child arrives and will be captured by pre-market income gaps. Provided the household productivity parameters are unchanged or linearly related over time, then \( y_{it} - y_{-it} \) controls for \( \frac{w_i}{k_i} - \frac{w_{-i}}{k_{-i}} \).
sickness and parental leave and benefits. In subsection 5.2 we also discuss additional labor market outcomes related to employment spells from the FD-Trygd database. From these spells, we construct the following measures of monthly labor supply: Dummies for the employment spell exceeding 4, 20 and 30 contracted hours per week, whether the primary employment is in the public sector (2003 - 2014 only) and a proxy measure of the family friendliness of the firm. 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 Norway was the second country in the world to legally recognize same-sex partnerships in 1993 through the Partnership Act, so we have a longer panel of same-sex couples compared to most countries. However, there were restrictions regarding children until 2002, when same-sex couples could legally adopt the children of their partners. Thus, we take extra caution in the data to identify children of same-sex couples and heterosexual couples.

---

24 We set negative incomes to 0, comprising less than 0.2% of the observations.
25 For parental leave and sickness absence spells we pull data from FD Trygd, the register of the Norwegian Public Insurance system. For sickness absence, we measure the number of sickness days due to physician-certified spells of leave that exceed 16 days in a given month, scaled by the grade in the case of graded sickness absence to measure efficient days lost. For parental leave spells we measure how many weeks of leave were taken for a particular child. Details on this measure is provided in Appendix A.
26 The database covers most important employment spells from 1992 - 2003 and all employment spells (excluding self-employment) from 2003 - 2014. To create comparable measures across most of the sample period, we exclude spells of self-employment from the pre-2003 data and include only the employment spell with the most contracted hours for the post-2003 data. In more than 95% of the cases, the spell considered most important in the pre-2003 data is the one with the longest contracted hours.
27 Family friendliness is the leave-out-mean of mothers with children below 15 years that work in the firm.
28 Couples were not eligible for domestic adoptions, were not eligible for publicly subsidized assisted fertility treatment, and the registered spouse of a woman giving birth was not automatically registered as the second parent (as the *pater est* principle established for married heterosexual couples).
ples and when comparing child penalties for same-sex and heterosexual couples, restrict the analysis to children born after 2001. We describe these steps in more detail in Appendix Section A.

We keep only first-born children to both parents. In 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 in both samples only couples where the first child appears at ages 22 to 60 for both parents, giving us some time before and after birth to observe earnings. This leaves us with a sample of 250,296 heterosexual couples, 634 same-sex female couples and 32 same-sex male couples. The number of same-sex male couples with children is unfortunately too small to yield precise estimates, so we focus only on heterosexual and same-sex female couples.

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, to investigate labor market responses to the child’s birth. 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 Appendix Table A6 columns 2 and 3.

Same-sex female couples are slightly older than heterosexual couples at first birth, and are also slightly more educated. Reflecting the rules on establishing legal co-parent status (see Appendix Section A), the age at adoption is slightly delayed for same-sex female couples compared to heterosexual couples, as it takes some time for the co-mother to be legally registered. We find that same-sex female mothers have higher pre-birth labor earnings relative to heterosexual mothers, and the

---

29 The lack of same-sex male couples with children is consistent with the difficult same-sex male couples face when trying to have children. See Section A.

21
gap between mothers is smaller than the gap within heterosexual couples. While some of this is likely driven by older age at first birth, this result 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.

In Section 6, we estimate the impact of paternity leave and child care availability on the child penalty. Following birth, Norwegian parents have been entitled to a generous paid parental leave since 1977. Total parental leave is currently 49 weeks at 100% replacement or 59 weeks at 80% replacement rate, but the length of leave has been steadily increased since the mid 1980’s, reforms that we exploit and describe in more detail in Section 6.1. Benefits are capped at around 600,000 NOK or 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 among the parents.

We additionally exploit data on child care use and availability. Following parental leave, Norway has a well developed, regulated, and highly subsidized child care sector, as documented in 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 child care slots, we use administrative data from the child care sector.

---

30 Same-sex couples where one of the partners was legally step-parent adopting the child was granted this right in 2007.

31 In order to qualify for leave, a parent must have been employed for at least 6 of the 10 months prior to birth, and the annual earnings must exceed a low threshold of around 50,000 NOK or 6,000 USD. 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.
centers on the number of slots for children of different ages by December 15th each year. At the individual level, however, we can measure the exact use of child care at ages 13 to 35 or 36 months, depending on cohort. For these ages, a cash for care benefit was given to children who did not attend formal care in a given month. If we assume that all children who do not use child care apply for the benefit, which is relatively generous, we know exactly which children attended how much care for each month. From these data, we construct precise measures of full-time equivalent years of child care use from ages 13 to 35 or 36 months.

5 Heterosexual and same-sex child penalties

In Figure 1 we present the main results. The graphs report estimates of $C_{jk}$ (see equation (12)) generated by the simple event study in equation (11). The results for heterosexual couples are shown on the left and same-sex female couples on the right. As has been shown in many other papers, we find that mothers in heterosexual couples experience large income penalties in the range of 20% of their counterfactual earnings in the absence of children upon the birth of their first child. Fathers experience no income penalty upon the birth of the first child.

The graph for same-sex female couples is strikingly different. We find that both mothers experience a drop in income the year after the child is born, but initially the woman who gives birth has a larger drop in income. These drops in income, however, are much smaller than that of heterosexual mothers, at around 13% and 5% of counterfactual earnings for mothers and co-mothers, respectively. Moreover,

---

32 Throughout 2001-2008, which is the period we exploit, the benefit was around 3,500 NOK or 420 USD per month.
33 In Appendix Figure B1 we also report the raw mean earnings by event time for each couple type, without imposing any of the structure from equation (11), and the patterns are the same.
2 years after birth the woman who gives birth catches up and her penalty is no longer statistically significantly different from her partner’s. By five years after birth, the income penalty for both women has largely disappeared.

Figure 1: Estimated child penalties across couples types

Note: Figures in the top panel show the estimated child penalties from equation 11, scaled as described in eq. 12. Sample construction and data as defined in section 4. Figures in the bottom 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 income gaps within couple. Bootstrapped 95% 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
impose additional restrictions 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 to heterosexual couples.

The child penalty experienced by women in heterosexual couples is so large, it would seem to imply an overall household income penalty. In Figure 2 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 same-sex female and heterosexual couples experience statistically indistinguishable initial income declines on the birth of the first child. However, this drop in income persists for heterosexual couples while it decreases over time for same-sex female couples.

\[34\] We have also estimated the child penalties using the gap in pre-birth education instead of the pre-birth income gaps, another measure of labor market productivity, see Appendix Figure B3. The results are almost identical.

\[35\] In the appendix we also graph the child penalties for heterosexual couples where the woman makes more than the man pre-birth. In these cases, comparative advantage should cause the man to specialize in child care and have a larger penalty. This is not what we find in Figure B4, the woman continues to experience a large and similarly sized drop in income. 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 thousand NOK the year before birth, this large premium for fathers disappear, while the child penalty form mothers remain.
Based on these results a number of conclusions can be drawn. First, and perhaps most obvious, the striking difference in response to child birth 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. Thus, the fact that only the woman can give birth in a heterosexual couples cannot, on its own, explain the child penalty, since the mother who gives birth in same-sex female couples experiences a smaller income penalty after birth that is only significantly different than her partner’s income penalty in the year of and the year after birth. Moreover, by five years after birth the mother who gave birth in same-sex female couples no longer experiences an income penalty. 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 the income differences between couples with event dummies and removing that effect from the estimates, as we do in the bottom panel of Figure 1 the strikingly different patterns remain\[36\]. This suggests that the comparative advantage mechanism cannot explain the child penalty. Thus, we are able to reject two out of the four most commonly suggested explanations for the child penalty in heterosexual couples. However, the conclusions regarding female preferences and gender norms are less clear cut. The fact that the same-sex co-mother also takes an income penalty in the first year after birth suggests that she values time with the child, consistent with a female preferences for child care explanation\[38\]. The fact that over time the child penalties experienced by the same-sex female mothers decrease suggest that gender norms may also play a role, as the long term penalty cannot be explained by strong female preferences for spending time with children alone.

5.1 Robustness checks

In the Appendix, we report results from a number of robustness checks. First, one might be concerned that the differences between heterosexual and same-sex female couples are explained by differential fertility. In Appendix Figure B7 we show that fertility patterns for same-sex female couples are almost identical to

\[36\] There are some differences in the impact before birth that is likely caused by autocorrelation of incomes over time for the income differences specification.

\[37\] We get the same results when we control for comparative advantage using education differences within the couple pre-birth, see Appendix Figure B3.

\[38\] Note that an alternative formulation of the model might assume that this preference \(\eta\) is larger for the mother who gives birth within a same-sex female couple than the mother who does not, given that which mother gives birth is endogenous in same-sex female couples. However, if this is the case then we would expect to see a persistent gap between same-sex mothers in later years, in contrast to the catch up that we find.
heterosexual couples. Second, same-sex female couples may switch who gives birth over time, and this could explain the catch up experienced by the woman who initially gives birth. In 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 same-sex female couples, so we should be careful in 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. In the summary statistics Table A6 in the Appendix, we show that both same-sex mothers take more time for sick leave before birth, and the sickness leave they take is not significantly different from each other, which suggests this is not an issue.

Another concern is that there are other differences between couple types that explain the differences in child penalties, which are not captured by the model driven controls for comparative advantage above. To address this concern in Figure B9 in the Appendix we present child penalties for heterosexual and same-sex female parents using both a propensity score matching and also a nearest neighbor matching exercise. Details on these procedures are provided in Appendix B. We match on a variety of pre-birth characteristics such as municipality of residency, both spouses 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 sample of same-sex female couples and matched heterosexual samples. Although precision is lower in the sample of same-sex female couples, the results are similar to the baseline estimates for both exercises.
5.2 Decomposing the child penalties

To further understand the anatomy of the child penalty and what same-sex female 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.3. We find that same-sex mothers are equally likely to switch to flexible careers compared to heterosexual mothers, but do not have long term income penalties from children. This suggests that occupational flexibility alone cannot explain the large and sustained income penalties from children experienced by heterosexual women. The most important difference in terms of choices made by same-sex and heterosexual mothers is that heterosexual mothers experience more sustained drops in labor market hours. To summarize the main take away from these results, the differences in the child penalties between 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 market or switching occupations).

5.3 Child test scores

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 Table 1 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 children of same-sex couples do much better than children of heterosexual couples, in the range of 0.4 to 0.6 standard deviations in the three subjects. Moving right, we gradually add more controls for observable pre-birth differences between same-sex couples and heterosexual couples. 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.\textsuperscript{39}

\textsuperscript{39}Although a further analysis of the relative performance of children from same-sex and heterosexual couples is beyond the scope of this paper, these results might also indicate stronger positive selection into child bearing among same-sex female couples that is not accounted for by our rich set of controls.
Table 1: Impact on children: Test scores at age 10

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.395***</td>
<td>0.363***</td>
<td>0.283***</td>
<td>0.0893</td>
<td>0.0766</td>
</tr>
<tr>
<td></td>
<td>(0.0858)</td>
<td>(0.0853)</td>
<td>(0.0853)</td>
<td>(0.0835)</td>
<td>(0.0838)</td>
</tr>
<tr>
<td>Reading</td>
<td>0.410***</td>
<td>0.352***</td>
<td>0.263***</td>
<td>0.146*</td>
<td>0.170**</td>
</tr>
<tr>
<td></td>
<td>(0.0832)</td>
<td>(0.0833)</td>
<td>(0.0836)</td>
<td>(0.0821)</td>
<td>(0.0810)</td>
</tr>
<tr>
<td>English</td>
<td>0.565***</td>
<td>0.529***</td>
<td>0.433***</td>
<td>0.248***</td>
<td>0.235***</td>
</tr>
<tr>
<td></td>
<td>(0.0800)</td>
<td>(0.0794)</td>
<td>(0.0803)</td>
<td>(0.0773)</td>
<td>(0.0777)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pre-birth controls</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Child gender</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Birth year dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Age dummies (mother × father)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Municipality dummies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Education level dummies (mother × father)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Income (mother, father, interacted)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Observations (min. over course type) 316,039 315,880 315,880 315,879 302,468
Children of same-sex female couples 134 134 134 134 133
Children of same-sex male couples 4 4 4 4 4

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

6 The impact of family friendly policies on the child penalty

Despite the persistence of the child penalty within heterosexual couples, Appendix Figure B12 suggests that decreases in the child penalty are possible, as we find large decreases in the child penalty from 1971-2010. In the last part of the paper we explore the causal impacts of two policies that occurred during this period, paternity leave and early subsidized childcare, that could theoretical be effective at reducing the child penalty, given the mechanisms identified in the previous section.
6.1 Paternity leave

As means for 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. 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 (increasing fathers’ $\eta_i$ in equation 1) and might also decrease the distaste fathers have for mothers working outside the home (reducing $\alpha$ in equation 2). 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, paternity leave could theoretically reduce the child penalty, but only if it actually changes these parameters in practice. In this section we test that hypothesis.

40In 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), see OECD (2014). A number of firms in the United States also offer paternity leave.

41Paternity 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.
### Table 2: Parental leave reforms in Norway

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

Parental leave benefits and quotas in Norway in weeks. In parenthesis is the total leave including additional weeks compensated at 80% compared to the main leave amounts that are compensated at 100%. Source: [NOU 2017:6 (2017)].

In Table 2 we report every leave reform in Norway from 1992 - 2014. The maternal and paternal quota columns report the amount of parental leave in weeks that is reserved exclusively for the mother and father. The remaining leave can be shared among parents however they choose and is reported in column 6. The reforms were generally announced in October the year before implementation as part of the budgeting process, making it nearly impossible to plan conception in response to the announcement of the quota change in order to manipulate birth dates around the cutoff in April or July. In Figure C1 in the Appendix, we verify that there is no statistically significant change in the density of births around the cutoff for each reform.

In this paper, we exploit the 2005, 2006, 2009, 2011, 2013 and 2014 reforms using a regression discontinuity design. Identification relies on continuity in the underlying regression functions at the cutoff. Our identification strategy exploits

---

\[42\] We exclude the 1992 reform because it requires a donut-RD framework to identify the effects, which is not necessary for the other results.
the fact that parents of children born just before the reforms were not subject to the changes in parental leave quotas, while parents of children born right after each reform were subject to the changes. For this exercise, we draw on heterosexual couples with first children born in each reform year.\footnote{Because we want to capture mothers and fathers exposed to the leave reforms, we include in the sample only couples where the mother took some leave, indicating that she is eligible, because users of the alternative one-time benefit would not be affected.} We set leave to zero for fathers where we observe no leave take-up.\footnote{Appendix A provides additional details on the construction of our parental leave measure.} We rely on the following fuzzy RD setup separately for both mothers’ and fathers’ earnings measured at each event time $t$ relative to child birth:

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

$$L_i = \gamma_1(x_i \geq 0) + g(x_i) + \eta_{it}$$

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 \cite{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.\footnote{Many models in this section are estimated using the robust RD commands for Stata written by Matias D. Cattaneo and coauthors, whom we owe thanks. These include rdrobust, rddensity, rdbwselect and others. These packages are documented in Calonico et al. \cite{Calonico et al. 2018} and Cattaneo et al. \cite{Cattaneo et al. 2018}.} 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 Table C1 in
the Appendix.  

Figure 3: Fuzzy RD first stage estimate

Note: This figure graphically depicts first stage estimates around each reform date, using
local linear polynomials, triangular weights and optimal bandwidths.

An important imbalance revealed in this table is maternity leave takeout, as some of the
reforms 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 takeout
from already high levels to be driving our results. In Appendix C.3, we exploit the fact that some of
these reforms expanded the paternity leave use at the expense of the maternity leave, while others
expanded the total leave length. This allows us to instrument for both the maternity and paternity
leave use, confirming the baseline results of the effects of paternity leave.
Table 3: RDD first stage estimates

<table>
<thead>
<tr>
<th>Reform year Pooled Stacked</th>
<th>Pooled</th>
<th>Stacked</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD estimate per week</td>
<td>1.26***</td>
<td>0.88***</td>
</tr>
<tr>
<td>conventional standard error</td>
<td>(0.13)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>robust standard error</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>conventional p-value</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>robust p-value</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>93,508</td>
<td>93,508</td>
</tr>
<tr>
<td>Optimal bandwidth</td>
<td>35.3</td>
<td></td>
</tr>
<tr>
<td>Efficient observations</td>
<td>19,751</td>
<td>31,302</td>
</tr>
</tbody>
</table>

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

We see clear effects of all reforms on the take up of paternity leave in Figure 3. In Table C2 in the Appendix we show that these first stage estimates are always significant, whether using robust bias-correcting inference or conventional inference that only accounts for heteroskedasticity 47. In order to increase precision, we next combine the six reforms 48. 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. We call this the pooled estimate, and report the first stage specification for this procedure in Table 3. This estimate, however, restricts the functional form of the local linear polynomials to be the same for all cutoffs, potentially increasing the approximation error and lowering precision. An alternative and more straightforward

47Note that all results have been scaled to reflect one week of quota expansion. In order to increase precision, we next combine the six reforms.
48The second stage for all reforms are represented in Appendix Figure C3 and the same conclusions hold.
ward way to stack the estimates is to allow the local polynomials of the running variable to vary by cutoff and use the cutoff-specific optimal bandwidths and kernel weights from the individual specifications. The results are scaled to reflect one week of quota expansion. Scaling is secured by using an indicator of the number of weeks of quota increase rather than a dummy at the cutoff. Unfortunately, we cannot calculate bias-corrected standard errors for this specification, but we argue that the problem should be relatively minor. Our preferred first stage estimate from the stacked specification indicates that granting fathers another week of paternity leave quota increases leave takeout by .9 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 exactly reproduces the cutoff-specific first stage estimates for each reform reported in Table 3 and so is a natural way to stack the reforms. When interpreting these fuzzy RD estimates, it is important to keep in mind that these estimates are local average treatment effects: they capture the effects of additional leave use on earnings for people induced to use more leave because they were exposed to the 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 2). In case of heterogeneous treatment effects, the average effect for the compliers need not be the same as the average effect in the population. Despite this, we argue that the LATE is

---

First, notice that the difference between the conventional and the robust standard error estimate for the pooled specification is very small, indicating that the variance coming from the approximation error is relatively minor. Second, the approximation error should be smaller for the stacked than the pooled specification because we allow the local polynomials to differ between cutoffs and thus approximate the unknown functions better. Nonetheless, inference from this specification is only correct if the model is well specified, so that approximation error vanishes asymptotically.
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.

Figure 4: Fuzzy RD estimates of paternity leave use on the child penalty

Note: Figure shows fuzzy RD estimates of the impact of an additional week of paternity leave use on the child penalty, using all six reforms. Pooled estimate refers to the simple reentered estimate, while the stacked estimate stacks the cutoff-specific specifications. Robust bias-correcting inference reported for the pooled estimate, conventional, heteroskedasticity-robust inference for the stacked estimate.

Figure 4 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 of the child penalty, as estimated from the event studies from the first half of the paper. 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 con-

---

Effects on mothers’ and fathers’ annual incomes, reported in Appendix Figure C2 show no estimate on mothers’ or fathers’ incomes, suggesting that the main result is driven by lack of impact overall, rather than impacts of fathers and mothers canceling each other out.
Confidence intervals rules out reductions larger than around 5 - 7% of the female child penalty per week of paternity leave use for ages 1 through 5. While these results suggest paternity leave does not substantially reduce the child penalty, paternity 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 of spending time with 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 takeout for subsequent children for all children born up until and including 2014 in a setup similar to the peer effects estimates from Dahl et al. (2014). We cluster standard errors on the father to account for the fact that each father may have multiple treated kids and may get multiple subsequent kids for which we can measure outcomes. Notice that we cannot use the 2014 reform for this exercise, as we cannot reliably measure paternity leave use for kids born after 2014.

Notice that if fertility is endogenous to the parental leave reforms, this might constitute an endogenous sample selection criteria. Hart et al. (2019) investigates fertility response to the 2009 reform, finding no evidence of such effects.
Table 4: Paternity leave norms: Fuzzy RD of paternity leave use on leave for subsequent kids

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RD estimate</td>
<td>-0.15</td>
<td>-0.25</td>
<td>-0.064</td>
<td>-0.47</td>
<td>0.71*</td>
<td>-0.21</td>
<td>-0.11</td>
</tr>
<tr>
<td>conventional standard error</td>
<td>(1.49)</td>
<td>(0.38)</td>
<td>(0.15)</td>
<td>(0.38)</td>
<td>(0.34)</td>
<td>(0.16)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>robust standard error</td>
<td>1.81</td>
<td>0.45</td>
<td>0.18</td>
<td>0.45</td>
<td>0.40</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>conventional p-value</td>
<td>0.92</td>
<td>0.51</td>
<td>0.67</td>
<td>0.21</td>
<td>0.037</td>
<td>0.20</td>
<td>0.39</td>
</tr>
<tr>
<td>robust p-value</td>
<td>0.95</td>
<td>0.57</td>
<td>0.72</td>
<td>0.15</td>
<td>0.039</td>
<td>0.18</td>
<td></td>
</tr>
</tbody>
</table>

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

Table 4 provides the results of this exercise, for each reform separately and the pooled and stacked estimates for all reforms. Across the rows of Table 4 we see little evidence of any permanent impact on norms as measured by takeout of paternity leave for later kids: Except for the 2013 reform, where the efficient sample size is only 150 children and we find a marginally significant effect, none of the reforms provide statistically significant results, and point estimates are negative. Focusing on our preferred stacked estimates, the results indicate non-significant effect of .11 less weeks of leave for subsequent children for each week of leave for the first child, where the top of the 95% confidence interval rules out effects larger than around 0.1 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 of providing high quality subsidized child care to mothers in Norway on the child penalty. 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% of Norwegian 4 - 5 year olds attending care. For toddlers (ages 1-3), however, coverage was much lower at less than 50% and 30%, respectively, and the market was strongly rationed. These facts are documented in greater detail in Andresen and Havnes (2019), including additional evidence from surveys on the actual and preferred modes of child care for children at these ages. The underrepresentation of children between ages 1-3 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 1-2 year olds over the next years as shown in Appendix Figure D5a. However, the expansion varied considerably between municipalities and over time (see 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). We use the same variation to estimate the impact of increasing access to high quality formal child care on the child penalty in this section.

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 2002-2008.53 We assign

---

52 The prevalence of care is the result of a reform and gradual expansion of formal care for these children in the 1970’s (Havnes and Mogstad, 2011).
53 This includes a few thousand twins. Clustering at the municipality level accounts for within-family clustering.
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 endogenous outcome of the reform, and therefore do not restrict the sample to youngest children. To be consistent with the results we have presented thus far, we look at the effects on maternal and paternal income when the child is between the ages 0 through 5, and use the years before birth as placebo outcomes. This leaves us with a sample of around 103,000 couples\(^{54}\).

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:

\[
\begin{align*}
  y_{it} &= \pi_k + \gamma_{T,it} + \beta_{a_it} + \phi_t m_i + \epsilon_{it} \\
  m_i &= \tilde{\pi}_k + \tilde{\gamma}_{T,it} + \tilde{\beta}_{a_it} + \gamma_{1} CC_{1}^{k} + \gamma_{2} CC_{2}^{k} + \tilde{\epsilon}_{it}
\end{align*}
\]

(14)

Where \(\gamma_{T,it}\) are calendar year fixed effects, \(\pi_k\) are municipality fixed effects, \(\beta_{a,it}\) are age fixed effects for the parent (in years) and \(m_i\) is our measure of child care use from ages 13 - 36 months from the cash for care data. The instruments are \(CC_{1}^{k}\), the share of slots for 1-year olds in the municipality at age 1 to the population of

\(^{54}\)Notice that because we restrict to first born children, the sample size in this paper is a little less than half the size of the samples of cohabiting mothers and fathers in Andresen and Havnes (2019). This gives us less precision but is consistent with the rest of the paper. Because of the inherent focus on labor supply over time, we also measure child care use throughout the full 13 - 35 or 36 months period we can measure, in contrast to the preceding paper that is mostly concerned with child care use and labor supply during the calendar year the child turns 2.
one year olds, and $CC_k^2$, the same share for 2-year olds, measured at the relevant age of the child.

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

Table 5: First stage estimates, formal care use

<table>
<thead>
<tr>
<th>Years of child care use at ages 13 - 35 or 36 months</th>
<th>Coverage rate at age 1</th>
<th>Coverage rate at age 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.787*** (0.0543)</td>
<td>0.764*** (0.0552)</td>
</tr>
<tr>
<td></td>
<td>0.630*** (0.0629)</td>
<td>0.632*** (0.0660)</td>
</tr>
<tr>
<td>Municipality fixed effects</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Age profiles</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Education-specific age profiles</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>$N$</td>
<td>103,172</td>
<td>103,157</td>
</tr>
<tr>
<td>mean dep. var</td>
<td>1.031</td>
<td>1.031</td>
</tr>
<tr>
<td>$F$</td>
<td>167.9</td>
<td>138.5</td>
</tr>
</tbody>
</table>

Note: First stage estimates of eq. 14 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 5 column 1, where we see that the availability of slots in care has a strong influence on years of early care use. Expansions of care both at age 1 and at age 2 have a strong impact on 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 5: Impact of a year of child care use at ages 13-36 months scaled in percentages of estimated child penalty

Note: IV results from equation 14 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, where we additionally take these results and scale them with the estimated child penalties from the first part of the paper to estimate the change in the child penalty.
In Figure 5, 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 25-30% for mothers when their children are between the ages 2-3, but the impact appears only in the years of treatment. In 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 14. The first stage from this specification is hardly affected by this, as documented by column 2 in table 5. 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 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 same-sex couples experience a very different child penalty compared to heterosexual 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 turned 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 norms. In contrast, we show that early child care use reduces the child penalty for mothers by around 25% per year of use in the years of treatment. These results suggest that if policy makers wish to decrease the 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


MOBERG, Y. (2016). Does the gender composition in couples matter for the division of labor after childbirth?


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


For online publication: Appendix

This appendix is intended for online publication along with "What Causes the Child Penalty and How Can It Be Reduced? Evidence from Same-Sex Couples and Policy Reforms" by Martin E. Andresen and Emily Nix. Appendix A provides details on how we identify same-sex and heterosexual 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 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

When we compare child penalties for same-sex and heterosexual couples, we restrict to years where we are able to identify children born to same-sex couples.

While Norway was the second country in the world to legally recognize same-sex partnerships in 1993 through the Partnership Act, there were restrictions regarding children until 2002, when same-sex couples could legally adopt the children of their partners. The guidelines required a stable relationship and having had a de facto parenting role for the child in question for some period of time, most often five years, as well as consent from the existing parent. In practice the increasing use and availability of assisted fertility treatments among same-sex female couples

[Couples were not eligible for domestic adoptions, were not eligible for publicly subsidized assisted fertility treatment, and the registered spouse of a woman giving birth was not automatically registered as the second parent (as the pater est principle established for married heterosexual couples).]
challenged this five-year rule, as planned children of same-sex female couples conceived through assisted fertilization abroad became increasingly common. Therefore, in 2006 the Norwegian government clarified the rules so that the five-year rule would not apply in cases where the fatherhood cannot be established, such as with IVF treatment.

In 2009, a new marriage act was introduced which equalized same-sex and heterosexual marriages in all but one respect: A same-sex spouse cannot later adopt the child of his/her spouse that was in turn adopted from a country that does not allow adoptions to same-sex couples. The new marriage act from 2009 also gave same-sex female couples the right to IVF treatment in Norway, but only when using a non-anonymous donor, as the law requires all children conceived through IVF in Norway to have the possibility of knowing the identity of the donor father at age 18. Before this, same-sex female couples often traveled abroad to get IVF treatment, most often in Denmark. Even after the new law was passed, many couples still travel abroad either to speed up the process or because they want to use an anonymous donor. If conception happens through IVF treatment with a non-anonymous donor in a recognized (private or public) fertility clinic, co-mothership can now be registered at birth, but otherwise the couple must go through an adoption process in order for the partner to be formally registered as the co-mother.

For gay couples, getting children is naturally more complicated. Surrogacy is illegal in Norway. Gay couples can adopt children since 2009, but in practice

---

56 Some gay couples still enter into surrogacy agreements with surrogate mothers from abroad. No special rules apply to these children, and parenthood must be established according to the law when returning with the child. Typically, this means that the (most often biological) father will declare fatherhood upon returning to Norway and be registered as the father, and that the other spouse will then have to start the adoption process to be registered as co-father.
this is difficult since domestic adoption at birth is very rare in Norway and few international countries are willing to adopt children to these couples. As a result, we find very few gay couples with infants in the data. While the results for gay couples are consistent with our main conclusions, they are very imprecise so we do not include those results.

To construct our sample, we rely on registrations of legal parent status in the population registers. In practice, we therefore observe children appearing in same-sex couples at various times following birth, 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 where both parents were legally registered as parents

---

57Ruling out adoptions by near family and adoptions of foster- and step-children, as few as two to three children are adopted away at birth or right thereafter per year in Norway. In addition, the biological parents are given a say on prospective adoptive parents, and their opinion is given considerable weight in the decision among potential adoptive parents. This makes matters worse for same-sex couples if the biological parents prefer a heterosexual couple. In practice, this means that this option is not very relevant for same-sex couples.

58The first adoption from abroad to a same-sex couple in Norway happened in the fall of 2017, when Colombia became the first donor country to approve an adoption to a Norwegian same-sex couple following a controversial Supreme Court ruling from 2015. In the empirical analysis, we restrict attention to children born in 2014 at the latest, so that foreign adoptions to same-sex couples will not be relevant for this paper.

59These results are available upon request.
at the latest in the year the child turns 1 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 same-sex female couples who receive 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 A6 we graph the adoption of children by age and year to same-sex female couples. In Table A6 we report summary statistics for the heterosexual and same-sex female couples used in Section 5.

Figure A6: Registered children 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.
Table A6: Summary statistics by couple type

<table>
<thead>
<tr>
<th>Birth year (first child)</th>
<th>2001-2014</th>
<th>2001-2014</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Child characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth year</td>
<td>2007.7</td>
<td>2010.6</td>
</tr>
<tr>
<td></td>
<td>(4.00)</td>
<td>(2.89)</td>
</tr>
<tr>
<td>Multiple birth</td>
<td>0.020</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>Female child</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.49)</td>
</tr>
<tr>
<td>Age at adoption</td>
<td>0.028</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.81)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>B: Parent characteristics, year before birth</strong></th>
<th>Mother 1</th>
<th>Father 2</th>
<th>Mother 3</th>
<th>Co-mother 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent type (K)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at first birth</td>
<td>27.8</td>
<td>30.3</td>
<td>32.2</td>
<td>32.8</td>
</tr>
<tr>
<td></td>
<td>(4.23)</td>
<td>(5.03)</td>
<td>(4.12)</td>
<td>(5.64)</td>
</tr>
<tr>
<td>Labor income (1,000s of 2017 NOK)</td>
<td>339.6</td>
<td>471.9</td>
<td>488.9</td>
<td>480.0</td>
</tr>
<tr>
<td></td>
<td>(206.0)</td>
<td>(1055.8)</td>
<td>(196.7)</td>
<td>(308.3)</td>
</tr>
<tr>
<td>Years of education†</td>
<td>15.1</td>
<td>14.6</td>
<td>16.4</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td>(2.91)</td>
<td>(3.01)</td>
<td>(2.42)</td>
<td>(2.65)</td>
</tr>
<tr>
<td>Days of sickness absence year $t-2$</td>
<td>7.25</td>
<td>6.7</td>
<td>11.7</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>(31.7)</td>
<td>(31.2)</td>
<td>(37.8)</td>
<td>(46.5)</td>
</tr>
</tbody>
</table>

$N$ couples: 251,490, 634

**Note:** Summary statistics on estimation samples constructed as described in this Section. Standard deviations in parentheses.
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 shows means of annual labor earnings for the years before and after birth of the first child. Bottom panel shows simple event study estimates without year and age profiles. Sample construction and data as defined in Section 4.

Figure B2 provides subsample analysis by (birthing) mother’s education, revealing relatively similar effects across groups. Figure B3 show that the same
patterns for comparative advantage still hold when we use pre-birth education differences as a proxy for comparative advantage instead of pre-birth income difference within the couple. Last, Figure B4 shows that the pattern remains similar even when we restrict to couples where the woman in the couple makes more than her male partner.

(a) Heterosexual couples, high ed. mother
(b) Same-sex female couples, high ed. mother
(c) Heterosexual couples, low ed. mother
(d) Same-sex female couples, low ed. mother

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 years of education differences in $t - 1$

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 11) we specify an event study model that controls for age profiles and yearly shocks. While our sample of same-sex female couples is large relative to most other papers studying 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 11 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 the 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 better educated than heterosexual couples, we might worry that they have a different age profile entering into 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 are 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 not only be gender-specific, but also education level specific, where education is measured in 9 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

\[ \text{footnote: We discuss this in the context of age profiles, but similar arguments can be made for yearly shocks, and when implementing alternative models, we relax both.} \]
catch-up among heterosexual couples when accounting for these flexible age profiles, the contrast to same-sex female couples is still remarkable and comparable to the baseline estimates.

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 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 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 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 shows that the restrictions we make are reasonable.

In the bottom left panel of figure \(\text{B6}\) 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 to 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.
Figure B6: Fully flexible age profiles and yearly shocks on a (more) balanced sample

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 similar completed fertility. This suggests that the differences in child penalties are not driven by the differences in number of additional children following the first child. Figure B8 shows that the estimated child penalties are similar to the baseline and show the same stark differences across couple types when we restrict to the 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 $t + 5$
Figure B9: Child penalties in matched samples

Note: Figure shows child penalties estimated from the baseline model in a sample matched to the same-sex female couples on pre-birth characteristics in a propensity score matching exercise (top panel) and a nearest neighbor matching exercise (bottom panel). In the sample of heterosexual and same-sex female couples before birth, we estimate a logit model for being a same-sex female couple. Covariates include each partner’s age and their interaction, each spouse’s years of education and their interaction and a full set of municipality dummies. In the top panel we then weight the baseline model with the propensity score from this model, so that heterosexual couples who look more like same-sex female couples on observables are given higher weight. The entire procedure is bootstrapped, clustering on couple. In the bottom panel, we conduct a nearest neighbor matching exercise, selecting heterosexual couples who look most like the same-sex female couples.
B.3 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 like in equation 12. Therefore, the estimates are interpretable as the effects of children at age (in months) \( j \), relative to the effect 12 months before birth.

Results are presented in figures B10 and B11. We begin in figure B10 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 to the baseline employment rate 12 months before birth. In panel (c) we estimate impacts on a dummy indicating a full time job, as defined by contracted weekly hours above 30. The fact that the impact on this measure is larger than on the employment measure, at around a 20 percentage
points reduction, indicates that there is response both on the extensive and intensive margins of labor force participation: some mothers drop out of the labor force entirely while others reduce labor supply and work part time following child birth. As before, we find little response among heterosexual fathers for these measures.

For 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 B11, 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 B11. We see strong positive trends in public sector employment for mothers in heterosexual couples around child bearing. Ignoring the dip in the year of birth that is likely caused by the very low employment rates of new mothers, mothers move into the public sector in anticipation of - and following - child birth, while this trend is flat for men. The trend in this outcome is relatively similar for both partners in 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.\footnote{Despite this, we occasionally see non-employed individuals in these data. We exclude the few}
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.\textsuperscript{62} 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.

\textsuperscript{62}Note that some of this could be caused by subsequent pregnancies.
Figure B10: Decomposition I: Child penalties for heterosexual (left) and same-sex female (right) couples
Figure B11: Decomposition II: Child penalties for heterosexual (left) and same-sex female (right) couples
B.4 Child penalties over time

Figure B12 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.

![Figure B12: The child penalty in income over time for mothers and fathers in heterosexual couples](image)

*Note: Child penalties estimated separately by birth cohort of first child in 5-year intervals. Estimated using the event study framework from equations 11 and 12.*
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% of the leave spells recorded are for regular leave spells, and we focus on these.

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

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

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

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

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

C.1 Balancing tests

Table C.1 provides sharp RD balancing tests for a range of covariates in the baseline RD model. Figure C.1 provides 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 C.3.
Table C1: Sharp RD balancing tests

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

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

Table C2 presents first stage results for each reform separately. Figure C2 reports the impacts on mothers’ and fathers’ annual incomes over time using 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 takeout 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.
Table C2: RDD first stage estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RD estimate per week</td>
<td>0.83***</td>
<td>1.09***</td>
<td>0.96***</td>
<td>0.88***</td>
<td>0.74**</td>
<td>0.78***</td>
</tr>
<tr>
<td>conventional standard error</td>
<td>(0.34)</td>
<td>(0.33)</td>
<td>(0.094)</td>
<td>(0.28)</td>
<td>(0.24)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>robust standard error</td>
<td>0.42</td>
<td>0.40</td>
<td>0.11</td>
<td>0.33</td>
<td>0.28</td>
<td>0.13</td>
</tr>
<tr>
<td>conventional p-value</td>
<td>0.015</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>robust p-value</td>
<td>0.024</td>
<td>0.006</td>
<td>0.000</td>
<td>0.004</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td>14,658</td>
<td>15,138</td>
<td>16,556</td>
<td>16,558</td>
<td>16,268</td>
<td>14,330</td>
</tr>
<tr>
<td>Optimal bandwidth</td>
<td>60.7</td>
<td>55.2</td>
<td>74.4</td>
<td>45.3</td>
<td>63.0</td>
<td>43.3</td>
</tr>
<tr>
<td>Efficient observations</td>
<td>5006</td>
<td>4,901</td>
<td>6,993</td>
<td>4,467</td>
<td>6,120</td>
<td>3,815</td>
</tr>
<tr>
<td>Weights in pooled</td>
<td>0.15</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
<td>0.18</td>
<td>0.16</td>
</tr>
<tr>
<td>Weights in stacked</td>
<td>0.16</td>
<td>0.16</td>
<td>0.22</td>
<td>0.14</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td>Quota increase</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>-4</td>
</tr>
</tbody>
</table>

Notes: Robust semiparametric sharp RD estimates of the effect of paternity leave reforms on paternity leave takeout using optimal bandwidths, triangular kernel and local linear polynomials on either side of the cutoff. All estimates are scaled to reflect one week of quota increase. Stacked estimates are the stacked individual cutoffs, allowing polynomials to vary over cutoffs and using the cutoff-specific bandwidths and weights. Conventional standard errors are heteroskedasticity-robust, but not bias-corrected. ***p < 0.01, **p < 0.05, *p < 0.1, using conventional, heteroskedasticity-robust standard errors.
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. Pooled estimate refers to the simple re-centered estimate, while the stacked estimate stacks the cutoff-specific specifications. Robust bias-correcting inference reported for the pooled estimate, conventional, heteroskedasticity-robust inference for the stacked estimate.
Figure C3: Robust RDD estimates, paternity leave reforms

Notes: First and third columns show binned plots of the weeks of paternity leave against birth date of the child in days after the reform, overlaid with the estimated local linear polynomials. Second and fourth panels show sharp RD estimates of the impact of an additional week of quota on maternal and paternal income by year. Optimal MSE-reducing bandwidths, triangular kernel and local linear polynomials on either side of cutoff. Confidence intervals are robust and bias-corrected.
C.3 Accounting for effects of maternal leave

As evident from Table 2, 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 to the change in paternity leave, and at much higher margins, we might worry that it is partly the changed maternity leave that causes any changes in later labor market outcomes, not paternity leave.

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

\[ y_{irt} = \beta^L_l L_{ir} + \beta^M_t M_{ir} + \varphi^0_r x_i \mathbb{1}(x_i < 0) + \varphi^1_r x_i \mathbb{1}(x_i \geq 0) + \pi_r + \epsilon_{irt} \]
\[ L_{ir} = \gamma_{LQ} Q_{ir} + \gamma_{LS} S_{ir} + \varphi^0_r x_i \mathbb{1}(x_i < 0) + \varphi^1_r x_i \mathbb{1}(x_i \geq 0) + \eta^L_{ir} \]
\[ M_{ir} = \gamma_{MQ} Q_{ir} + \gamma_{MS} S_{ir} + \varphi^0_r x_i \mathbb{1}(x_i < 0) + \varphi^1_r x_i \mathbb{1}(x_i \geq 0) + \eta^M_{ir} \]

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 C3. 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.

Table C3: First stage effects of maternity and paternity leave quotas

<table>
<thead>
<tr>
<th>Weeks of leave</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mother</td>
</tr>
<tr>
<td></td>
<td>reform</td>
</tr>
</tbody>
</table>

A: 50-day bandwidth

| Paternity leave quota | 0.066 | 1.00*** | 2005 | 50 | 4,037 |
| (Q_{rt})             | (0.14) | (0.16)  |      |    |       |
| Remaining leave for mother | 0.77*** | 0.18 | 2011 | 50 | 3,656 |
| (S_{rt})             | (0.21) | (0.21)  |      |    |       |
| joint F              | 21.1   | 76.5    | 2014 | 50 | 3,902 |

N = 24,520

B: Maternity leave-optimal bandwidth

| Paternity leave quota | 0.069 | 0.96*** | 2005 | 66.9 | 5,418 |
| (Q_{rt})             | (0.14) | (0.15)  |      | 2006 | 61.3  |
| Remaining leave for mother | 0.79*** | 0.12 | 2009 | 42.4 | 4,017 |
| (S_{rt})             | (0.20) | (0.20)  |      |     |       |
| joint F              | 24.8   | 79.7    | 2014 | 52.4 | 4,532 |

N = 29,028

C: Paternity leave-optimal bandwidth

| Paternity leave quota | 0.055 | 0.98*** | 2005 | 58.0 | 4,770 |
| (Q_{rt})             | (0.14) | (0.15)  |      | 2006 | 68.4  |
| Remaining leave for mother | 0.73*** | 0.12 | 2009 | 44.6 | 4,192 |
| (S_{rt})             | (0.21) | (0.21)  |      |     |       |
| joint F              | 18.0   | 74.9    | 2014 | 40.8 | 3,502 |

N = 27,344

Note: First stage results from stacked specification of all six parental leave reforms, instrumenting for weeks of paternity and maternity leave take up as described in eq. [15] 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. 15. 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 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. Estimates are significant at the 5% level at age 3 and 10% level at age 2, and thus indicate that there is some immediate effects of child care use on earnings during the years of treatment, perhaps driven by allowing mothers to return to work earlier after child birth.

Results for fathers are noisy, but point, if anything, to negative impacts on earnings, which could also reduce 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.

63This 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 D5: Child Care Coverage

Source: Statistics Norway Statistikkbanken, tables 09169 and 07459.

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 municipality- and year fixed effects and an interaction of pre-reform characteristics interacted with year dummies, in a sample of municipalities over time. Plotted are the year-specific impact of the pre-reform characteristics on expansion of care in a particular year. 95% confidence intervals in grey, clustered at the municipality level.
Figure D7: Impact of early child care use on parents’ earnings

Note: IV results from equation IV 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. IV in the main paper.