Online Appendix

Children and Gender Inequality: Evidence from Denmark

Henrik Kleven, Princeton University and NBER
Camille Landais, London School of Economics
Jakob Egholt Søgaard, University of Copenhagen

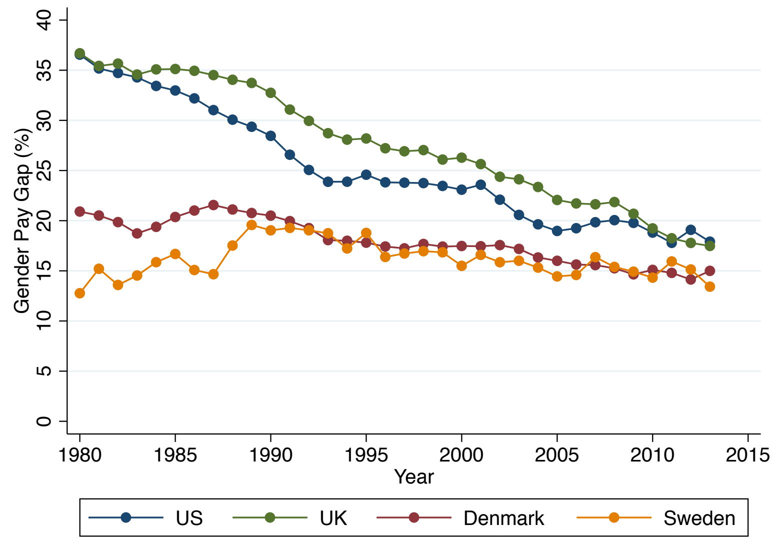
October 2018

A Supplementary Figures and Tables

Figure A.I: Gender Gaps Across Countries 1980-2013

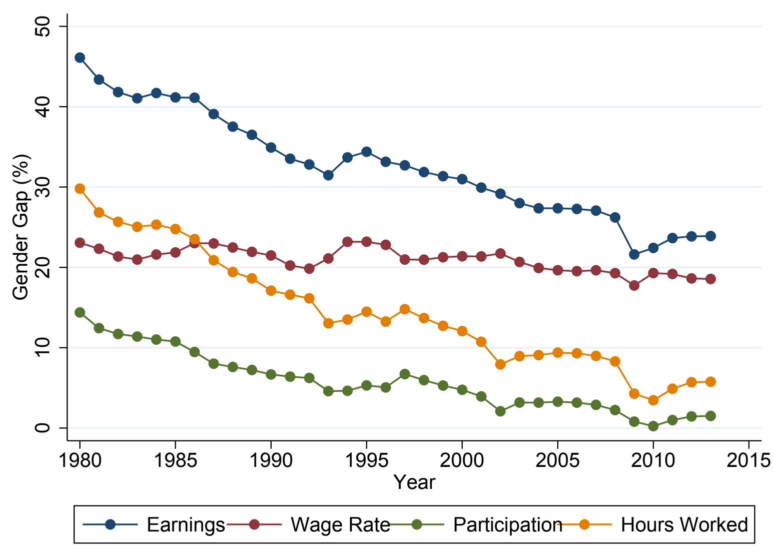
A: Convergence of the Gender Pay Gap Across Countries

Median Earnings for Full-Time Workers



B: Evolution of Gender Gaps in Denmark

Means for All Workers

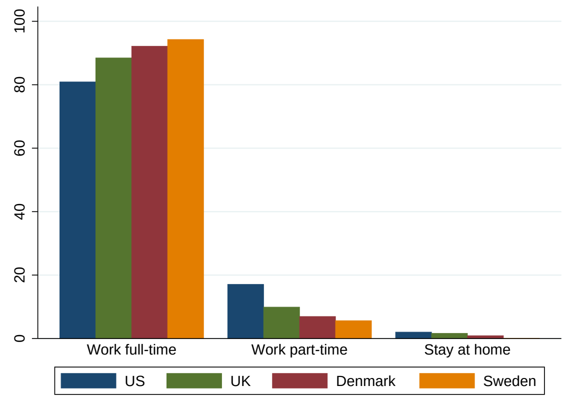


Notes: The time series in Panel A are drawn from OECD.org, except for Denmark where we use our own calculations of median earnings for full-time workers aged 16-64 (where full time is defined based on the ATP hours measure described in section 2.2). Our calculations for Denmark uses the same underlying data as the official OECD series, but is more consistent with the sample definitions used for the other countries. In Panel B the gaps in earnings and participation are calculated for the entire population aged 16-64 regardless of employment status, while the gaps in hours worked and wage rates are calculated conditional on participation.

Figure A.II: Gender Norms Across Countries

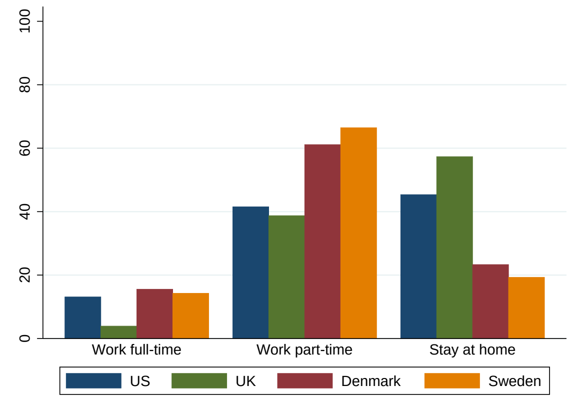
A: Women Without Children

Do you think that women should work outside the home full-time, part-time or not at all when they are married but with no children?



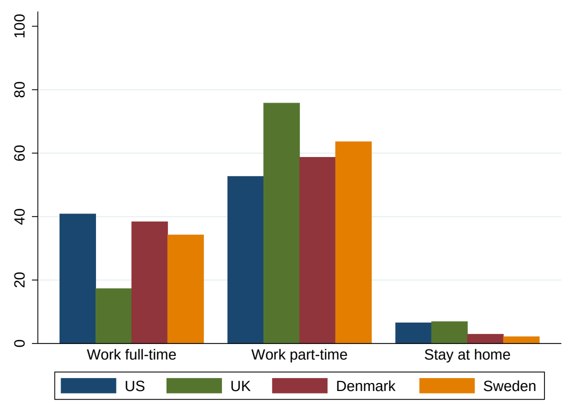
B: Women With Children Under School Age

Do you think that women should work outside the home full-time, part-time or not at all when there is a child under school age?



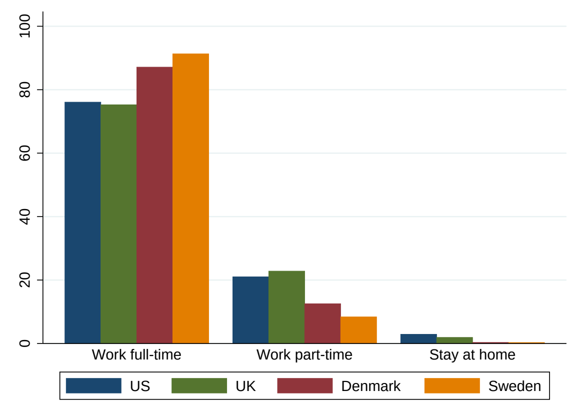
C: Women With Children In School

Do you think that women should work outside the home full-time, part-time or not at all when the youngest child is still in school?



D: Women With Children Who Have Left Home

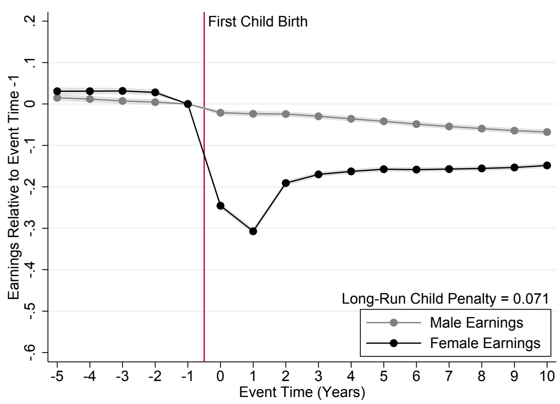
Do you think that women should work outside the home full-time, part-time or not at all when the child has left the home?



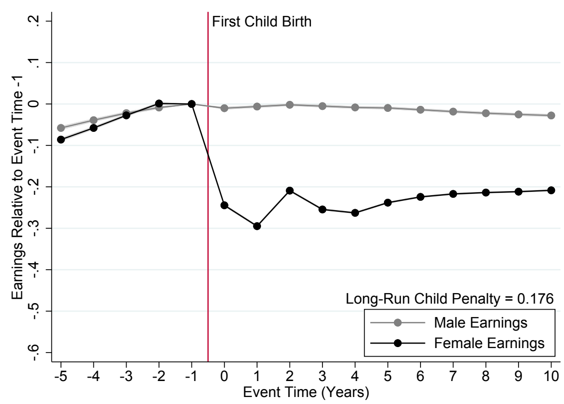
Notes: The figure is based on data from the International Social Survey Program (ISSP) in 2002. Each panel shows shares (in percent) choosing each of the 3 listed categories.

Figure A.III: Earnings Impacts by Number of Children

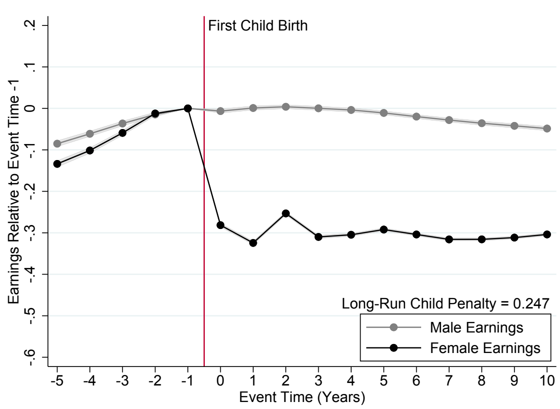
A: One-Child Mothers



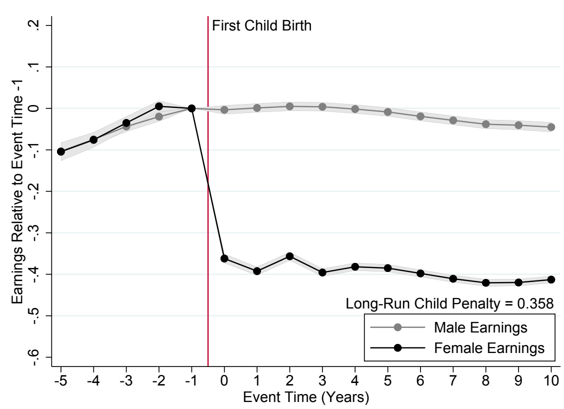
B: Two-Child Mothers



C: Three-Child Mothers

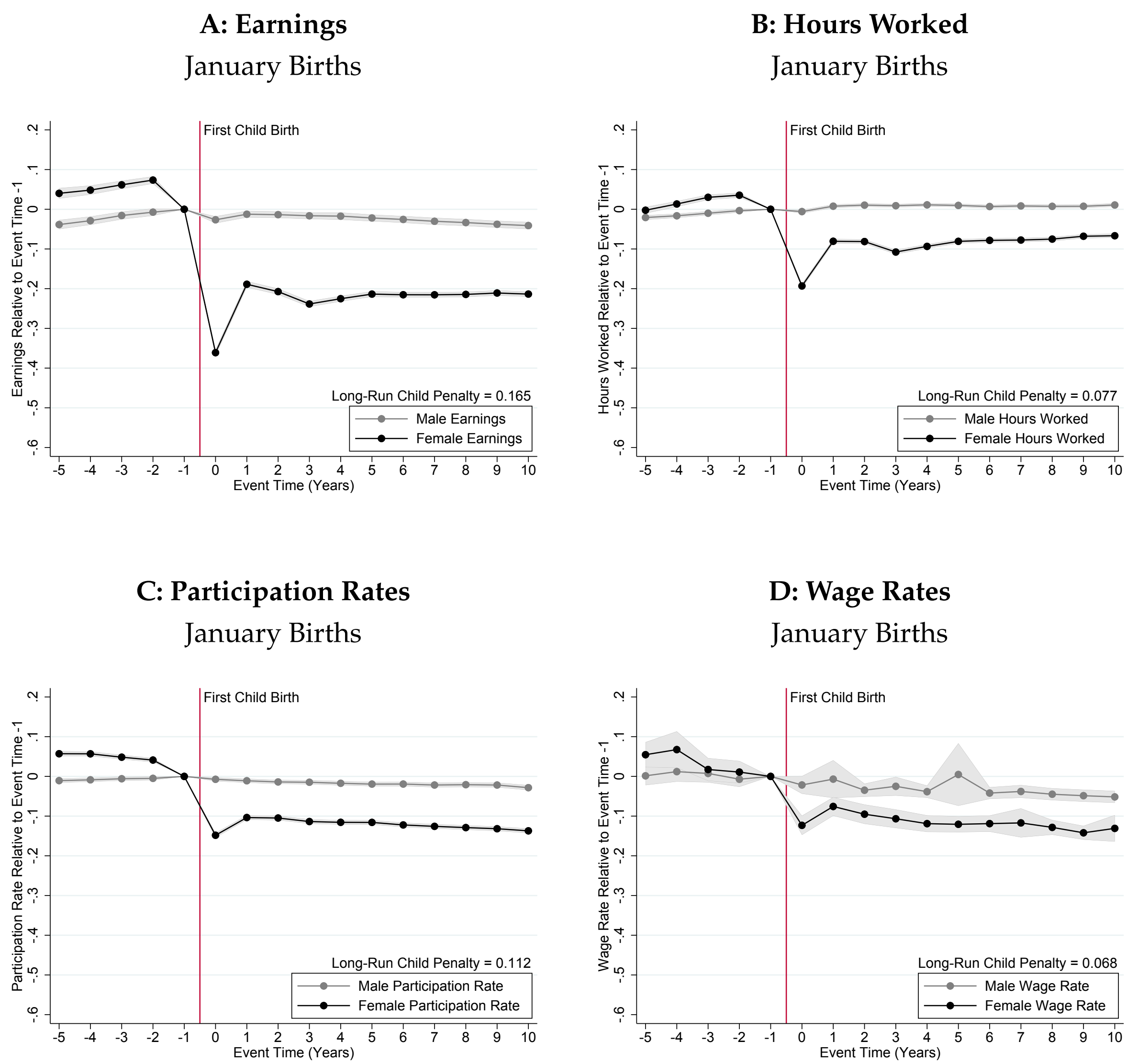


D: Four-Child Mothers



Notes: The figure shows the impact of children on earnings exactly as in Figure 1A, but splitting the sample by the woman's total number of children as of 2013 (1, 2, 3, or 4 children).

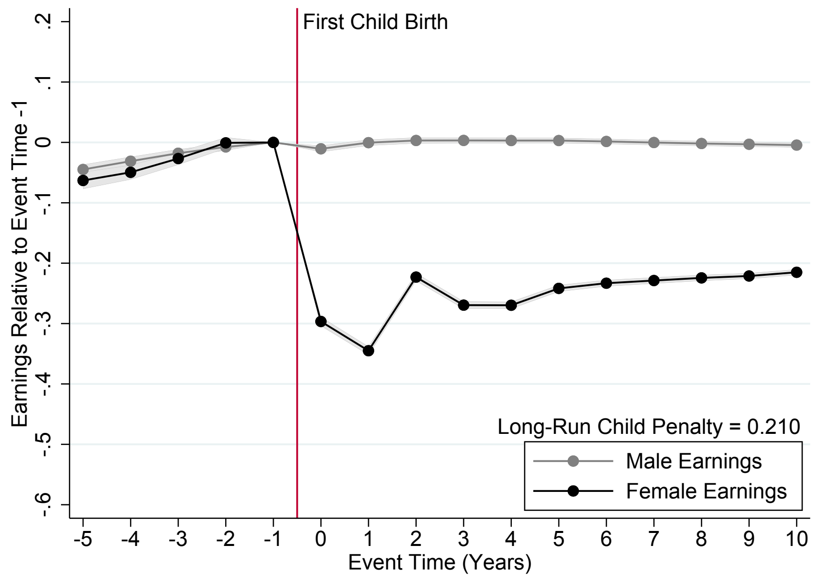
Figure A.IV: Impacts of Children Born in January



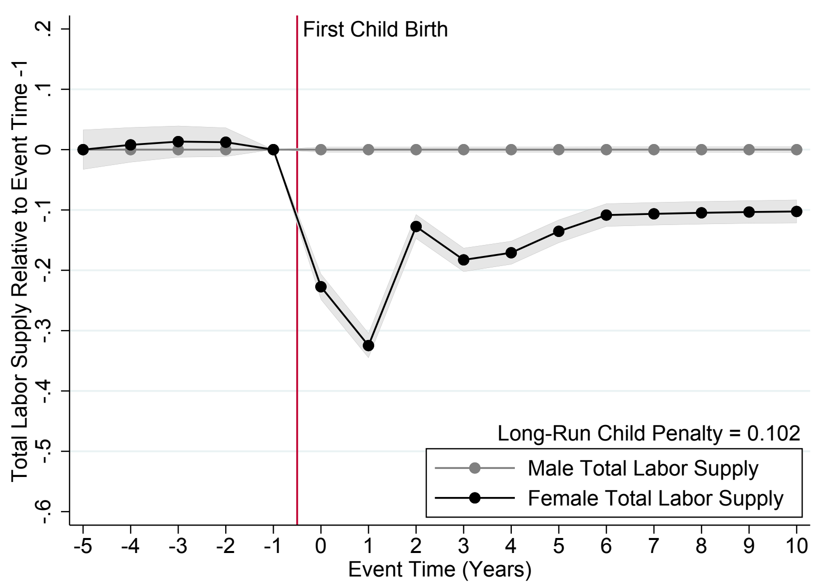
Notes: The figure is constructed as Figure 1, but estimated on the subsample of individuals who have their first child in January.

Figure A.V: Median Impacts of Children

A: Earnings



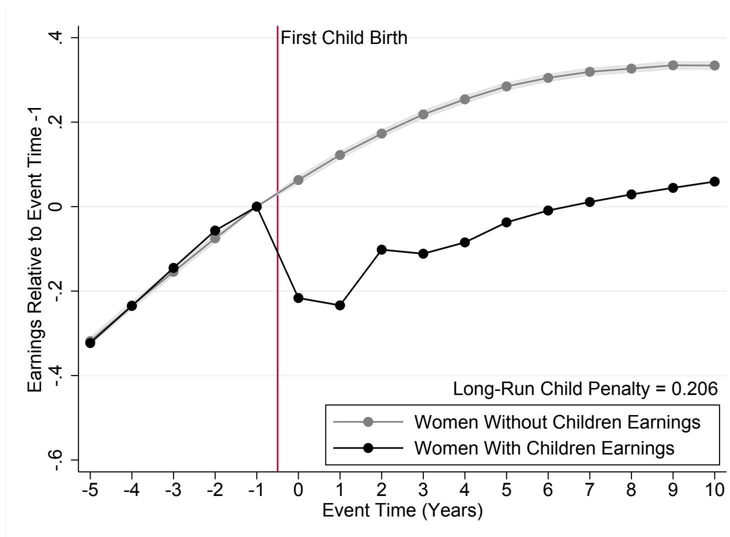
B: Total Labor Supply



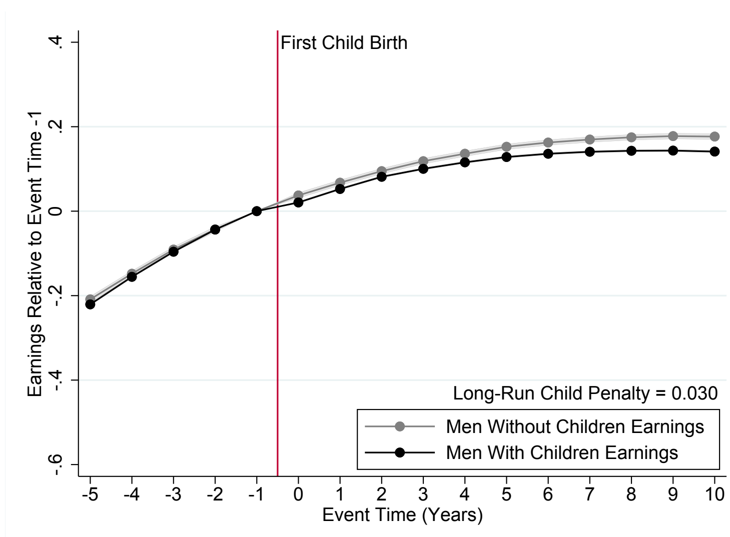
Notes: The figure is constructed in the same way as in Figure 1, but showing median impacts on earnings and total hours worked (including the zeros, thereby combining the intensive and extensive margins). These quantile regressions are based on a 1/7 subsample, which makes the confidence bands somewhat larger.

Figure A.VI: Earnings Impact of Children in a Difference-in-Differences Event Study Design

A: Women Who Have Children vs Women Who Don't



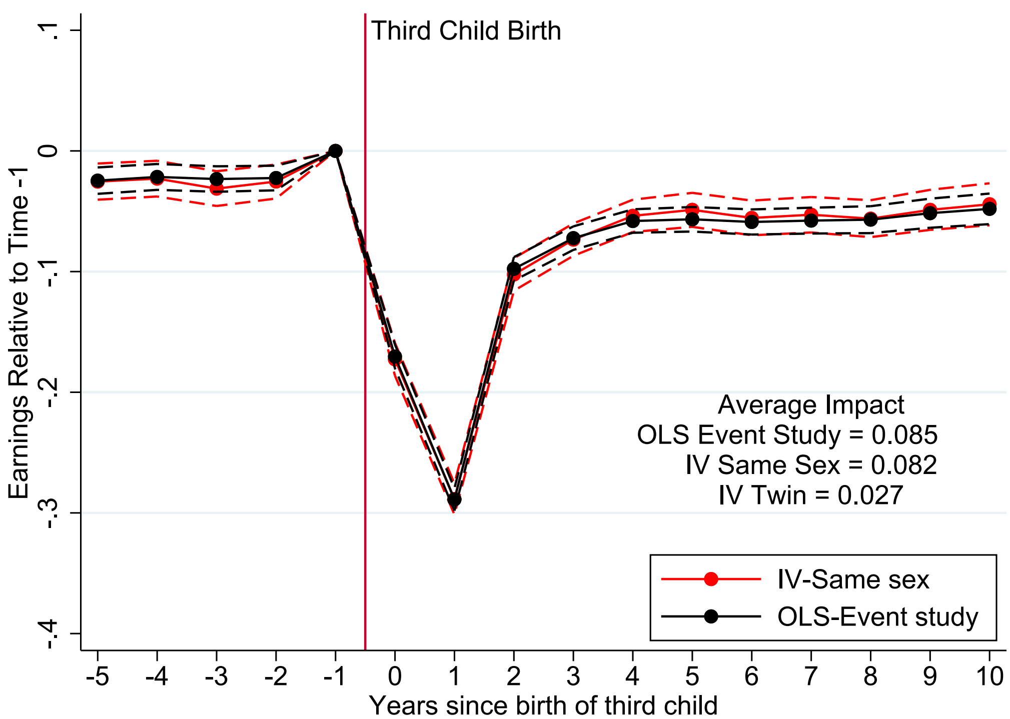
B: Men Who Have Children vs Men Who Don't



Notes: The figure show the evolution of earnings relative to the year before the birth of the first child for individuals with children compared to those who never have children (assigning placebo births based on the observed distribution of age at first child among those who have children). Panel A shows the evolution in earnings for women and Panel B the evolution for men. In the figure we control for year fixed effects. The details of how we construct the control groups of men and women who never have children are described in Appendix B.2. The figure reports long-run child penalties for men and women separately, estimated as a difference-in-differences between those who have children and those who never have children (as opposed to previous penalty measures based on comparing men and women, both of whom have children). The shaded 95 % confidence intervals are based on robust standard errors.

Figure A.VII: Earnings Impact of Third Child in Event Study vs IV Designs

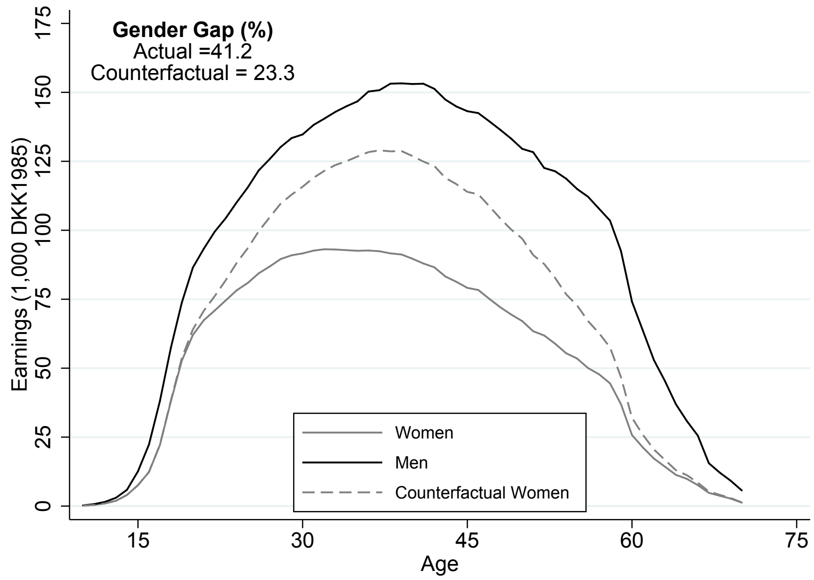
Sibling Sex Mix or Twin Births as Instruments



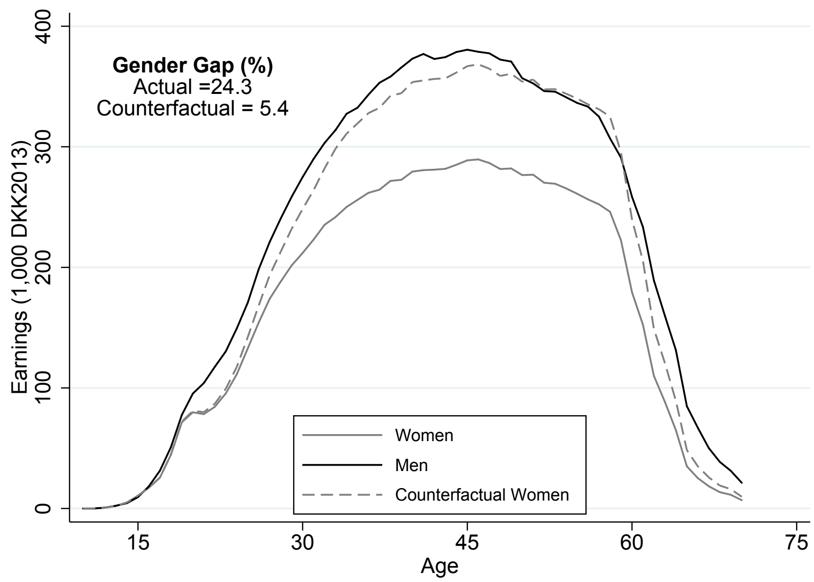
Notes: The figure shows the earnings impact of a third child on women ($P^w_{t'}$ defined in section B.2) obtained from the OLS event study specification (black series) and the IV same-sex specification (red series) as a function of years since the birth of the third child. The event study estimates are based on specification (6) in Appendix B.3. The IV-specification is based on the same specification using the sex mix of the first two children as instrument. The 95 % confidence intervals are based on robust standard errors. The figure also compares the average impact (i.e., across event times 0-10) obtained from the event study, the IV using sibling sex, and an IV using twins in the second birth.

Figure A.VIII: Decomposing the Age Profiles of Gender Inequality in Earnings

A: 1985-Profile Decomposed



B: 2013-Profile Decomposed

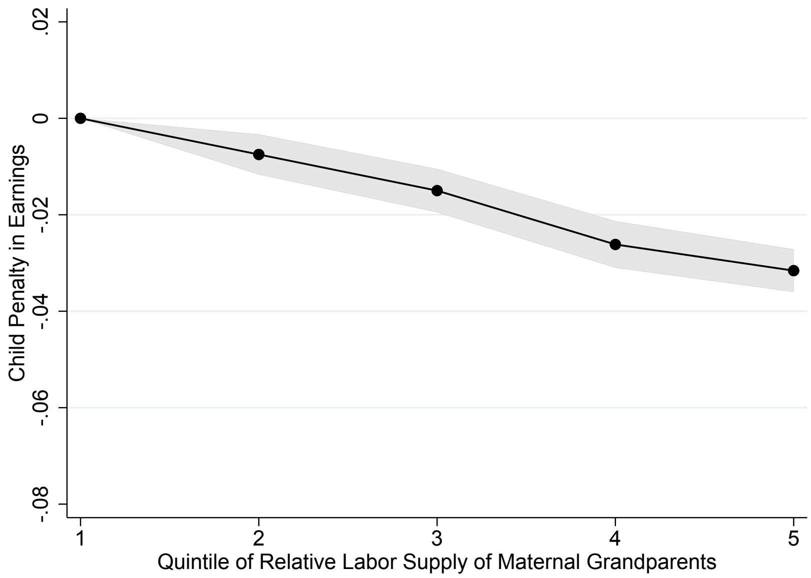


Notes: Based on the child penalties estimated from specification (3) in the main paper, this figure decomposes the within-year age profiles of gender inequality in earnings for 1985 and 2013. The difference between actual male earnings (solid black) and counterfactual female earnings (dashed grey) corresponds to non-child inequality, while the difference between counterfactual female earnings (dashed grey) and actual female earnings (solid grey) corresponds to child-related inequality. These within-year age profiles aggregate to the averages shown in Figure 5A for those two years.

Figure A.IX: Intergenerational Transmission of Child Penalties with Richer Controls

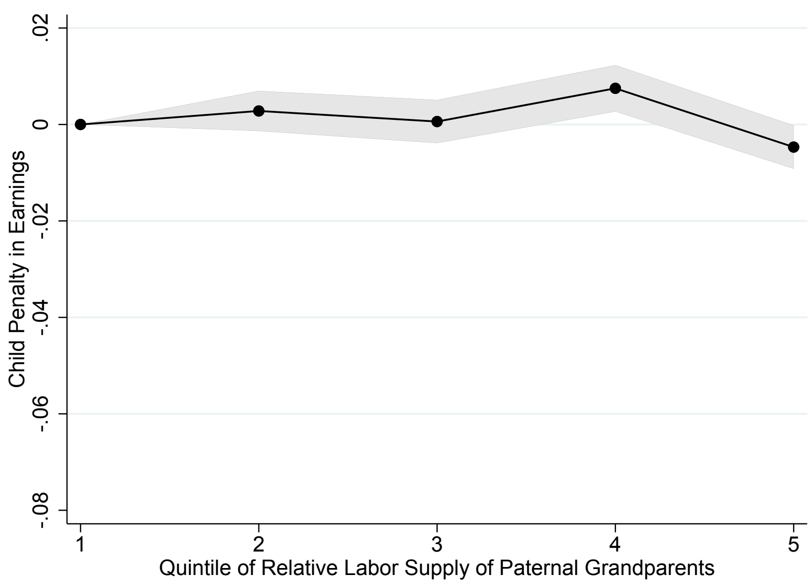
A: Maternal Grandparents

With Education-by-After Controls



B: Paternal Grandparents

With Education-by-After Controls

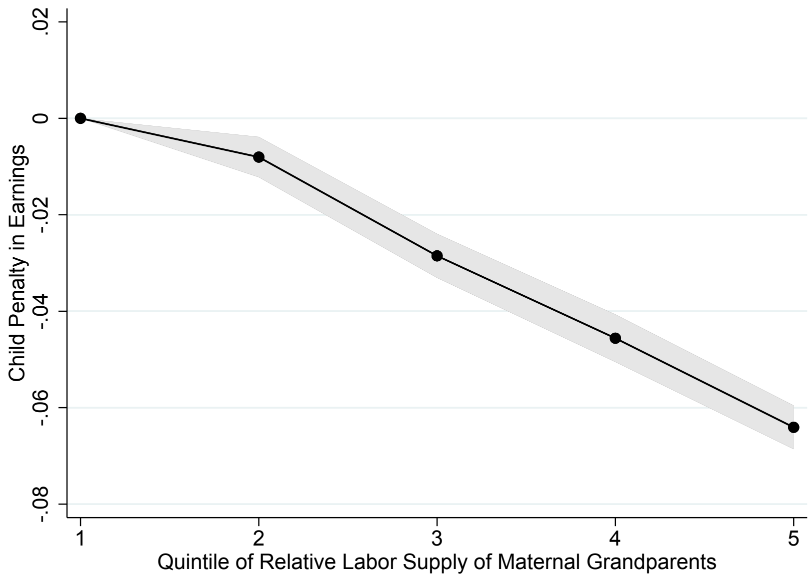


Notes: The figure shows the child penalty in earnings against quintiles of the relative labor supply distribution of the grandparents. The relative labor supply of grandparents is based on cumulated hours worked over the period 1964-79 (obtained from ATP pension contributions). The child penalties are estimated using equation (5) in the main paper and the statistic reported is P_q as defined in Section 5.2 extended with a rich set of non-parametric controls for the education level/fields of the grandparents interacted with the after dummy. The shaded 95 % confidence intervals are based on robust standard errors.

Figure A.X: Intergenerational Transmission of Child Penalties Controlling for the Fertility Channel

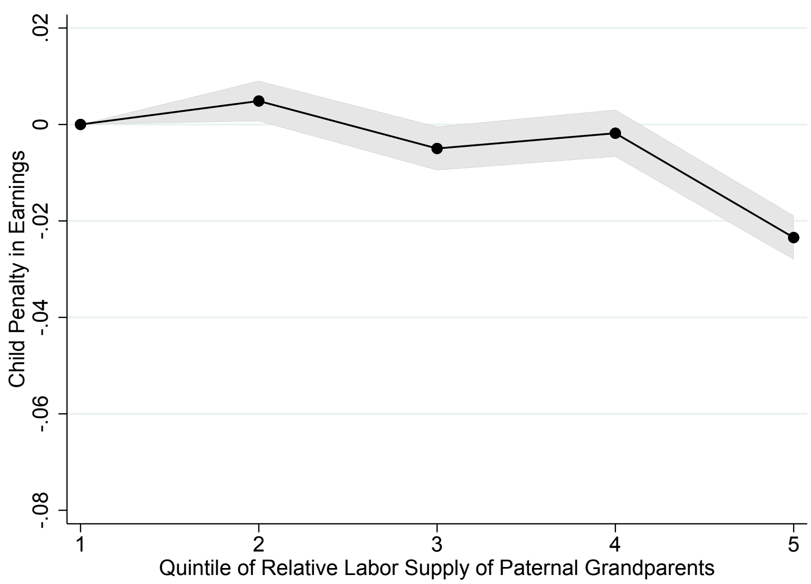
A: Maternal Grandparents

Controlling for Parental Fertility



B: Paternal Grandparents

Controlling for Parental Fertility



Notes: The figure shows the child penalty in earnings against quintiles of the relative labor supply distribution of the grandparents. The relative labor supply of grandparents is based on cumulated hours worked over the period 1964-79 (obtained from ATP pension contributions). The child penalties are estimated using equation (5) in the main paper and the statistic reported is P_q as defined in Section 5.2 extended with non-parametric controls for the complete fertility of the women. The shaded 95 % confidence intervals are based on robust standard errors.

Table A.I: Robustness Checks: Intergenerational Transmission of Child Penalties

Child Penalty Estimates by Quintiles of the Relative Labor Supply of the Grandparents

Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Maternal Grandparents							
2nd Quintile	0.000 (0.002)	-0.001 (0.002)	-0.009 (0.002)	-0.008 (0.002)	-0.008 (0.002)	-0.008 (0.002)	-0.008 (0.002)
3rd Quintile	-0.013 (0.002)	-0.014 (0.002)	-0.029 (0.002)	-0.029 (0.002)	-0.028 (0.002)	-0.015 (0.002)	-0.029 (0.002)
4th Quintile	-0.029 (0.002)	-0.027 (0.002)	-0.041 (0.003)	-0.044 (0.003)	-0.045 (0.003)	-0.026 (0.002)	-0.046 (0.003)
5th Quintile	-0.061 (0.002)	-0.061 (0.002)	-0.066 (0.002)	-0.065 (0.002)	-0.065 (0.002)	-0.032 (0.002)	-0.064 (0.002)
Panel B: Paternal Grandparents							
2nd Quintile	0.007 (0.002)	0.007 (0.002)	0.003 (0.002)	0.005 (0.002)	0.005 (0.002)	0.003 (0.002)	0.005 (0.002)
3rd Quintile	0.003 (0.002)	0.002 (0.002)	-0.007 (0.002)	-0.005 (0.002)	-0.005 (0.002)	0.001 (0.002)	-0.005 (0.002)
4th Quintile	0.01 (0.002)	0.011 (0.002)	0.005 (0.002)	0.001 (0.002)	-0.001 (0.002)	0.007 (0.002)	-0.002 (0.002)
5th Quintile	-0.023 (0.002)	-0.023 (0.002)	-0.025 (0.002)	-0.024 (0.002)	-0.025 (0.002)	-0.005 (0.002)	-0.024 (0.002)
Grandparental Controls							
Birth Cohort	X	X	X	X	X	X	X
Wealth Quintiles		X	X	X	X	X	X
Wealth Quintiles × After			X	X	X	X	X
Region				X	X	X	X
Region × After				X	X	X	X
Education					X	X	X
Education × After						X	
Parental Controls							
Fertility							X
Fertility × After							X

Notes: This table shows child penalty estimates by quintiles of the relative labor supply of the maternal grandparents (Panel A) and the paternal grandparents (Panel B) under different sets of controls in columns (1)-(7). Wealth controls consist of dummies for the within-generation wealth rank (quintiles) of the grandparents. Region controls add dummies for the region of residence of the grandparents. Education controls add dummies for both the the length and field of education (22 dummies) for both the grandmother and the grandfather. Fertility controls consist of dummies for the total number of children of the parents (measured in 2014). Specification (1) corresponds to the estimates shown in Figure 6, specification (5) to Figure 7, specification (6) to Figure A.IX, and specification (7) to Figure A.X.

B Identification

B.1 Conceptual Framework

In this section we set out a simple conceptual framework to clarify what is being estimated in the event studies. We denote by $\mathbf{k_i} = (0, ..., k_{it}, ..., k_{iT})$ the anticipated lifetime path of fertility for individual i: the individual starts life with zero children, has k_{it} children at time t, and ends up with k_{iT} children over the lifetime. Earnings at time t are chosen based on the number of children present at time t as well as the anticipated lifetime path of fertility. Specifically,

$$\begin{aligned} Y_{it} &= F\left(k_{it}, \mathbf{x_{it}}, \mathbf{z_{it}}\right) \\ &= F\left(k_{it}, \mathbf{x}\left(k_{it}, \mathbf{k_i}, \mathbf{z_{it}}\right), \mathbf{z_{it}}\right), \end{aligned} \tag{1}$$

where Y_{it} is earnings, $\mathbf{x_{it}} = \mathbf{x}\left(k_{it}, \mathbf{k_i}, \mathbf{z_{it}}\right)$ is a set of earnings determinants that are chosen based on children, and $\mathbf{z_{it}}$ is a set of earnings determinants that do not depend on children. Compared to the empirical specification (1) in the main paper, we simplify notation by leaving out indexation of gender and calendar time. The elements of $\mathbf{x_{it}}$ include variables such as hours worked, occupation, sector and firm — variables which we have seen respond to children in the event studies — while the elements of $\mathbf{z_{it}}$ include factors such as age, ability and preferences. Hence, in this framework earnings may respond directly to children conditional on choices (e.g. the impact of being tired or distracted at work) and indirectly through labor market choices $\mathbf{x_{it}}$ (e.g. the impact of switching to a lower-paying, but more family-friendly firm). Furthermore, we allow labor market choices $\mathbf{x_{it}}$ to respond both to the contemporaneous number of children k_{it} and to the entire path of past and future fertility. The latter effect captures for example that some women may take less education or opt for family-friendly career tracks knowing that they will eventually have many children.

While we do not specify the demand for children, we make the assumption that children k_{it} are exogenous to the outcome variable Y_{it} conditional on the set of underlying determinants $\mathbf{z_{it}}$. The assumption that "the event" (in our case, child birth) is not determined by the outcome variable is fundamental to any event study analysis. The graphical evidence presented above lends support to this assumption: there is no indication that outcomes respond *prior* to child birth (or prior to pregnancy as discussed above); the sharp breaks in career trajectories always occur *just after* having children.

This framework allows for two conceptually different effects of children on earnings. One is a *pre-child effect* of future children, conditional on the current number of children k_{it}, which operates through the dependence of labor market choices $\mathbf{x_{it}}$ on anticipated lifetime fertility $\mathbf{k_i}$. The other is a *post-child effect* of current children, conditional on anticipated lifetime fertility $\mathbf{k_i}$, which operates through both the direct effect of k_{it} and the effect of k_{it} on labor market choices $\mathbf{x_{it}}$. An obvious but important point is that the event studies cannot capture pre-child effects — these are incorporated in the pre-event levels that are differenced out — and is designed to identify only post-child effects. If women are investing less in education and career in anticipation of motherhood (as the child penalty sharply reduces the return to such investments), then the pre-child effect on female earnings is negative and the event study provides a lower bound on the total effect.[1]

Under what conditions do the event studies correctly identify the post-child impacts? It is important to distinguish between short-run and long-run impacts. The short-run impact is estimated by comparing event times *just* before and after time zero. Denoting these event times by t_-, t_+ and using equation (1), the short-run event study estimates capture

$$\mathsf{E}\left[Y_{it_+} - Y_{it_-}\right] = \mathsf{E}\left[F\left(1, \mathbf{x}\left(1, \mathbf{k_i}, \mathbf{z_{it_+}}\right), \mathbf{z_{it_+}}\right)\right] - \mathsf{E}\left[F\left(0, \mathbf{x}\left(0, \mathbf{k_i}, \mathbf{z_{it_-}}\right), \mathbf{z_{it_-}}\right)\right], \tag{2}$$

when we do not directly control for elements for $\mathbf{z_{it}}$ through for example age and year dummies. Assuming smoothness of the average non-child earnings path, i.e. $\mathsf{E}\left[F\left(0, \mathbf{x}\left(0, \mathbf{k_i}, \mathbf{z_{it_-}}\right), \mathbf{z_{it_-}}\right)\right] \approx \mathsf{E}\left[F\left(0, \mathbf{x}\left(0, \mathbf{k_i}, \mathbf{z_{it_+}}\right), \mathbf{z_{it_+}}\right)\right]$, equation (2) identifies the short-run effect of the first child conditional on $\mathbf{z_{it_+}}$. With direct controls for $\mathbf{z_{it}}$, the smoothness assumption can be relaxed.

The long-run impact is obtained by considering an event time t_{++} long after time zero, i.e.

$$\mathsf{E}\left[Y_{it_{++}} - Y_{it_-}\right] = \mathsf{E}\left[F\left(k_{iT}, \mathbf{x}\left(k_{iT}, \mathbf{k_i}, \mathbf{z_{it_{++}}}\right), \mathbf{z_{it_{++}}}\right)\right] - \mathsf{E}\left[F\left(0, \mathbf{x}\left(0, \mathbf{k_i}, \mathbf{z_{it_-}}\right), \mathbf{z_{it_-}}\right)\right]. \tag{3}$$

There are two differences between this impact measure and the previous one. The first difference is that the long-run impact captures the effect of total lifetime fertility k_{iT} as opposed to the effect of only the first child. The second difference is that the smoothness assumption is no longer sufficient for identification as we can still have large changes in non-child earnings components over a long event time window. Hence, if we are not fully controlling for $\mathbf{z_{it}}$, then the long-run child penalty may be a biased estimate of the true post-child impact. Allowing for non-parametric age and year controls as we do in specification (1) in the main paper may go a long way in alleviating this

[1] On the other hand, if women are engaging in intertemporal substitution of work effort around the event of having a child, then the pre-child effect could be positive. However, our event graphs feature very stable pre-trends that are identical for men and women, indicating that no significant intertemporal substitution is taking place.

problem, but we cannot be certain that there is no remaining bias. There are two potential solutions to the problem. One is to use a control group — naturally women and men who never have children — to account for the non-child earnings trend in a difference-in-differences design. The other is to leverage an instrument for child birth within our event study approach. In the following sections we consider both approaches.

B.2 Identification Check: DD Event Study

In this section we lay out a difference-in-differences event study design that uses men and women who never have children as controls. The design is based on assigning placebo births to individuals who never have children, drawing from the observed distribution of age at first child among those who do have children (within cells of cohort and education).

Two technical issues arise when defining the control group and assigning placebo births. First, individuals observed without children include two types: those who will never have children and those who have not had children yet. The first group is the cleanest possible control group, but we face a truncation issue in identifying them. Taking age 40 as the latest age at which people have their first child (as only small fractions of both men and women have their first child after that age), the fertility of cohorts born after 1973 is truncated as they are younger than 40 when we last see them in 2013. Hence, for individuals observed without children in the later cohorts, we select those who are most likely never to have children based on a linear probability model of zero lifetime fertility ($k_{iT} = 0$) as a function of observables, estimated on the (non-truncated) cohorts born between 1955–1973.[2] In each of the later cohorts, the number of individuals we select as having zero lifetime fertility is such that the probability of never having children is the same after 1973 as the average between 1955–1973 (as this probability has been quite stable during this time).[3] Our control group consists of these selected individuals from the post-1973 cohorts along

[2]Specifically, we estimate the following model separately for men and women:

$$P[k_{iT} = 0] = X'\beta, \tag{4}$$

where $k_{iT} = 0$ is a dummy for zero lifetime fertility and X includes the following dummy controls: quartiles of the income distribution of the individual's cohort, quartiles of the wealth distribution of the individual's cohort, quartiles of the wealth distribution of the spouse's cohort, education length/degree (8 categories), decade of generation of the maternal grandmother, decade of generation of maternal grandfather, and region of residence. The model is estimated on the non-truncated cohorts from 1955-1973. We predict the probability of zero lifetime fertility for the truncated cohorts after 1973 as $\hat{P} = X'\hat{\beta}$.

[3]That is, for each cohort c born after 1973, we rank individuals based on their $\hat{P}$s. Among those from cohort c who have no children by 2013, we pick the n_c individuals with the highest $\hat{P}$s such that $\frac{n_c}{N_c} = P_{1955-1973}$, where N_c is the total number of individuals (men and women, respectively) in the cohort and $P_{1955-1973}$ is the average fraction of individuals with $k_{iT} = 0$ for cohorts 1955-1973. These n_c individuals are assumed to have zero lifetime fertility. The rest of cohort c observed with no children are assumed to have children later, and are therefore not included among the

with all individuals without children from earlier cohorts.

Second, we need to allocate placebo births to those in our control group. Here we also have to distinguish between truncated cohorts born after 1973 and non-truncated cohorts born before that time. For the older cohorts, the distribution of age at first child A is approximated by a log-normal distribution within cells of birth cohort c and education e. That is, we assume $A_{c,e} \sim \mathcal{LN}\left(\hat{\mu}_{c,e}, \hat{\sigma}^2_{c,e}\right)$ where the mean $\hat{\mu}_{c,e}$ and variance $\hat{\sigma}^2_{c,e}$ are obtained from the actual distributions within each cohort-education cell. Individuals in older cohorts without children get a random draw from this distribution. For the younger cohorts, we draw a random age at first child from $\mathcal{LN}\left(\tilde{\mu}_{c,e}, \hat{\sigma}^2_{c,e}\right)$ where the mean $\tilde{\mu}_{c,e}$ is the predicted average age at first child obtained by estimating a linear trend on the older cohorts. That is, consistent with the stylized pattern observed for the older cohorts, we allow for an upward linear drift in the age at first child while keeping the variance constant.

With this setup, we are able to implement event studies that compare our treatment group (a balanced panel of those who have their first child between 1985–2003 and are observed in a 15-year window around the first child birth) to a control group (a balanced panel of those who never have children, but have been assigned a placebo birth between 1985–2003 and are observed in a 15-year window around the placebo). The impact of children can be estimated as a difference-in-differences, i.e.

$$\mathsf{E}\left[Y_{i,t>0} - Y_{i,t<0} \mid k_{iT} > 0\right] - \mathsf{E}\left[Y_{i,t>0} - Y_{i,t<0} \mid k_{iT} = 0\right]. \tag{5}$$

The identification assumption is a standard parallel trends assumption, which in the notation established above implies $\mathsf{E}\left[\Delta F\left(0, \mathbf{x}\left(0, \mathbf{k_i}, \mathbf{z_{it}}\right), \mathbf{z_{it}}\right) \mid k_{iT} > 0\right] = \mathsf{E}\left[\Delta F\left(0, \mathbf{x}\left(0, \mathbf{0}, \mathbf{z_{it}}\right), \mathbf{z_{it}}\right) \mid k_{iT} = 0\right]$. Given the parallel trends assumption — the validity of which we can verify from the pre-trends — it is not necessary to introduce controls for $\mathbf{z_{it}}$. Therefore, we drop the age dummies in the specification discussed below.[4]

Figure A.VI shows the earnings impacts of children in this difference-in-differences design. Panel A shows women while Panel B shows men. The event studies are very sharp and confirm the key qualitative findings from the baseline specification. Women with children and women without children are on identical pre-trends, diverge sharply at the time of the first child birth, and the impact is very stable over time. The impact of children 10 years after equals 20.6%, slightly larger than the baseline impact of 19.4% shown in Figure 1A. The baseline graph already suggested that

controls.

[4]We do keep the year dummies (in order to show changes in real terms), but the child penalty estimates are virtually unchanged when dropping the year dummies as well.

we were slightly underestimating the career cost of children on women: it showed a weak upward pre-trend for women (compared to men) that the child penalty estimate did not take into account. Using women without children as a control group accounts for this trend.

A new insight that emerges from the DD event study design is that men *are* affected by parenthood, although the effect is very small. Men who have children and men who don't are on identical pre-trends and then diverge at event time 0. Even though the effect is tiny, the perfect pre-trends and precision of our data makes the effect very clear. The long-run child penalty in male earnings is equal to 3%.

B.3 Identification Check: IV Event Study

As another identification check, we compare our event study approach to an IV approach using the sex mix of the first two children as an instrument for having a third child (Angrist and Evans 1998). The idea is that parental preferences for variety make it more likely to have a third child when the first two have the same sex, while the children's sex should have no independent impact on labor market outcomes and thus satisfy the exclusion restriction.[5] As the sibling sex mix instrument gives the local average treatment effect of a third child, we have to modify our event study approach to also provide the local impact of the third child in order to compare the two approaches.

We consider the following event study specification for estimating the effect of a third child:

$$Y_{istt'} = \sum_{j \neq -1} \alpha_j \cdot \mathbf{I}\left[j = t\right] + \sum_k \beta_k \cdot \mathbf{I}\left[k = \text{age}_{is}\right] + \sum_y \gamma_y \cdot \mathbf{I}\left[y = s\right] + \sum_{n \neq -1} \delta_n \cdot \mathbf{I}\left[n = t^{'}\right] + \nu_{istt'} \quad (6)$$

where the index t still denotes event time with respect to the first child, while the new index $t^{'}$ denotes event time with respect to the third child. The first three terms on the right-hand side corresponds to our baseline specification: it gives the effect of the first child, controlling for a full set of age and year dummies. The fourth term on the right-hand side is new and it includes event time dummies around the birth of the third child (omitting $t^{'} = -1$). Even though the objective is to estimate the effect of the third child, we keep the event time dummies around the birth of the first child in the specification. This is because past child dynamics may matter for the impact of the

[5]The validity of this exclusion restriction can be verified based on our event study approach: we find no differences in the impacts of the first child depending on whether it is a boy or a girl, suggesting that child gender is not important for parental labor market outcomes. While the exclusion restriction is thus compelling, there might be a problem with the assumption of no defiers underlying LATE. Some parents may have preferences for a specific sex — typically boys — and are therefore less likely to have a third child if they start out with two boys. The presence of defiers due to boy-bias is arguably less of an issue in Denmark than in more traditional societies.

third child.[6]

We run the specification on a sample of women who had their first child between 1985-2003 (as before) and who have completed fertility of two or three when we last observe them in 2013.[7] In order to capture the long-run effect of a third child, we expand from the previously balanced panel that included a 10-year window after the first child birth to an unbalanced panel that includes the longest possible window for each individual. For example, a woman who had her first child in 1985 and is observed until 2013 is included in the sample with event times up to $t = 28$ with respect to that child, allowing for the longest possible period after the third child.

The IV-specification is the same as equation (6), but there we instrument the event time dummies around the third child birth using the sex mix of the first two children. Specifically, we instrument each dummy $\mathbf{I}\left[n = t'\right]$ by the interaction $\mathbf{I}\left[n = t'\right] \times \mathbf{I}\left[\text{same sex siblings}\right]$, which takes the value of one when the woman is at event time t' with respect to the third child and her first two children have the same sex. In contrast to previous implementations of such IV approaches, this specification traces out the full dynamic pattern of the effects of the third child.

The results are presented in Figure A.VII, which shows the earnings impacts of a third child obtained from the event study specification (black series) and the IV specification (red series) as a function of years since the birth of the third child. Analogous to the previous graphs, the statistic shown here is $P_{t'}^{w} \equiv \hat{\delta}_{t'}^{w} / \mathsf{E}\left[\tilde{Y}_{istt'}^{w} \mid t'\right]$ where $\tilde{Y}_{istt'}^{w}$ is the counterfactual outcome absent the effect of the third child, but not the other children, i.e. $\tilde{Y}_{istt'}^{w} \equiv \sum_j \hat{\alpha}_j \cdot \mathbf{I}\left[j = t\right] + \sum_k \hat{\beta}_k \cdot \mathbf{I}\left[k = \text{age}_{is}\right] + \sum_y \hat{\gamma}_y \cdot \mathbf{I}\left[y = s\right] + \hat{\eta}_i$.

The following key insights emerge from the figure. First and foremost, the event study estimates and the IV estimates are almost perfectly aligned through event time, providing strong support for our empirical approach. The fact that the pre-event coefficients are very similar (and close to zero) for the two approaches suggests that anticipation effects of children are not important. The fact that the post-event coefficients are very similar shows that the impacts on women who have a third child because of the sex mix of the first two (the IV-compliers) are the same as the impacts on all treated women. Second, the short-run effect of a third child is similar to the short-run effect of the first child, an earnings reduction of 20-30%. Third, the long-run effect of a third child is about 5%. This is less than the long-run effect of the first child among those who only have one child (equal

[6] The reason why we can separately identify event time dummies for the first child and the third child, while simultaneously controlling for a full set of age and year dummies, is that there is enough independent variation in the ages at which women have their first and third child, respectively.

[7] As we do not include men here, we have dropped the superscript g in specification (6).

to 7.1% as shown in Figure A.III), suggesting that the marginal effect of children is declining in the number of children.

Besides sibling sex mix, a number of studies have considered the occurrence of twins at first birth (e.g. Rosenzweig and Wolpin 1980; Bronars and Grogger 1994) or at second birth (e.g. Angrist and Evans 1998) as instruments for children. As an alternative to the strategies discussed above, we may consider twins at second birth as an instrument for the third child. Figure A.VII therefore compares the average impacts obtained from the event study and same-sex IV to the average impact obtained from a twin IV. We see that the twin estimate is considerably smaller than the event study and same-sex estimates. A natural interpretation of this difference is that twins represent a more efficient child production technology (e.g. due to economies of scale) and therefore impose smaller penalties on women. This implies that even though the occurrence of twin births is an exogenous event, it is not a valid instrument for having an extra child in the standard sequential way: it does not satisfy the exclusion restriction if it changes aspects of the child care technology that have their own direct impact on earnings.[8] Our results are consistent with those of Angrist and Evans (1998), who also find smaller effects when using the twin instrument than when using the same-sex instrument.[9]

References

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review*, 88: 450–477.

Bronars, Stephen G., and Jeff Grogger. 1994. "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment." *American Economic Review*, 84: 1141–1156.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 1980. "Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models." *Journal of Political Economy*, 88: 328–348.

[8] To be clear, the exogeneity of twins implies that it does give the causal effect of a twin birth relative to a singular birth (the reduced-form impact of a twin birth dummy), but it does not give the effect of increasing the number of children through the standard sequential birth technology.

[9] Angrist and Evans (1998) argue that, besides economies of scale, the smaller size of the twin estimates could be driven by differences in the age of the third child: in cross-sectional comparisons, a third child triggered by twins will be older than a third child triggered by same sex siblings. Our dynamic IV approach allows us to separate the age and economies-of-scale hypotheses by estimating the impact of the third child at each event time after birth. Such an exercise reveals that the twin estimates are smaller conditional on event time, which cannot be explained by age and suggests that economies of scale are important.