Causal Inference Report: Review on Child Penalties Across Countries: Evidence and Explanations

Juan Pablo Brasdefer Fabian Pawelczyk

2023-05-24

Contents

1 Abstract

The paper we analyzed examines the influence of having children on the gross labor earnings of both mothers and fathers (the estimand). Specifically, it delves into the *child penalty* - a phenomenon referring to the potential decrease in earnings experienced by individuals, particularly women, upon becoming parents. This 'penalty' essentially represents the earnings disparity between those with children and those without.

The authors aim to contribute to the extensive literature on gender inequality by applying an *event study methodology*, with the birth of the first child serving as the central event. Utilizing a rich panel dataset that provides information on labor market outcomes and children, the authors navigate through the complex dynamics of parenthood and earnings, aiming to shed light on how gender inequality in earnings might be perpetuated by the advent of parenthood.

2 Basic DAGs to illustrate the causal pipeline

In Figure 1, we introduce a fundamental Directed Acyclic Graph (DAG), which illustrates the estimand in the simplest manner. However, it's essential to acknowledge that this model offers a simplified representation in two ways:

a) The use of panel data introduces a temporal component to the data, which isn't explicitly depicted in this model.

b) The authors also clarify that earnings penalties may stem from three distinct labor supply margins: the extensive margin (employment, abbreviated as `Emp`), the intensive margin (hours worked, abbreviated as `HW`), and the wage rate (denoted by `w`).

To better represent the nuanced composition of earnings, we incorporate these components into a new DAG (Figure 2). This adaptation provides a clearer visualization of the multiple elements contributing to earnings.

3 Data

The study works as a form of Natural Experiment for a few key reasons, though the authors highlight the fact that *the ideal experiment for the estimand would be to randomize fertility.*

First, there is exogenous assignment of treatment, meaning that assignment is not subject to manipulation. Here it is important to note that the authors develop counterfactuals statistically to some extent instead of as natural data points. This is further explored in the Identification section later on. Second, data collection for purposes of the survey takes place after the data itself has been generated, meaning that there is no opportunity for the researchers to influence the true data or reported response of individuals. Third, bias is partially reduced by the pool of data. Something could be said about participation bias when it comes to those who do not answer surveys, but the authors do not see this as a hindrance, given a very big sample size and frequent (panel) data.

Data for Austria, Denmark, and Sweden is taken from administrative registers that hold data for the full population while Germany, the US, and the UK use surveys with sufficiently large sample sizes and long time series. For the remainder of the Methods, and Identification sections, it is important to note that this paper bases its approach on a previous paper by Kleven et al. from 2018. Data sources are dissimilar, but the statistical formulations and approaches to bias reduction and identification are the same.

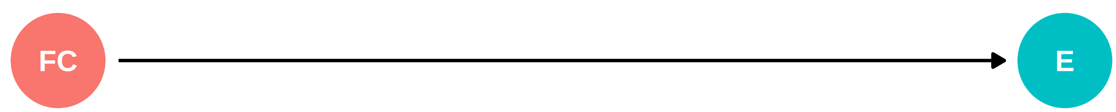

Figure 1: DAG showing the simplified relationship between having a First Child (FC) and Earnings (E). This model is a simplification, and the effects of additional children are also captured in the analysis.

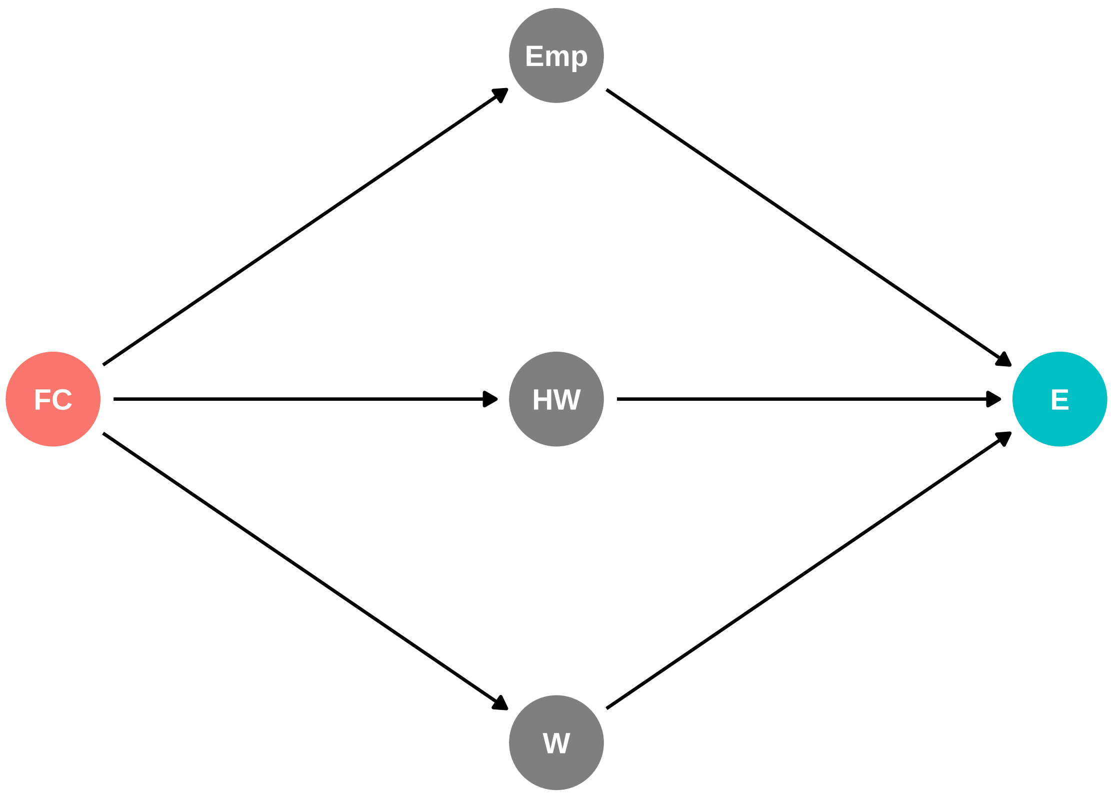

Figure 2: DAG showing the hypothesized relationships between having a First Child (FC), Employment (Emp), Hours Worked (HW), Wage (W), and Earnings (E). This model is a simplification, and the effects of additional children and other potential confounding factors are also captured in the analysis.

4 Methods

4.1 Event Study Framework

The paper follows an Event Study Framework, meaning that the focus of the study is to note changes in variables before and after a precise event. With regards to choosing Event Study as the study design, the authors recognize that, *although fertility choices are not exogenous*, the effects of having a first child create changes in labour market outcomes that are *arguably orthogonal to unobserved determinants of those outcomes*. Said differently, we expect the unobserved determinants to evolve smoothly over time, regardless of birth.Thus, this event study relies on the Smoothness Assumption (especially for identifying the short-run effects)

The Event Window extends backwards 5 years and forward 10 years around the birth of the first child, with the birth being $t = 0$. Thus, $t = (-5, -4...0...9, 10)$ denotes the distance from the birth of each individual's *(i)* first child, which is of the births of other children. One thing to note is that $t = -1$ is not included in measurements, as it incorporates the effects of pregnancy on a woman that men do not experience.

4.2 Baseline Regression Model

Figure 3 represents a simple understanding of the potential relationships at play in this study. Specifically, it highlights the potential for confounding by age (LCT) and calendar year (TT).

In order to circumvent these confounders in the baseline model, the authors create the regression equation below, where the outcome variable Y ('Earnings') for individual i in calendar year s with relative event-time t (-5,. . . -10). The regression is calculated separately for men and women, denoted by superscript g.

$$Y^g_{\text{ist}} = \sum_{j \neq -1} \alpha^g_j \cdot I[j = t] + \sum_k \beta^g_k \cdot I[k = \text{age}_{is}] + \sum_y \gamma^g_y \cdot I[y = s] + \nu^g_{\text{ist}}$$

The equation has three dummy terms (functioning through Indicators). The first dummy term deals with the 'event-time', which is to say that it captures the coefficient for distance from the birth of the child. The second dummy is the 'age' dummy, and helps us control for underlying life-cycle trends that occur at different points of an individual's life, such as entering the workforce or leaving it. Age is also important to include because it allows us to control for the fact that men and women often have children at different ages. The final dummy is the 'year' dummy, which helps us control for time trends such as economic recessions, inflation, and world events that may interfere with our outcome.

Given that we include 'age' and 'year' in the model, the DAG now looks like the following, showing that we break the confounding back-door paths.

A quick note: if the age and year dummies were not included, the estimated event coefficients

$$\hat{\alpha}$$

would correspond to the mean value of the outcome at event time, relative to $t = -1$. It is also important to note that both 'age' and 'year' are included nonparametrically, meaning that the authors do not assume a specific functional form or distribution of their variables. This allows for more flexible modelling of the relationship with each other and the estimand.

4.3 Child Penalty Percentage

The results of our baseline regression (the estimated level effects Ym and Yw) are then converted into percentages for later comparison via the following equation:

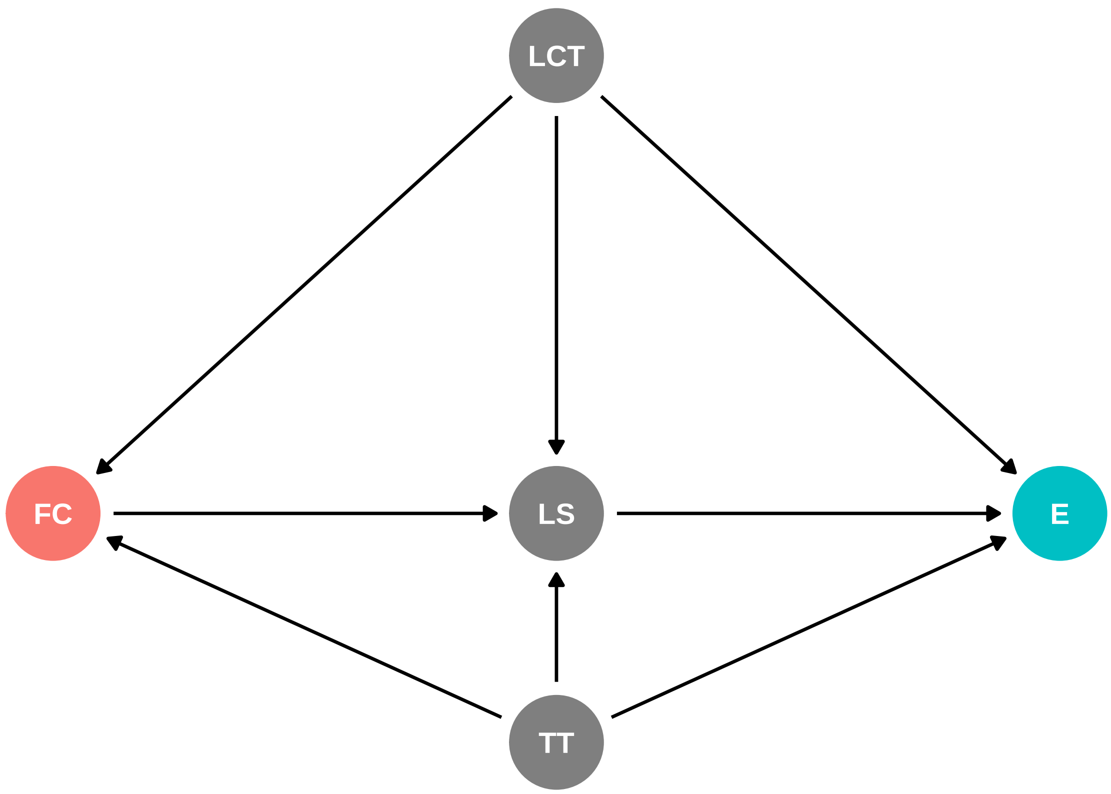

Figure 3: DAG showing the hypothesized relationships between having a First Child (FC), Labor Supply (LS), Life-Cycle Trends (LCT), Time Trends (TT), and Earnings (E). LS represents the combined effect of employment status, hours worked, and wage. LCT and TT are confounders that are controlled for in the analysis. This model is a simplification, and the effects of additional children are also captured in the analysis.

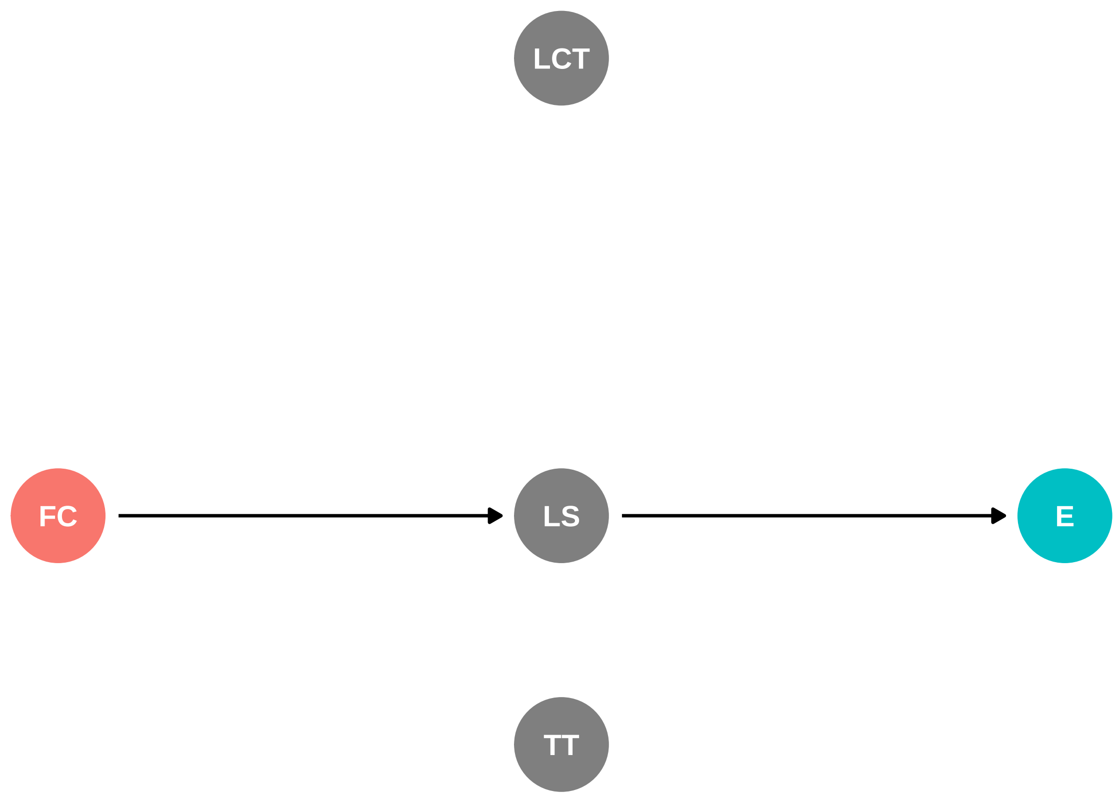

Figure 4: This DAG outlines the causal relationships between the birth of the first child (FC), labor supply (LS), and earnings (E). The labor supply (LS) is a mediator variable that encapsulates the combined influence of employment status, hours worked, and wage rate. Notably, the model controls for Time Trends (TT) and Life-Cycle Trends (LCT), which are recognized confounders in this context. In this simplified representation, no direct links are drawn from TT and LCT to the other variables, reflecting the control for these confounders in the statistical model.

$$P_t^g \equiv \frac{\hat{\alpha}_t^g}{E[\tilde{Y}_{ist}^g|t]}$$

Where $\tilde{Y}^g$ (in the denominator) is the predicted outcome of our baseline regression when we omit the dummy responsible for the child-birth event. This is crucial, and is effectively a counterfactual regression that looks like:

$$\tilde{Y}_{\text{ist}}^g = \sum_k \hat{\beta}_k^g \cdot I[k = \text{age}_{is}] + \sum_y \hat{\gamma}_y^g \cdot I[y = s]$$

Given this counterfactual, Pgt captures the year-t effect of children as a percentage of the counterfactual outcome, absent children.

Finally, an equation to determine the Child Penalty on women, relative to men, at event time t is given by:

$$P_t \equiv \frac{\hat{\alpha}_t^m - \hat{\alpha}_t^w}{E[\tilde{Y}_{ist}^w|t]}$$

Where the alpha coefficients correspond to the coefficients of the baseline-men and baseline-women models respectively and the denominator corresponds to the women's counterfactual regression model absent-children.

5 Identification Strategy

5.1 Conceptual Framework

Beyond the baseline model, the authors do the responsible thing and explore the degree to which their study and assumptions are causally correct. They begin with establishing a conceptual framework, which, parallel to the causal relationships outlined in the DAGs above, is a more general view of how earnings fluctuations are constructed.

They beging this framework by establishing some precision around the idea of the 'event' and how it relates to the total number of children that an individual will have over their lifetime. The authors assign a variable that corresponds to the number of children 'k' for individual *i*. This *ki*, is named the "anticipated lifetime fertility path" (ALFP) for one person, and is a set of values beginning with 0 and ending with the total number *kiT* at the end of a person's life *'T'*, where the value of *kit* at any point t is the number of children of the individual at that time.

$$k_i = (0, k_{it}, ...k_{iT})$$

An important thing to note here is that the authors DO NOT discriminate individuals based on the number of children had after the first. In other words, there are conceivably data points with event-time 't' greater than 1 that have underlying child-number values also great than 1. This means that the medium-to-long-run penalties (P5,. . . P10) also include the effect of having more children beyond the firstborn, and so have the potential of capturing the total effect of children on gender-based earning inequalities.

Using this ALFP, the authors then conceptualize Earnings 'Y' as a function of three factors: the ALFP first, and two sets of "earnings determinants", the first of which (x) are chosen based on the presence of children (eg: hours worked, industry) and the second of which (z) are independent of children (eg: age of individual, education. . .).

$$Y_{it} = F(k_{it}, x_{it}, z_{it})$$

Given that earning determinants 'xit' are chosen based on the presence (or absence) of a child, the authors expand this formulation to reflect x's dependency on: number of children at time *t*, ALFP, and the non-child determinants 'z'. By including ALFP (entire path of past and future fertility) in this formulation for 'xit', the equation includes 'pre-child' effects such as women potentially choosing to invest less in education knowing they plan to have children in some future. This cannot be captured explicitly, however, as the event study is designed to only capture 'post-child' effects.

$$x_{it} = x(k_{it}, k_i, z_{it})$$

The post-child effect thus works *through* the *kit* variable directly *AND* via the *kit* variable present in *xit*. With this, the final conceptual framework can be written mathematically as:

$$Y_{it} = F(k_{it}, x(k_{it}, k_i, z_{it}), z_{it})$$

Which we use in order to create our Short-Run ($kit = 1$) Post-Child Impacts, where times t+ and t- indicate post-event and pre-event, and corresponding z-preferences adjust accordingly:

$$E[Y_{it+} - Y_{it-}] = E[F(1, x(1, k_i, z_{it+}), z_{it+})] - E[F(0, x(0, k_i, z_{it-}), z_{it-})]$$

And Long-Run Post-Child Impacts, where time t++ indicates long-run post-event times:

$$E[Y_{it++} - Y_{it-}] = E[F(k_{iT}, x(1, k_{iT}, z_{it++}), z_{it++})] - E[F(0, x(0, k_i, z_{it-}), z_{it-})]$$

Once again, it's good to note that this long-run impact captures the effect of total lifetime fertility *kiT* as opposed to the effect of only the first child. As well, it should be obvious that the smoothness assumption is no longer sufficient for identification, as we begin to have large changes in the *zit++* component with a long enough time window. Thus, if we are not fully controlling for *zit*, the long-run child penalty may be a biased estimate of the true post-child impact. This means that longer term penalties such as P20 require much stronger identification techniques such as a defined control group.

For this, the authors decide to pursue a Difference in Differences (DiD) and Instrumental Variable strategy.

5.2 Difference in Differences

For the (DiD) component, which requires a control group in order to be able to locate the 'Differences', the authors designate 'men a and women that never have children in their lifetimes' ($kit = 0$) as the control group.

To do this, they first classify those aged 40 and older as $kit = 0$, arguing that a negigible amount of people over this age go on to have children. Then, since the data only runs from 1955-2013, those in the 1973-2013 bracket (ie: sub-40 at any point in the dataset) are assigned an estimated value of being $kit = 0$ across the rest of their lives. The authors do this through a simple (but useful) Linear Probability Model,

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

where the probability is a function of X', a conjuction of 7 variables including education level, region of residence, generation of grandparents... that help determine the probability of never having any children. It is then trained on the 1955-73 cohorts, and used to assign values of $kit = 0$ for individuals in the 1973+ cohorts. The finalized control group is thus a mix of "Non-Truncated" (1955-73) true non-parents and "Truncated" (1973+) probable non-parents.

Next, to 'Allocate Placebo Births' to the control group (and thus create a group for comparison in differences), the authors distinguish once more between the Non-Truncated and Truncated groups. For the older, truncated groups, the distribution of Age at First Child, denoted 'A', is approximated by a log normal distribution calculated from the combinations of Birth Cohorts (c) and Education Cohorts (e). Mathematically, this is represented as:

$$A_{c,e} \sim \mathrm{LN}(\hat{\mu}_{c,e}, \hat{\sigma}^2_{c,e})$$

Where the mean and variance are taken from the actual groups of birth/education combinations. Individuals in this 1955-73 Non-Truncated group are then given a random draw from the distribution to decide at what age they are 'allocated' their placebo first child.

As for the younger, Truncated group, the placebo is allocated by drawing at random from:

$$A_{c,e} \sim \mathrm{LN}(\tilde{\mu}_{c,e}, \hat{\sigma}^2_{c,e})$$

Where the mean is now a predicted average that is born from a contunuation-of-trend from the older cohorts. Said differently: the authors allow the pattern that pervades through the older cohorts (such as, perhaps, people nowadays becoming parents later in life) to continue to an "upward linear drift" (while holding variance constant). Pretty cool stuff.

Finally, they are able to carry out event studies that compare a treatment group (those who have their first child between 1985–2003 and are observed in a 15-year window around the event) to a control group (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 event).

$$E[Y_{i,t>0} - Y_{i,t<0} | k_{iT} > 0] - E[Y_{i,t>0} - Y_{i,t<0} | k_{iT} = 0]$$

Said in plain English, the above difference-in-differences yields an estimate of the impact of children as:

[The expected value of (Earnings post-birth) minus (Earnings pre-birth) for those who had at least one child] minus [The expected value of (Earnings post-birth) minus (Earnings pre-birth) for those who had no children throughout their lifetime but have been assigned a placebo]

5.3 Instrumental Variable Event Study

To verify whether the authors' identification strategy is valid, they run two main identification checks and therefore address robustntess/ potential bias.

In this identification check, the authors propose a variation of their event study approach by introducing an Instrumental Variable (IV) strategy. They use the gender composition of the first two children as an instrument for having a third child, following the reasoning that parents with two same-sex children are more likely to have a third child. The validity of this instrument relies on the assumption that the sex of the first two children doesn't independently impact labor market outcomes (the exclusion restriction).

The authors modify their event study specification to estimate the Local Average Treatment Effect (LATE) of having a third child in order to compare it with the IV approach. This means they are focusing on the sub-population influenced by the instrument (the "compliers"). For the interested reader we created a DAG to visualize this IV approach (see Figure 5).

To facilitate a valid comparison, the authors modify their event study model as follows:

$$Y_{\mathrm{istt'}} = \sum_{j \neq -1} \alpha_j \cdot I[j = t] + \sum_k \beta_k \cdot I[k = \mathrm{age}_{is}] + \sum_y \gamma_y \cdot I[y = s] + \sum_{n \neq -1} \delta_n \cdot I[j = t'] + \nu_{\mathrm{istt'}}$$

In this model, t' represents the event time with respect to the third child, while the new term $\sum \delta_n ů I[n = t']$ introduces event time dummies for the birth of the third child. The authors also keep the event time dummies for the first child, as previous childbearing dynamics may impact the effect of the third child.

The IV specification is identical, except it includes an additional instrumenting strategy which uses the event time dummies around the third child birth, incorporating the sex mix of the first two children:

$$I[n = t'] = I[n = t'] * I[samesexsiblings]$$

In other words, $I[n = t']$ takes the value of one when the woman is at event time t' with respect to the third child and her first two children are of the same sex.

Following the logic from the baseline model, the counterfactual outcome (denoted $\tilde{Y}^w_{istt'}$), is calculated excluding the effect of the third child but including the effects of the first child and other controls. It represents the expected labour market outcome for a woman if she did not have a third child, providing a baseline against which to compare the actual outcomes of those who did have a third child.

The findings suggest that the event study and IV estimates align closely, indicating robustness in their original empirical approach. The short-run effect of the third child is similar to that of the first, showing an earnings reduction of 20-30%. However, the long-run effect of a third child is about 5%, lower than the long-run effect of the first child for those who only have one child, suggesting a diminishing marginal effect with the addition of more children.

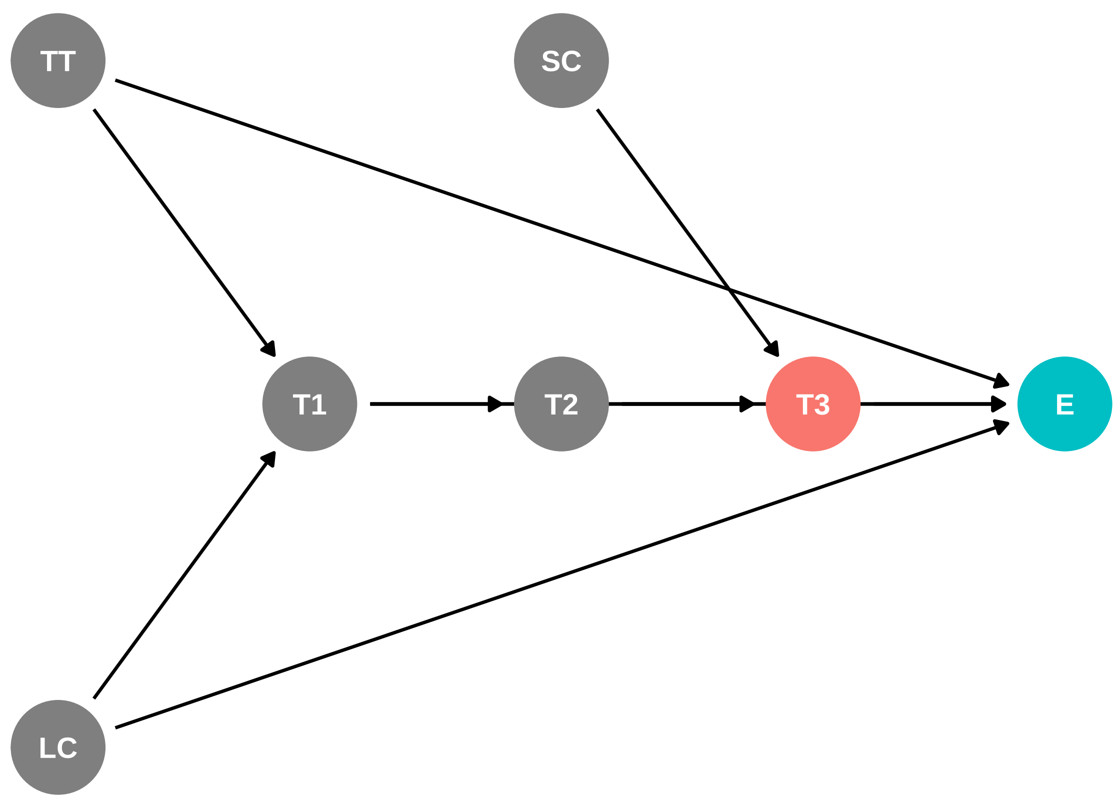

Figure 5: Note: This DAG represents a way to think about the aforementioned IV approach. T3 denotes the timing of the third child that is instrumentalized through the sex composition (SC). Also, to get rid of the confunders LC and TT the authors are controlling for it. E - Earnings, T1 - First Child, T2 - Second Child, T3 - Third Child, SC - Sex Composition, TT - Time Trends, LC - Life Cycle Trends

6 Critiques

While the authors are able to identify a causal effect and have a rather strong methodological foundation there are still some things one can discuss.

6.1 Comparision among countries and different sources of earnings

A criticism, not entirely attributable to the measurement process itself but nonetheless significant when contrasting the metrics across different countries, pertains to variations in parental leave benefits. As the authors' selected earnings measure incorporates any non-statutory parental leave benefits, there is a possibility that cross-country comparisons may reflect variations in policies such as those related to parental leave benefits. To account for this, we propose an expansion of the initial DAG to include a new pathway incorporating `earnings` in the model (see Figure 6).

It's important to note this is a simplification. It's plausible that parental leave benefits could directly influence the decision to have children by reducing the financial stress associated with parenthood. Therefore, adjusting for the extent of such benefits would be advisable.

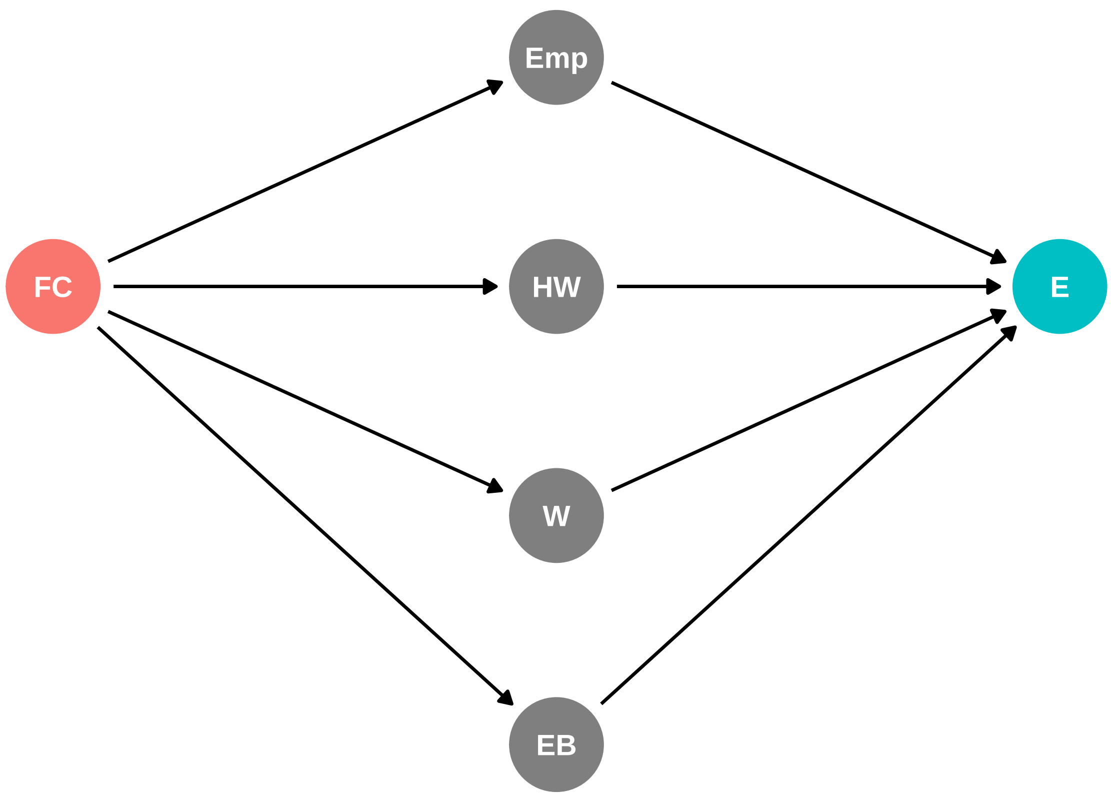

Figure 6: DAG demonstrating the hypothesized relationships between the birth of the First Child (FC), Employment (Emp), Hours Worked (HW), Wage (W), Employer Benefits (EB), and Earnings (E). Note: This model is a simplification, and the effects of additional children and other potential confounding factors (TT and LC) are also captured in the analysis.

6.2 Causal Effect vs Causal Mechanism: Choices, Policies, Culture

While the authors deliver rather strong evidence on the existence of a causal effect, the underlying mechanism is still unclear. The name itself, "child *penalty*" suggests punishment towards mothers and reduces the earnings drop to exactly this: a financial drop. Butm what if mothers decide to take jobs with more flexibility in order to raise their children? This would increase their net utility even if the wage is lower. To better understand the mechanism it would be good to extend the research here and see how women navigate the employment landscape after bearing children. Is it towards a family-friendly sector with a lower wage, or is the industry from pre-event maintained but simply with a lower (pre-inflation) wage?

Another interesting critique to explore is that of government policies. One explanation for the differences in the penalties could have to do with government policies. While numerous studies have been conducted to evaluate the effect on gender gaps and female labor supply by public policies such as taxes, transfers, parental leave, and child care provisions, these policies do not adequately explain the "child penalties" experienced by mothers in the workforce.

As the authors hint towards in the paper, another potentially fruitful area to research would be a look into gender-norms and broader cultural differences across given societies, as these are far more difficult to capture that an explicit government policy, but are arguably more influential in determining te magnitude of a "child penalty".

6.3 Controls missing

As we saw multiple times, the authors' analysis does not control for those parents that only have one child, instead providing only the lower bound in an inequality. As a result, the long-term child penalties that we observe include the impacts of having additional children, and therefore, are influenced by total fertility. To overcome this in the short-term, we would suggest to run an additional model with a `only-one-child` variable.