NBER WORKING PAPER SERIES

THE GEOGRAPHY OF CHILD PENALTIES AND GENDER NORMS: A PSEUDO-EVENT STUDY APPROACH

Henrik Kleven

Working Paper 30176
http://www.nber.org/papers/w30176

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2022, revised August 2025

I thank Charles Brown, Raj Chetty, Amy Finkelstein, John Friedman, Peter Ganong, Pat Kline, Ilyana Kuziemko, Camille Landais, Ale Marchetti-Bowick, Gabriel Leite Mariante, Petra Moser, Isaac Sorkin, Christopher Walters, Owen Zidar, and two anonymous referees for comments. I also thank Valentina Andrade, Eva Demsky, Niklas Hein, Ragini Jain, Madhavi Jha, and Aarti Malik for outstanding research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The Geography of Child Penalties and Gender Norms: A Pseudo-Event Study Approach
Henrik Kleven
NBER Working Paper No. 30176
June 2022, revised August 2025
JEL No. J13, J16, J21, J22, J61

ABSTRACT

This paper develops a pseudo-event study approach to estimating child penalties using cross-sectional data. The approach produces estimates that align closely with those obtained from true event studies using panel data and is statistically more precise. This allows for providing more granular evidence and study mechanisms. The paper presents a detailed investigation of the variation in child penalties across time, geography, and demographic/cultural groups in the US. There is large variation in these dimensions. Using a variety of approaches — including epidemiological studies of movers and immigrants— the paper finds that gender norms have sizable effects on child penalties.

Henrik Kleven
Princeton University
Department of Economics
and CEPR
and also NBER
kleven@princeton.edu

1 Introduction

A recent literature on gender inequality highlights the importance of child penalties: the effects of parenthood on women relative to men. In developed countries, child penalties account for most of the remaining gender inequality in the labor market (Kleven, Landais and Søgaard 2019; Kleven *et al.* 2019; Cortés and Pan 2023). A crucial question is why child penalties are so large even in modern societies? Fundamentally, this amounts to asking what explains the persistence of the traditional homemaker-breadwinner institution. This paper contributes methodologically and empirically to this question.

Research on the mechanisms behind child penalties is still in its infancy. The existing evidence casts doubt on explanations rooted in biology or comparative advantage (Kleven, Landais and Søgaard 2021) and suggests that the effects of government policy are modest (Kleven *et al.* 2024). Yet we lack conclusive evidence on the core drivers of these penalties. A key reason is the data-intensive nature of standard estimation methods, which rely on high-quality panel data and event study designs around childbirth. As a result, estimates are available for fewer than a dozen countries, and there is hardly any evidence on how child penalties vary across space and time within countries. To address this gap, this paper develops a new method to estimate child penalties using widely available cross-sectional data.

The first part of the paper develops the cross-sectional approach to estimating child penalties using data from the Current Population Survey (CPS 1968-2020) and the American Community Survey (ACS 2000-2019). The objective is to provide event studies around first childbirth, indexed as event time $\tau = 0$. The main challenge of using cross-sectional data is that negative event times are unobserved. That is, the data does not reveal if and when those observed without children will eventually have a child. To circumvent this problem, I use a simple matching algorithm to create a pseudo-panel: each person observed at event time $\tau = 0$ is matched to a childless person n years younger n years before and with the same demographic characteristics to obtain a synthetic observation for $\tau = -n$. With this pseudo-panel, the event study specification of Kleven, Landais and Søgaard (2019) can be implemented.

To clarify the assumptions for causal identification, I develop a potential outcomes framework for both event studies and pseudo-event studies of child penalties. Besides parallel trends and no anticipatory effects, the pseudo-event study approach relies on a conditional independence assumption. This assumption allows researchers to convert cross-sectional data into a synthetic panel and conduct event studies. The assumption can be directly tested in panel data. To validate the approach, I use data from the Panel Study of Income Dynamics (PSID 1968-2019) and the National Longitudinal Survey of Youth (NLSY 1979-2018). The results from pseudo-event studies and true event studies are virtually identical — in the full sample and in subsamples — but the cross-sectional approach is much more precise due to superior sample size.[1]

Considering the US as a whole, child penalties are currently equal to 20% in annual employment, 24% in weekly employment, and 31% in earnings. These penalties account for most of the remaining gender inequality in labor market outcomes. Similar estimates exist in the literature (Kleven *et al.* 2019; Cortés and Pan 2023), but the methodology developed here greatly expands the range of questions that can be studied. Because of its minimal data requirements and higher statistical precision, it allows for granular analyses of heterogeneity and mechanisms.[2]

The paper presents four main findings on the variation in child penalties. First, child penalties have fallen substantially over the last five decades. They were extremely high in the 1970s — 46% in annual employment and 70% in earnings — but have declined by more than half since then. Importantly, almost all of this decline occurred prior to the mid-1990s, followed by a long period of stagnation. This sheds light on a stylized fact documented elsewhere in the literature: the slowdown of gender convergence in labor market outcomes since the 1990s (Blau and Kahn 2006, 2017; Kuziemko *et al.* 2018). The evidence presented here points to a simple explanation: gender convergence stalled in large part because the decline in child penalties stalled.

Second, child penalties vary enormously across geography. The annual employment penalty ranges from 12% in the Dakotas (rural states with Scandinavian heritage) to 38% in Utah (a religiously and culturally conservative state). The earnings penalty ranges from 21% in Vermont, another rural state, to 61% in Utah. Overall, the US map of child penalties highlights two potential mechanisms: urbanization and cultural norms. More urban places tend to have larger penalties, likely because urban jobs offer less flexibility than rural jobs.[3] More culturally conservative places also tend to have larger penalties, but many conservative places are relatively rural and this pulls

[1] The equivalence of the two approaches is confirmed both through visual inspection of the event studies and through a formal specification test.

[2] The approach also allows for studying child penalties in low- and middle-income countries where evidence has been scarce. In follow-up work, Kleven, Landais and Leite-Mariante (2024) use the approach to construct a global atlas of child penalties.

[3] Job flexibility has been shown to be important for gender gaps (see e.g., Goldin 2014; Goldin and Katz 2016).

in the opposite direction. An example is the Bible Belt in the American South. The remainder of the paper delves into the effect of gender norms and culture on child penalties, addressing the confounding effects of urbanization and other factors.[4]

Third, the relationship between child penalties and gender norms is analyzed using General Social Survey data (GSS 1972-2018). The analysis constructs an index of gender progressivity using survey questions regarding gender roles in families with children. Gender progressivity has increased substantially over time, but most of this increase occurred prior to the mid-1990s. As a result, the time series in gender progressivity is a mirror image of the time series in child penalties. Gender progressivity also varies substantially across geography. States in the Bible Belt and Utah are among the most conservative, while states in the Northern Midwest and New England are among the most progressive. An analysis using both time and spatial variation suggests that gender norms have a strong influence on child penalties. An increase in the gender progressivity index of one standard deviation reduces the child penalty in annual employment by 18pp, and the child penalty in weekly employment and earnings by 23pp.

Finally, the paper provides an epidemiological study of gender norms using US-born movers and foreign-born immigrants.[5] This analysis provides striking graphical evidence, leveraging the enormous variation in child penalties across states in the US and countries around the world. The child penalty for US movers is strongly related to the child penalty in their state of birth, controlling for selection in their state of residence. Parents born in high-penalty states (such as Utah or Idaho) have much larger child penalties than those born in low-penalty states (such as the Dakotas or Hawaii). The effect is quantitatively large: a 10pp increase in the employment penalty in a woman's state of birth translates into an increase in her employment penalty of about 7pp. Similarly, the child penalty for foreign immigrants is strongly related to the child penalty in their country of birth. Immigrants born in high-penalty countries (such as Bangladesh, Mexico, or Switzerland) have much larger child penalties than immigrants born in low-penalty countries (such as China, Cuba, or Portugal).[6] The magnitude of this effect is also large: a 10pp increase in the employment penalty in a woman's country of birth translates into an increase

[4] The paper provides evidence on heterogeneity in other dimensions than geography. There is virtually no heterogeneity in child penalties by female education level, which suggests against specialization based on comparative advantage (see also Kleven, Landais and Søgaard 2021). Conversely, there is lots of heterogeneity by marital status (much larger child penalties on married women than on single women) and by race (much larger child penalties on white women than on black women).

[5] See Fernández (2011) for a review of the epidemiological approach to studying norms and culture.

[6] The country-of-birth child penalties used in this epidemiological study come from Kleven, Landais and Leite-Mariante (2024).

in her employment penalty of about 5pp. These results are consistent with important effects of childhood culture on child penalties. It is worth noting that, even if there is bias in the pseudo-event study estimates of individual child penalties, explaining the epidemiological results by violations of the empirical design requires a convoluted selection story.

This paper contributes to a large literature on gender inequality, reviewed by Altonji and Blank (1999) and Bertrand (2011). It relates most directly to a burgeoning literature studying the impact of childbirth on gender gaps, including Angelov, Johansson and Lindahl (2016), Kleven, Landais and Søgaard (2019), Kleven *et al.* (2019), Kleven, Landais and Søgaard (2021), Kleven *et al.* (2024), Andresen and Nix (2022), and Cortés and Pan (2023). These papers provide event study evidence on child penalties using panel data — often administrative data from Scandinavian countries — and the empirical framework has been validated using instruments for fertility from sibling sex mix (Kleven, Landais and Søgaard 2019), IUD failure (Gallen *et al.* 2023), and IVF treatment (Lundborg, Plug and Rasmussen 2017).[7] While this research has produced valuable insights, our understanding of generalizability, heterogeneity, and mechanisms has been hampered by the data-demanding approach used to estimate child penalties.

I advance the literature in two directions. The first advance is to develop a pseudo-event study approach based on cross-sectional data, validating it against a true event study approach based on panel data. The approach is related to the synthetic-cohort approach developed by Deaton (1985), but it uses a granular matching algorithm to assign synthetic event times among untreated individuals and conduct staggered event studies.[8] Given the availability of large cross-sectional datasets with information on labor market outcomes and children, the approach allows for estimating child penalties across most countries of the world and over the long run of history (Kleven, Landais and Leite-Mariante 2024). Beyond the study of child penalties, the pseudo-event study approach may be applicable to other settings where panel data is unavailable.

The second advance is to provide granular evidence on child penalties across time, geography, and demographic/cultural groups. The paper documents large variation in these dimensions and provides striking evidence on the explanatory power of gender norms. These findings relate to an existing literature estimating the effects of social norms on female labor supply (e.g., Fernán-

[7] The IUD instrument is particularly compelling in terms of exogeneity and the exclusion restriction. It replicates the child penalty estimates obtained from event studies almost exactly (Gallen *et al.* 2023).

[8] We show that the estimates are not biased by treatment-effect heterogeneity, a concern raised in the recent econometrics literature on staggered difference-in-differences and event study designs (see e.g., de Chaisemartin and D'Haultfœuille 2020; Sun and Abraham 2021; Goodman-Bacon 2021; Callaway and Sant'Anna 2021; Roth *et al.* 2023; Borusyak, Jaravel and Spiess 2024).

dez, Fogli and Olivetti 2004; Fortin 2005; Fernández and Fogli 2009; Blau, Kahn and Papps 2011; Bertrand 2020).[9] The epidemiological study of US movers overlaps with two recent studies using mover designs: Charles, Guryan and Pan (2022) estimate the effect of sexism on female labor market outcomes using within-US movers, and Boelmann, Raute and Schönberg (2023) estimate the effect of culture on maternal employment using movers between East and West Germany. The mover analysis presented here has a different focus — understanding what drives child penalties — and relies on sharp event studies of childbirth at a granular geographic level. Even stronger evidence on the effect of gender norms is provided by the epidemiological study of foreign immigrants. This analysis is based on event studies of childbirth among US immigrants from 81 diverse countries, featuring source-country employment penalties ranging from 0% to 64%. These analyses are feasible only because of the pseudo-event study approach.

The paper is organized as follows. Section 2 describes the data. Section 3 develops the empirical methodology. Section 4 provides baseline estimates and validates the approach. Section 5 presents evidence on the variation in child penalties across time, geography, and demographic groups. Section 6 investigates the effect of gender norms on child penalties using difference-in-differences and epidemiological approaches. Section 7 concludes.

2 Data

The pseudo-event study approach developed below is implemented using pooled data from the Current Population Survey between 1968-2020 (CPS 1968-2020) and the American Community Survey between 2000-2019 (ACS 2000-2019). The CPS component includes data from both the basic monthly files and the Annual Social and Economic Supplement (ASEC), or "March files".[10] The pooled dataset includes about 44 million households over the entire period, which gives sufficient statistical power for granular event studies.

Three different labor market outcomes are considered: annual employment (worked last year), weekly employment (worked last week), and earnings (wages and salary last year). Annual em-

[9] In the context of child penalties, recent studies provide correlational evidence consistent with effects of social norms, including cross-country evidence (Kleven *et al.* 2019; Moriconi and Rodríguez-Planas 2021) and within-country evidence from the Netherlands (Rabaté and Rellstab 2022). This body of work relates to the correlational analysis of elicited gender progressivity provided here. Based on the pseudo-event study methodology, this paper is able to exploit both time and spatial variation at a granular level.

[10] March files from 1968-2020 are included in the analysis, whereas monthly files are included only from 1989 onwards. Although the monthly files go back to 1976, they do not allow for accurately identifying the presence and number of children prior to 1989. See Kleven (2024) for details.

ployment and earnings are observed in the CPS March files and ACS, but not in the CPS monthly files. The presence of children is measured using information on own children living in the household, including biological children, step children, and adopted children. The event time of parents is measured using information on the age of the oldest child living in the household. For studying the impact of social norms and culture, a key feature of the data is that it includes information on state of birth (ACS data) and country of birth (ACS data and CPS data since 1994). This allows for epidemiological studies of both movers within the US and immigrants from abroad.

The pseudo-event study approach is validated against a true event study approach using pooled data from the Panel Study of Income Dynamics between 1968-2019 (PSID 1968-2019) and the National Longitudinal Survey of Youth between 1979-2018 (NLSY 1979-2018). The pooled panel dataset includes about 17,000 households. This gives enough data for conducting validation exercises in the full sample and in broad subsamples, but the PSID/NLSY data are underpowered for more granular analyses.

3 Methodology: Event Studies and Pseudo-Event Studies

3.1 Potential Outcomes Framework

This section develops a potential outcomes framework to elucidate the identification of child penalties using either event studies or pseudo-event studies. The setup and notation are similar to Roth *et al.* (2023). Individuals are indexed by i and time by $t = 1, ..., T$. The treatment of interest is parenthood — having had a first child — which may occur at any time and is an absorbing state. The treatment indicator $D_{it} = \{0, 1\}$ denotes whether individual i is a parent in period t, and $F_i = \min\{t : D_{it} = 1\}$ denotes the year of first childbirth. If an individual does not have a first child during the sample, then $F_i \equiv \infty$.

To define potential outcomes, let $\mathbf{0}_s$ and $\mathbf{1}_s$ denote s-dimensional vectors of zeros and ones, respectively. The potential outcome for individual i at time t, if their year of first childbirth is $F_i = f$, can be defined as $Y_{it}(f) \equiv Y_{it}(\mathbf{0}_{f-1}, \mathbf{1}_{T-f+1})$. The potential outcome if they do not have a first child within the sample period is defined as $Y_{it}(\infty) \equiv Y_{it}(\mathbf{0}_T)$. The framework allows potential outcomes to depend on the entire path of treatment assignments, but because parenthood is an absorbing state, the treatment path can be summarized simply by the year of first treatment f.

The object of interest is the average treatment effect on the treated (ATT). This effect may vary by year of first child f and by year t. We have

$$\text{ATT}(f,t) = \mathbb{E}\left[Y_{it}(f) - Y_{it}(\infty) \,|F_i = f\right]. \tag{1}$$

The identification challenge is that the untreated potential outcome $Y_{it}(\infty)$ is not observed for the treated group ($F_i = f$). Difference-in-differences and event study methods overcome this challenge through assumptions that allow researchers to impute the mean counterfactual outcome. I first describe the assumptions for causal identification in event studies using panel data, and then describe the (additional) assumption for identification in pseudo-event studies using cross-sectional data.

Event Study Approach: Event study designs rely on two main assumptions: parallel trends and no anticipatory effects. In our staggered event study setting, the parallel trends assumption can be formulated as follows.

Assumption 1 (Parallel Trends). *We assume that, on average, the trend in untreated potential outcomes $Y_{it}(\infty)$ is parallel between groups who have their first child at different times F_i. That is,*

$$\mathbb{E}\left[Y_{it}(\infty) - Y_{it'}(\infty) \,|F_i = f\right] = \mathbb{E}\left[Y_{it}(\infty) - Y_{it'}(\infty) \,|F_i = f'\right], \tag{2}$$

for all $t \neq t'$ and $f \neq f'$.

The other main assumption — no anticipation — states that, if an individual is untreated at time t, their outcome does not depend on the timing of future treatment. In other words, individuals without children do not act on any knowledge of whether and when they will have children. This may be formalized as follows.

Assumption 2 (No Anticipatory Effects). *We assume that, on average, the treated potential outcome $Y_{it}(f)$ is the same as the untreated potential outcome $Y_{it}(\infty)$ prior to childbirth. That is,*

$$\mathbb{E}\left[Y_{it}(f)\right] = \mathbb{E}\left[Y_{it}(\infty)\right] \text{ for all } t < f. \tag{3}$$

Under these assumptions, we may rewrite the average treatment effect on the treated in terms of observables. Using equations (2)-(3), the definition of ATT in equation (1) can be expressed as a difference-in-differences:

$$\begin{aligned} \text{ATT}(f,t) &= \mathbb{E}\left[Y_{it}(f) - Y_{i,f-1}(f) \,|F_i = f\right] \\ &\quad - \mathbb{E}\left[Y_{it}(f') - Y_{i,f-1}(f') \,|F_i = f' > t\right], \end{aligned} \tag{4}$$

where all of the potential outcomes correspond to *realized* outcomes observed in panel data. This can be estimated based on difference-in-differences or event study regressions. As described below, the baseline specification pools all ages at first birth to recover the average effect of parenthood on all parents. A stacked specification that allows for heterogeneous effects by age at first birth is also considered.

Child penalties are defined as effects of parenthood on women relative to men (Kleven, Landais and Søgaard 2019). This reflects that child penalties are intended to measure effects on gender gaps. Importantly, this implies that Assumptions 1-2, while necessary for estimating the effects on women and men *separately*, are stronger than required for estimating child penalties. Denoting the average treatment effect on gender $g = w, m$ by ATT (f, t, g), the assumptions may be relaxed as follows.

Corollary 1 (Weaker Assumptions). *Child penalties defined as ATT (f, t, m) − ATT (f, t, w) can be causally identified under violations of parallel trends and no anticipation (Assumptions 1-2), provided the magnitude of bias in ATT (f, t, g) is unrelated to gender $g = w, m$.*

To put it differently, because child penalties represent differences in the treatment effects of two groups, the event study approach described here amounts to a triple-differences approach.

Pseudo-Event Study Approach: Now consider a setting without panel data, but with repeated cross-sectional data. Such data do not allow for event studies where the timing of treatment varies by individual, because, for those observed before treatment, we do not know when treatment will take place. In the context of child penalties, we do not know whether and when individuals observed without children will eventually become parents. Denoting event time relative to the year of first childbirth by $\tau_{it} = t - F_i$, we observe positive event times $\tau_{it} \geq 0$ for parents but not negative event times $\tau_{it} < 0$ for non-parents. The pseudo-event study approach imputes observations of negative event times by matching parents at $\tau_{it} = 0$ — those closest in time to childlessness — to individuals without children.

To see how the matching algorithm works, consider parent i observed at time $t = F_i$ with age $a_{it} = a$ and a vector of (time-invariant) demographic characteristics $X_i = X$. In a panel, this parent would be observed in prior years $t-n$, where their event time would be $\tau_{i,t-n} = -n$, their age $a-n$, and their demographic characteristics X. To mimic this panel structure, we match parent i to a childless individual observed in year $t-n$ with age $a-n$ and demographics X.[11] This childless individual has the same observables as parent i would have at event time $-n$, had they been observed. Hence, they are assigned the synthetic event time $\tau^S_{i,t-n} = -n$.

Besides Assumptions 1-2, the pseudo-event study approach relies on the following conditional independence assumption:

Assumption 3 (Conditional Independence). *Parent i sampled in year t with year of first birth $F_i = t$, age $a_{it} = a$, and demographic characteristics $X_i = X$ is matched to a non-parent sampled in year $t-n$ with age $a-n$ and the same demographics X. The matched non-parent is used as a synthetic observation of parent i in year $t-n$, where we assume*

$$\begin{aligned}\mathbb{E}\left[Y_{i,t-n}(t)\,|F_i = t, a_{i,t-n} = a-n, X_i = X\right] &= \\ \mathbb{E}\left[Y_{i,t-n}(F_i)\,|F_i > t-n, a_{i,t-n} = a-n, X_i = X\right], & \qquad (5)\end{aligned}$$

for all t and $n > 0$. By construction, the two samples have the same composition on age and demographics, such that equation (5) implies $\mathbb{E}\left[Y_{i,t-n}(t)\,|F_i = t\right] = \mathbb{E}\left[Y_{i,t-n}(F_i)\,|F_i > t-n\right]$. Given Assumption 2, the treated potential outcomes $Y_{i,t-n}(t), Y_{i,t-n}(F_i)$ can be replaced by the untreated potential outcome $Y_{i,t-n}(\infty)$.

Under Assumption 3, the panel-data version of ATT can be recovered from cross-sectional data. Consider the formulation of ATT(f,t) provided in equation (4): the first term is an outcome at event time $t-f > 0$, while the remaining terms are outcomes at negative event times. The negative event times are not directly observed, but can be replaced by their synthetic counterparts. Specifically, the second term $\mathbb{E}\left[Y_{i,f-1}(f)\,|F_i = f\right]$ is obtained by setting $t = f$ and $n = 1$ in Assumption 3. The other terms correspond to different combinations of t and n.

To be clear, the approach is ultimately a form of cross-sectional matching. The novel feature of the approach is to match the time-varying variables — year and age — in a staggered manner to

[11] A parent will have multiple possible matches whenever there is more than one childless individual in the specified cell of observables (year, age, and demographics). We match the parent to all childless individuals in the given cell, each of them weighted by $1/k$ where k is the cell size.

represent the cross-sectional data as a synthetic panel and conduct event studies. The conditional independence assumption required for the approach to be valid is weaker than in traditional matching approaches. Specifically, the assumption states that *untreated* potential outcomes — but not the treated ones — are independent of treatment assignment, conditional on covariates.[12] This reflects that the assumption has a more limited function than in standard cross-sectional matching: it allows only for converting cross-sectional data into a panel, not for identifying causal effects by comparing treatments and controls in the cross-section.[13] Crucially, the conditional independence assumption of the pseudo-event study approach can be directly tested in panel data. Related to Corollary 1, because child penalties are defined as effects on women relative to men, the relevant test is whether any violations of equation (5) differ by gender. We provide such a test below.

Inspecting pre-treatment trends is important for evaluating the credibility of parallel trends (Assumption 1) and no anticipation (Assumption 2). Given Assumption 3, the pre-trends observed in pseudo-event studies are informative:

Corollary 2 (Pre-Treatment Trends). *It follows directly from Assumption 3 that the pre-trends in panel data,* $\mathbb{E}\left[Y_{i,f-n}\left(f\right)-Y_{i,f-n'}\left(f\right)|F_i=f\right]$ *for all* $n, n' > 0$*, are identical to the synthetic pre-trends in the matched cross-sectional data.*

Consistent with this result, we do not find statistically significant differences in the pre-trends of event studies and pseudo-event studies. To summarize, while Assumption 3 may seem strong, it can be directly tested in panel data and indirectly validated by considering its implications for pre-trends and child penalty estimates. Our results show that, at least in the context of child penalties, the approach works exceedingly well. Future work will show if the approach is portable to other settings.

3.2 Regression Framework

The potential outcomes framework developed above motivates a two-way fixed effects (TWFE) model, where outcomes are regressed on individual fixed effects, year fixed effects, and dummies

[12] Consider the standard conditional independence assumption in cross-sectional matching: $Y(1), Y(0) \perp D \,|\, X$. Using this static notation, Assumption 3 corresponds to $Y(0) \perp D \,|\, X$. As an aside, note also that Assumption 3 is stated as independence in the *means* of potential outcomes, as opposed to the (stronger) assumption of independence in the *full distributions* of potential outcomes.

[13] In exchange for the weaker conditional independence assumption, we have to make the standard event study assumptions as well (parallel trends and no anticipation).

for event time relative to the year of first childbirth $\tau_{it} = t - F_i$. In practice, most child penalty studies adopt a less saturated specification with age rather than individual fixed effects (following Kleven, Landais and Søgaard 2019).[14] This is strictly necessary in the cross-sectional approach developed here.[15] Hence, we consider the following regression specification run separately for men and women:

$$Y_{it}^{g} = \sum_{\tau'} \beta_{\tau'}^{g} \cdot \mathbf{1}\left[\tau' = \tau_{it}\right] + \phi_{a_{it}}^{g} + \psi_{t}^{g} + \nu_{it}^{g}, \tag{6}$$

where an outcome Y_{it}^{g} for individual i of gender g in year t is regressed on event time dummies, age fixed effects, and year fixed effects. In the first term, we omit a base year before childbirth such that the event time coefficient β_{τ}^{g} measures the impact of childbirth at event time τ relative to the omitted base year.[16] As described above, child penalties will be defined in terms of the difference in effects between women and men, β_{τ}^{w} and β_{τ}^{m}.

Our focus is on labor market outcomes such as earnings and employment. Equation (6) is specified in levels rather than in logs to keep observations with zero earnings and employment, thus capturing both intensive and extensive margin effects. The estimated level effects are converted into percentage effects by calculating

$$P_{\tau}^{g} = \frac{\hat{\beta}_{\tau}^{g}}{\mathbb{E}\left[\tilde{Y}_{it}^{g} \mid \tau_{it} = \tau\right]}, \tag{7}$$

where $\tilde{Y}_{it}^{g}$ is the predicted outcome when omitting the contribution of the event time coefficients, i.e., the counterfactual outcome absent children. The main argument for focusing on percentage effects is that these are easier to compare across time, geography, and groups. Evidence will be presented to show that the variation in percentage and level effects is very similar.

Finally, the child penalty is defined as the average effect of having children on women relative to men over a specified event time horizon, i.e.

$$\text{Child Penalty} \equiv \mathbb{E}\left[P_{\tau}^{m} - P_{\tau}^{w} \mid \tau \geq 0\right] - \mathbb{E}\left[P_{\tau}^{m} - P_{\tau}^{w} \mid \tau < 0\right]. \tag{8}$$

[14] In general, separately identifying individual, year, and event time effects can be challenging, because individual and year implies event time relative to year of first childbirth. It is not infeasible, however, given the individual and year fixed effects are additive rather than interacted, and the two types of specifications tend to give similar results.

[15] Note that, while we match parents at $\tau_{it} = 0$ to non-parents at $\tau_{it} < 0$, we do not match them to *other* parents at $\tau_{it} > 0$. That is, the matching algorithm creates a synthetic panel up to event time zero, but not after event time zero. The approach could be extended to match both ways in which case (synthetic) individual fixed effects would be feasible.

[16] In the empirical implementation, the omitted base year is the year before pregnancy. The choice of base year hardly affects the results as there is very little pre-trend in the data.

The penalty is specified as the average effect across treated (non-negative) event times net of the average effect across untreated (negative) event times. The second term is not strictly necessary due to having omitted a base year before childbirth, but it improves the estimation in some of the more granular (and thus noisier) heterogeneity analyses. A positive child penalty implies that parenthood increases the gender gap.

Treatment-Effect Heterogeneity: Besides the assumptions formalized above, staggered event studies make implicit restrictions on treatment-effect heterogeneity across units treated at different points in time. A recent econometrics literature studies the potential bias arising from such heterogeneity (e.g., de Chaisemartin and D'Haultfœuille 2020; Sun and Abraham 2021; Goodman-Bacon 2021; Callaway and Sant'Anna 2021; Roth *et al.* 2023; Borusyak, Jaravel and Spiess 2024). A paper by Melentyeva and Riedel (2023) investigates this issue in the context of child penalties.

Given the staggered event study specification in (6) relies on variation in the timing (age or year) of first childbirth, the threat to identification is heterogeneity in this dimension. Using an approach similar to Sun and Abraham (2021) and Melentyeva and Riedel (2023), we address this issue by considering a stacked event study that allows for heterogeneous treatment effects by age at first birth:

$$Y_{itf}^{g} = \sum_{f'} \sum_{\tau'} \beta_{\tau' f'}^{g} \cdot \mathbf{1}\left[\tau' = \tau_{it}\right] \cdot \mathbf{1}\left[f' = f\right] + \phi_{a_{it} f}^{g} + \psi_{tf}^{g} + \nu_{itf}^{g}, \tag{9}$$

where Y_{itf}^{g} is the outcome for an individual whose age at first birth belongs to group f. The effects of event time, age, and year are all allowed to vary with f. The coefficient $\beta_{\tau f}^{g}$ measures the effect of childbirth at event time τ by gender g and age at first birth f. Defining scaled impacts as $P_{\tau f}^{g} = \hat{\beta}_{\tau f}^{g} / \mathbb{E}\left[\tilde{Y}_{itf}^{g} \mid \tau, f\right]$, the weighted average treatment effect across birth cohorts can be calculated as

$$P_{\tau}^{g} \equiv \sum_{f} \omega_f P_{\tau f}^{g}, \tag{10}$$

where ω_f denotes the sample share of cohort f. The weighted average in (10) does not suffer from the key issue discussed in the recent econometrics literature, namely that staggered event studies may assign negative weights to some cohorts (de Chaisemartin and D'Haultfœuille 2020; Goodman-Bacon 2021; Borusyak, Jaravel and Spiess 2024). As we shall see, the estimates obtained from the stacked specification are virtually identical to the estimates obtained from the baseline specification that pools all ages at first birth.

3.3 Descriptive Statistics

This section provides descriptive statistics in the cross-sectional and pseudo-panel datasets. These are based on the pooled CPS and ACS data.

Table 1 compares men and women observed with and without children in the cross-section. The table highlights the main identification challenge when estimating the impact of children: selection into parenthood. To see the problem, it is particularly informative to consider the outcomes of men. Men with children have better labor market and demographic outcomes than men without children. For example, their employment rates and earnings are much higher. In light of recent evidence showing that parenthood has no impact on the labor market outcomes of men (e.g., Kleven, Landais and Søgaard 2019; Kleven *et al.* 2019), these patterns must reflect positive selection. A similar selection problem seems to exist for women: the earnings of women with and without children are almost the same, despite the fact that child penalties pull mothers down, all else equal.

The standard event study approach addresses such selection problems by relying on within-individual variation around first childbirth using panel data. Absent panel data, the solution proposed here is to create a pseudo-panel with variation around first childbirth for the same synthetic individuals. The pseudo-panel data is created from repeated cross-sectional data using the matching algorithm described in section 3.1. This algorithm matches parents to non-parents based on age and year (in a staggered manner) along with a set of demographic variables. The choice of demographic variables can be anchored in results obtained from panel data: the pseudo-event study approach should give the same results as an actual event study approach. A particularly useful moment is the effect of first childbirth on men. Because childbirth is a non-event for men in actual event studies, if the pseudo-event study is associated with a positive jump in the labor market outcomes of men at $\tau = 0$, this reflects bias from positive selection. The set of matching variables used here is chosen to avoid such selection bias. These are gender, education (4 categories), marital status (5 categories), race (4 categories), and state of residence.[17]

Table 2 provides descriptive statistics for matched men and women at event times $\tau = 0$ and $\tau = -1$ in the pseudo-panel.[18] By construction, these samples match exactly on education, marital

[17] The binned matching variables are specified as follows. Education categories: Below high school degree, high school degree, some college or associate's degree, and college degree or more. Marital status categories: Married with spouse present, married with spouse absent or separated, divorced, widowed, and never married. The race categories combine information on race and ethnicity: white (non-Hispanic), black (non-Hispanic), Hispanic, and all others (mostly Asian).

[18] The sample is restricted to parents whose age at first birth lies between 25 and 45.

status, race, age at first birth, and cohort. Also by construction, individuals at event time $\tau = 0$ are exactly one year older than those at event time $\tau = -1$. The samples do not match on labor market outcomes, nor are they supposed to: those observed at $\tau = 0$ are one year further in their lifecycle and in time (making their outcomes better), and they may be affected by child penalties (making their outcomes worse). To isolate the child penalty component, lifecycle and time trends are absorbed by age and year fixed effects as explained above.

4 Baseline Estimates and Validation

Cross-Section vs Panel: Figure 1 compares the results obtained from pseudo-event study and true event study approaches. The pseudo-event studies are based on CPS and ACS data over the period 1968-2020, while the true event studies are based on PSID and NLSY data over the same period. Each panel shows an event study for men and women around first childbirth at $\tau = 0$. The event time horizon shown in these and subsequent graphs goes from $\tau = -5$ to $\tau = 10$. The average child penalty over event times 0-10 is displayed in each panel. Three outcomes are shown: annual employment (top panels), weekly employment (middle panels), and earnings (bottom panels).[19]

The results from the two approaches are closely aligned, but the pseudo-event studies feature much greater statistical precision. The difference in standard errors is mainly due to the different sample sizes of the cross-sectional and panel datasets.[20] All of the event studies show parallel trends between men and women before childbirth, combined with sharp divergence immediately after. Having a child is a non-event for men, but leads to an immediate and persistent drop in the labor market outcomes of women. The child penalties obtained from the cross-sectional approach equal 23% in annual employment, 25% in weekly employment, and 33% in earnings. The ranking of these penalties corresponds to what one would expect, because weekly employment includes

[19] It should be noted that the matching algorithm described in section 3.1 has to be implemented differently for weekly employment (obtained from a question about work activities *last week*) and annual employment/earnings (obtained from a question about earnings *last year*). Because of the retrospective nature of the annual outcomes, matching parents at event time $\tau = 0$ (observed in year t with age a) to non-parents (observed in year $t - n$ with age $a - n$) gives a synthetic observation of event time $\tau = -n - 1$. For the same reason, annual outcomes at non-negative event times $\tau = 0, ..., 10$ are obtained from parents observed at event times $\tau = 1, ..., 11$.

[20] The CPS/ACS pseudo-event studies are based on estimation samples that are 143 times larger (weekly outcome) and 80 times larger (annual outcomes) than the PSID/NLSY event studies. For both the pseudo-event and event studies, I compute standard errors on the percentage effects P_τ^g (defined in equation 7) by treating the scaling factor as non-stochastic. The more correct approach of bootstrapping the standard errors is computationally very time-consuming. Importantly, bootstrapping tends to give marginally *smaller* standard errors in this context, implying that the approximation provided here is conservative.

effects on both extensive and intensive margin labor supply, and because earnings includes effects on both labor supply and wage rates. The child penalties obtained from the panel approach are very similar.

To evaluate the choice of matching variables, Figures A.1-A.3 in the appendix show results for more parsimonious specifications. For each labor market outcome, four specifications are shown: matching only on year, age, and gender (Panel A), adding education (Panel B), adding marital status (Panel C), and adding race and state (Panel D). The specification in Panel D corresponds to the baseline specification presented above. The main insight is that the more parsimonious specifications are associated with positive jumps in the labor market outcomes of men between $\tau = -1$ and $\tau = 0$. As discussed above, such jumps reflect selection rather than a causal effect of children. Adding matching variables reduces the size of these jumps, and the baseline specification in Panel D eliminates them entirely.

Panel vs Panel: The preceding validation exercise compares results from different datasets. This conflates differences due to methodology and sample selection. It is possible that the pseudo-event studies align with the true event studies due to offsetting effects from sample selection. A more direct validation uses only panel data, conducting pseudo-event studies by ignoring the information on negative event times. Figure 2 presents such a validation, comparing pseudo-event studies and true event studies using PSID/NLSY data for both.

The within-panel validation is exceedingly convincing. The pseudo-event studies look very similar to the true event studies, and they produce virtually identical child penalties. The difference in child penalty estimates obtained from the two approaches is at most 1 percentage point across the three outcomes. The results in this figure suggest that the (minor) differences in estimates seen in the previous figure were driven, not by methodology, but by differences in the CPS/ACS and PSID/NLSY samples.

Hausman Test: We may provide a formal statistical test of the equivalence between the pseudo-event study and event study approaches using a Hausman specification test. The idea is that we have two ways of estimating the same parameter, one of which is consistent (event study) and one of which is more efficient but potentially inconsistent (pseudo-event study). We focus on the main empirical target — the average child penalty — rather than the full set of event time coefficients for men and women separately. For this purpose, we consider a difference-in-differences version of equation (6) estimated on the sample of men and women together:

$$Y_{it}^{g} = \beta \cdot Post_{it} + \beta^{w} \cdot Post_{it} \cdot \mathbf{1}\left[g = w\right] + \phi_{a_{it}}^{g} + \psi_{t}^{g} + \nu_{it}^{g}. \tag{11}$$

Here, the event time dummies have been collapsed into a binary indicator $Post_{it}$ for having had a first child. The fixed effects for age and year are allowed to vary by gender as in equation (6). The object of interest is β^{w}, which captures the average effect of first childbirth on women relative to men. This is the (unscaled) child penalty. Estimating equation (11) in the panel and the pseudo-panel, we obtain the estimates $\hat{\beta}_{true}^{w}$ and $\hat{\beta}_{pseudo}^{w}$. The Hausman test statistic is given by

$$H = \frac{\left(\hat{\beta}_{true}^{w} - \hat{\beta}_{pseudo}^{w}\right)^{2}}{\text{Var}\left(\hat{\beta}_{true}^{w}\right) - \text{Var}\left(\hat{\beta}_{pseudo}^{w}\right)}, \tag{12}$$

which, under the null hypothesis that the pseudo-panel estimator is consistent, follows a chi-squared distribution with one degree of freedom.

It is most informative to focus on the within-panel comparison, even if the main efficiency gains of the pseudo-event study approach only materialize when moving to the CPS/ACS data. As described, the within-panel comparison isolates the effect of empirical methodology with no contamination from sample selection. Conducting the test in all three outcomes, we obtain $H = 0.023$ (p-value of 0.880) for annual employment, $H = 0.290$ (p-value of 0.590) for weekly employment, and $H = 0.099$ (p-value of 0.754) for earnings. These results strongly support the consistency of the pseudo-panel estimator.

We may also compare the two approaches across different datasets: $\hat{\beta}_{true}^{w}$ estimated from PSID/NLSY data and $\hat{\beta}_{pseudo}^{w}$ estimated from CPS/ACS data. This conflates the effects of methodology and sample selection. In this case, we obtain $H = 0.228$ (p-value of 0.633) for annual employment, $H = 17.328$ (p-value of 0.000) for weekly employment, and $H = 14.807$ (p-value of 0.000) for earnings. Hence, the differences are statistically significant for weekly employment and earnings, but not for annual employment. Importantly, the differences in the estimates are *economically* small as shown in Figure 1.[21]

[21] The last possible comparison is between the pseudo-event study estimates obtained from PSID/NLSY and CPS/ACS data, respectively. This comparison yields similar results: we reject equivalence of the estimates for weekly employment and earnings, but not for annual employment.

Test of Assumption 3: In addition to the Hausman specification tests, Assumption 3 can be directly verified in panel data. This assumption requires balance between actual and synthetic observations at negative event times, where the synthetic observations are obtained from the matching algorithm described above. Appendix Table A.1 investigates whether this holds in the pooled PSID/NLSY data. The table reports average labor market outcomes in the panel and pseudo-panel, as well as the differences between the two.[22] Outcomes are shown for men, women, and men relative to women. To keep the table manageable in size, we focus on event time $\tau = -2$ (omitted base year) and the average of all negative event times.[23] The test provides further support for the empirical approach: the differences between the panel and pseudo-panel are small, and most of them are statistically insignificant. The largest differences are found for earnings, but these are statistically insignificant when considering men relative to women. This is the critical test for estimating child penalties: any imbalance between actual and synthetic observations that is present in both the male and female samples has no consequences for child penalties.

Validation in Subsamples: The statistical precision of the pseudo-event study approach allows for studying child penalties at a granular level. However, while we have validated the approach in the full sample, it does not necessarily follow that it performs equally well in subsamples. The small sample size of the panel data limits the feasible granularity of validation exercises, but it is possible to validate the pseudo-event study approach in broad subsamples.

Figure 3 provides such an analysis. Using PSID/NLSY data, the figure plots pseudo-event study estimates against true event study estimates in subsamples. The sample is split by geography (4 census regions), time (5 decades), education (high school or less vs college), marital status (single vs married), and race (4 categories). This gives a total of 17 subsamples. Despite the added granularity, the validation results remain compelling. For all three labor market outcomes and across all subsamples, the child penalty pairs lie close to the 45-degree line. The R-squared from a regression of pseudo-panel estimates on panel estimates lies between 0.83 and 0.88 across the three different outcomes. This suggests that the pseudo-event study approach is consistent even in subsamples, lending support to the analyses presented below.

[22] For ease of interpretation, the earnings outcome is normalized by average counterfactual earnings across all years of the data (for men and women separately), i.e., by $\mathbb{E}\left[\tilde{Y}_{it}^{g}\right]$ defined above. This normalization makes the reported differences directly comparable to the child penalty estimates.

[23] Note that the assumption in equation (5) is stated as balance between actual and synthetic outcomes within each cell of demographics used for matching, which implies balance in the averages across all cells of demographics. Ultimately, the approach only needs balance in the averages (but for each event time). Hence, Table A.1 focuses on the averages at different event times.

Predicted Fertility: Why does the pseudo-event study approach work so well? A reason could be that the approach accurately predicts fertility (location in negative event time) among those observed without children. Appendix Figure A.4 investigates this point, comparing predicted and actual event times among childless people in PSID/NLSY data. The figure shows the distribution of within-person differences between predicted event times (obtained from matching) and actual event times (directly observed).[24] Event time is perfectly predicted for 34% of the data, and predicted with an error of less than four years for 74% of the data. This is arguably very good considering the simplicity of the approach, but not perfect. As implied by the preceding validation exercises, the discrepancies between predicted and actual fertility do not destroy the accuracy of the pseudo-event study approach. The approach ultimately relies on a conditional independence assumption (Assumption 3), which is not a statement about the accuracy of predicted fertility. As we have seen, the assumption is directly testable and holds in this setting.

Treatment-Effect Heterogeneity: We investigate the possibility of bias from treatment-effect heterogeneity by comparing results from the baseline and stacked specifications described in section 3.2. The stacked specification, instead of pooling all ages at first birth, allows for heterogeneous effects by age at first birth (equation 9) and calculates a weighted average treatment effect using the sample shares of each cohort (equation 10). If treatment-effect heterogeneity creates bias by assigning "weird" weights to some cohorts in the pooled specification, we expect the baseline and stacked specifications to produce different results. Because this issue is not about the alignment of pseudo-event and true event study approaches, we focus solely on our main design: pseudo-event studies using cross-sectional data.

The results are presented in Appendix Figure A.5. The figure shows results from the baseline and stacked specifications for each of the three labor market outcomes.[25] The two specifications produce almost the same results, both in terms of the shape of the event studies and in terms of the magnitude of the child penalty. In particular, the results for annual and weekly employment are virtually identical, while the results for earnings are only marginally different (a child penalty of 31% vs 33%). These results speak against any non-trivial bias from heterogeneous treatment effects.

[24]The distribution is based on individuals observed in the panel data after age 45, ensuring that completed fertility can be measured.

[25]The stacked specification splits age at first birth (which ranges from 25 to 45 in the full sample) into four groups: 25-28, 29-32, 33-36, and 37-45 years of age.

The robustness of the results is arguably unsurprising when considering the shape of the baseline event studies. They have two key features: pre-event trends are parallel *and* post-event effects are almost perfectly persistent following a sharp effect at event time zero. In other words, there are virtually no dynamics in the data, except for the sharp effect precisely at event time $\tau = 0$. Given the specification controls for age and year fixed effects, the variation in event time comes from variation in age/year of first birth. Hence, the absence of dynamics outside $\tau = 0$ implies that there is little heterogeneity by age at first birth. There *is* heterogeneity in the magnitude of the drop at $\tau = 0$, but the trajectories are flat elsewhere. This is particularly true for the employment outcomes, explaining why these are more robust than earnings. In general, concerns about the validity of staggered event study designs are warranted primarily in settings with significant dynamics outside the time of the treatment event.

5 Variation in Child Penalties

5.1 Child Penalties Over Time

Figure 4 shows the evolution of child penalties over time. To construct these time series, the sample of parents is split by year of interview, and the event study specification (6) is run for different time periods separately.[26] The event studies for each time period and labor market outcome are presented in Figures A.6-A.8 of the appendix.

Child penalties have fallen substantially over the last five decades. In the 1970s, the penalties were 46% in annual employment, 53% in weekly employment, and 70% in earnings. In the 2010s, the penalties were 20% in annual employment, 24% in weekly employment, and 31% in earnings. The decline is therefore larger than 50% in all three outcomes, albeit from an exceptionally high baseline level. Almost all of the decline in child penalties occurred prior to the mid-1990s, followed by a long period of stagnation.

The time series evidence in Figure 4 shows scaled child penalties (effects in percentage terms). These penalties may change over time either due to changes in unscaled child penalties (effects in absolute terms) or due to changes in the scaling factor (the level of the counterfactual outcome).

[26]Given the matching specification used (matching parents observed in a given year to non-parents observed in prior years), splitting the sample of parents into different time periods implies that some of their non-parent matches were observed before the time period in question. All sample splits shown in the paper are based on splitting the sample of parents by some characteristic and using their non-parent matches regardless of whether they share the same characteristic.

Appendix Figure A.9 investigates if the changes are driven primarily by one or the other. The figure shows that the evolution of unscaled penalties is qualitatively similar to the evolution of scaled penalties: a decline until the mid-1990s and then stagnation. The counterfactual outcomes used for scaling have remained relatively constant for employment, while they have increased gradually for earnings.[27] Hence, changes in the baseline hardly matter for the evolution of scaled employment penalties, but they play some role for the evolution of scaled earnings penalties.

These findings shed new light on a stylized fact documented in the literature: the slowdown of gender convergence since the 1990s (e.g., Blau and Kahn 2006, 2017; Kuziemko *et al.* 2018). This trend has been viewed as puzzling given the increases in female education and job experience over the same period. The literature has discussed a variety of explanations, but no conclusive evidence has emerged. The evidence presented here suggests an explanation: gender convergence stalled because the decline in child penalties stalled.

How much of gender convergence can be attributed to child penalties? For each labor market outcome, Appendix Figure A.10 shows the fraction of the gender gap for parents explained by child penalties over time.[28] These fractions are relatively stable and very high. Child penalties explain 90-100% of the gender gap in annual employment, 80-90% of the gender gap in weekly employment, and about 50% of the gender gap in earnings.[29] This implies that the evolution of gender inequality in labor market outcomes — especially the employment outcomes — can be explained largely by the evolution of child penalties.

This explanation is admittedly very reduced-form. We may define the gender gap in a given outcome as the sum of child-related inequality (child penalties) and residual inequality (non-child effects). These are reduced-form concepts that depend on a set of underlying and potentially over-

[27]When plotting unscaled penalties and counterfactual levels for earnings over time, it is important to adjust the estimates (in dollars) for nominal earnings growth. Therefore, in Figure A.9, the earnings estimates have been inflated to 2020 dollars using nominal earnings growth in the full sample of working-age men and women in CPS data.

[28]The fraction of the gender gap for parents explained by child penalties can be calculated as

$$\text{Fraction Explained} = \frac{\text{Child Penalty}}{\text{Gender Gap}} \times \frac{\mathbb{E}\left[\tilde{Y}_{it}^{w} \mid \tau \geq 0\right]}{\mathbb{E}\left[Y_{it}^{m} \mid \tau \geq 0\right]},$$

where Gender Gap $\equiv \frac{\mathbb{E}[Y_{it}^{m}|\tau\geq 0]-\mathbb{E}[Y_{it}^{w}|\tau\geq 0]}{\mathbb{E}[Y_{it}^{m}|\tau\geq 0]}$, and where the second term on the right-hand side adjusts for the fact that the child penalty and gender gap have different denominators (counterfactual outcome for mothers vs actual outcome for fathers). This formula relies on a steady-state assumption, namely that the child penalty stays constant outside the event study window considered in the estimation. The observed stability of the child penalty within the event study window (up to $\tau = 10$) lends support to this assumption.

[29]For annual employment, the fraction of the gender gap explained by child penalties was just above 100% in the late 1980s. This implies that, if not for the impact of parenthood, women would have had a larger annual employment rate than men.

lapping factors. For example, an underlying factor may be social norms, and these could operate both through child penalties and non-child effects. A key objective of this paper is to gain a better understanding of the mechanisms that drive child penalties, leaving aside the mechanisms that drive non-child effects. The fact that child penalties account for most of the observed gender inequality implies that non-child effects are a relatively small part of the story.

5.2 Child Penalties Across States

To study the variation in child penalties across states, the event time dummies in equation (6) are interacted with state dummies.[30] In this specification, the lifecycle and time trends are estimated at the level of census divisions by interacting the age and year dummies with census division dummies. Estimating lifecycle and time trends at the state level produces similar results, but the event studies for some of the smaller states (specifically for the earnings outcome) become noisier under such a granular specification.

As a first glimpse of the spatial variation, Figure 5 presents case studies of three states: North Dakota, New Jersey, and Utah. Results are shown for annual employment, weekly employment, and earnings. The impact of childbirth is sharp and persistent in all three states, but it varies greatly in magnitude. Child penalties are relatively small in North Dakota, intermediate in New Jersey, and very large in Utah. For example, the annual employment penalty equals 12% in North Dakota and 38% in Utah. Interestingly, the range of child penalties across these states corresponds to the range observed in European countries, with North Dakota resembling the small-penalty countries of Scandinavia and Utah resembling the large-penalty countries of central Europe (Kleven *et al.* 2019; Kleven, Landais and Leite-Mariante 2024).

Figures A.11-A.13 in the appendix provide event studies for all states in all three labor market outcomes. In general, these event studies look compelling: men and women are on parallel trends before childbirth, diverge immediately and sharply after childbirth, and the effects are persistent over time. As one would expect, the employment series are sharper and more precisely estimated than the earnings series, but the earnings effects are still clear and statistically significant. The results from the state-level event studies are summarized in heatmaps in Figure 6. In these maps, states are divided into deciles of the child penalty, with darker colors implying larger penalties. The annual employment penalty ranges from 11.7% to 37.8% across states, the weekly employ-

[30] To be precise, these are dummies for the 50 states plus the federal district of D.C. For simplicity, all of them will be referred to as "states."

ment penalty ranges from 14.2% to 39.9% across states, and the earnings penalty ranges from 21.1% to 60.8% across states.[31]

As discussed above, variation in scaled child penalties (effects in percentages) may reflect variation in either unscaled child penalties (effects in absolute terms) or in the counterfactual levels used for scaling. Figure 7 investigates if the spatial variation is driven mostly by one or the other, focusing on the two employment outcomes. The left panels plot unscaled child penalties against scaled child penalties across states, while the right panels plot counterfactual employment rates against scaled child penalties across states. The figure shows that the spatial variation is driven almost exclusively by variation in the effect of children in absolute terms. The relationship between scaled and unscaled penalties is almost perfectly linear with a slope close to one, whereas the counterfactual employment rate is virtually flat across states. For example, the large child penalties observed in Utah are not driven by small baseline levels of female outcomes, but rather by large effects of children on female outcomes.

Empirical Bayes Correction: A general concern with highly disaggregated analyses is that the estimated heterogeneity may overstate the true heterogeneity due to noise from smaller units. A standard approach to address such concerns is to adjust the OLS estimates using an Empirical Bayes (EB) approach. Because our main focus is on average child penalties rather than the full set of event study coefficients, we focus on the difference-in-differences specification in equation (11), extended to allow for state-specific effects of childbirth.[32] We denote the OLS estimate of the child penalty in state s by $\widehat{CP}_s$ and its standard error by SE_s. Assuming that the empirical design is valid, $\widehat{CP}_s$ is an unbiased, noisy estimate of the true child penalty CP_s. Furthermore, assuming that the true child penalty is normally distributed with mean μ and variance σ^2, the EB estimate can be calculated according to the following linear-shrinkage formula:

$$CP_s^* = \left(\frac{\sigma^2}{\sigma^2 + SE_s^2}\right) \cdot \widehat{CP}_s + \left(\frac{SE_s^2}{\sigma^2 + SE_s^2}\right) \cdot \mu, \tag{13}$$

[31] Appendix Figure A.14 shows that the spatial variation in child penalties aligns closely with the spatial variation in raw gender gaps. The figure provides scatter plots of child penalties against raw gender gaps for parents across states. There is a strong positive relationship between the two, with a slope coefficient of close to 1 in all three labor market outcomes. The strong relationship is not surprising because child penalties represent effects on the gender gap and, as we have seen, they explain a very large fraction of the gender gap. In fact, Figure A.14 can be viewed as a state-level validation of the previous findings.

[32] As before, the controls for gender-specific lifecycle and time trends are estimated at the level of census divisions.

a weighted average of the OLS estimate $\widehat{CP}_s$ and the mean μ. The mean and variance parameters are estimated as $\hat{\mu} = \frac{1}{S}\sum_s \widehat{CP}_s$ and $\hat{\sigma}^2 = \frac{1}{S}\sum_s \left[\left(\widehat{CP}_s - \hat{\mu}\right)^2 - SE_s^2\right]$. The estimated variance captures excess variance compared to what we would expect from noise. The idea of the approach is to pull the state-level child penalty estimates toward their mean based on a signal-to-noise ratio. This ratio is determined by the variance of OLS estimates across different states relative to the average variance of OLS estimates in each state.

The results are presented in Figures A.15-A.16 of the online appendix. These figures compare EB and OLS estimates of unscaled child penalties for each state and each labor market outcome. The EB shrinkage adjustment hardly changes the estimates. The reason is the high statistical precision of the pseudo-event study approach: the imprecision of the state-level OLS estimates is very small compared to the variation in OLS estimates across states, implying a high signal-to-noise ratio.[33]

Mechanisms: What are the underlying mechanisms behind the variation in child penalties across geography? Existing research suggests that labor market structure — in particular, the flexibility or family friendliness of available jobs — is a significant determinant of gender gaps (Goldin 2014; Goldin and Katz 2016; Kleven, Landais and Søgaard 2019). Hence, there is a general equilibrium aspect of child penalties that may be responsible for some of the variation across local labor markets. A proxy for labor market structure is the degree of urbanization: the family friendliness of jobs is presumably greater in rural areas (say, agriculture) than in urban areas (say, banking). The map of child penalties is consistent with such effects. Child penalties tend to be smaller in rural states (such as those in the Midwest and the South) than in urban states (such as those on the Pacific coast and the Northeast). A similar pattern is seen when focusing on smaller regions of the country. In the Northeast, for example, rural states like Maine and Vermont have smaller child penalties than urban states like New York, New Jersey, Massachusetts, and Connecticut.

Still, it is important to note that such general equilibrium effects cannot explain child penalties on their own. The family friendliness of jobs would affect mothers and fathers equally unless other factors tilt child care responsibilities toward women. In other words, this factor may serve as an amplification mechanism, not as a stand-alone explanation. Traditional explanations turn

[33] In later analyses, we consider specifications where estimated child penalties are used as regressors. In principle, such specifications should use the EB shrinkage estimates to avoid attenuation bias from statistical noise. But given the similarity of EB and OLS estimates, this makes virtually no difference to any of the results. We therefore take the simpler approach of using OLS estimates throughout.

on biology and comparative advantage in child care vs market work, but Kleven, Landais and Søgaard (2021) show that these factors have little impact on child penalties in Denmark. Evidence presented in the next section suggests that comparative advantage is also not critical for child penalties in the US. If these traditional explanations have little traction, then what drives the variation in child penalties? Evidence on the important role of social norms and culture will be presented below.

5.3 Child Penalties Across Demographic Groups

This section presents evidence on heterogeneity in child penalties by demographic characteristics. Three dimensions of heterogeneity are analyzed: education, marital status, and race. Figure 8 presents event studies of first childbirth by each of these demographics and for each of the three labor market outcomes. To construct the figure, the sample of women is split into different demographic groups and specification (6) is estimated separately for each group. Because childbirth is always a non-event for men, the sample of men is not split by demographics. The following paragraphs summarize the findings.

Education: The top row compares low-educated women (high school degree or less) and high-educated women (college degree or more). The short-run impacts of parenthood are larger for low-educated women than for high-educated women, but the impacts quickly converge to the same level. In fact, the long-run child penalty is marginally *smaller* for low-educated women.[34] This finding contradicts explanations based on comparative advantage. If this mechanism were important for explaining child penalties, we would expect larger child penalties on women with less education. The absence of such effects is consistent with findings for Denmark in Kleven, Landais and Søgaard (2021).

It is worth noting that, because female education has increased over time, the low-educated sample tends to be selected from earlier years than the high-educated sample. Because child penalties were larger historically than today, reweighting the samples to be identically distributed over time would reduce the child penalty on low-educated relative to high-educated women. This only strengthens the finding of no comparative advantage effects.

[34] The penalties for both education groups in Figure 8 are smaller than the penalties for the full sample in Figure 1. The main reason is that here we report *long-run* penalties (over event times 5-10) rather than average penalties over the full event time window.

Marital Status: The middle row compares single and married women. The category of single women includes all unmarried individuals (never married, separated, divorced, or widowed). The results are striking: single mothers have much smaller child penalties than married mothers, even though single motherhood is presumably associated with larger fixed costs of working. The patterns of heterogeneity are similar across all three outcomes, but the magnitudes are particularly stark for the employment outcomes. For example, the child penalty in annual employment equals 27% for married women and only 5% for single women. These findings highlight that child penalties are closely linked to the possibility of specialization between spouses, even if this specialization is not governed by comparative advantage as shown above.

Why are child penalties on single women much smaller than on married women in the US? Interestingly, the US evidence is exactly the opposite of the Danish evidence presented in Kleven (2021). In Denmark, child penalties on single mothers are *larger* than on married mothers. Kleven (2021) argues that this asymmetry is related to the welfare system, presenting quasi-experimental evidence from US welfare reform in the 1990s. Because single mothers cannot coordinate specialization with a spouse, they are forced to work unless the government supports their children. Denmark provides some of the most generous welfare benefits in the world (along with free education and health care), allowing single mothers to take a large child penalty in the labor market and still be able to support their children. The same is not true in the US.

Race: The bottom row compares black and white women. There is also strong heterogeneity in the race dimension, with much smaller child penalties on black women than on white women. The differences between black and white women are about as large as the differences between single and married women. In fact, the two phenomena are partly related: the rate of single motherhood is much higher among blacks than among whites (36% vs 11%). However, the higher incidence of single motherhood among blacks is not sufficient to explain all of the racial heterogeneity in child penalties. Other factors have to be at play, too. This may include cultural differences across racial groups, a mechanism studied in detail below.

6 The Effect of Gender Norms on Child Penalties

6.1 Child Penalties vs Gender Progressivity Over Time and Space

We first consider the relationship between child penalties and elicited gender norms across time and space. This analysis is correlational and descriptive. The main objective is to investigate if existing cross-country evidence on the relationship between child penalties and elicited norms (e.g., Kleven *et al.* 2019) survive when taking a more granular within-country perspective using state$\times$time variation. As we shall see, similarly strong correlations are present within the US.

The analysis starts by creating a measure of gender progressivity using General Social Survey data between 1972-2018 (GSS 1972-2018). A number of GSS questions elicit attitudes regarding gender roles in families with children. To measure gender progressivity consistently over time, we focus on three questions available in all five decades of the data. These questions ask respondents if they strongly agree, agree, disagree, or strongly disagree with the following statements:

- It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family
- A working mother can establish just as warm and secure a relationship with her children as a mother who does not work
- A pre-school child is likely to suffer if his or her mother works

A Gender Progressivity Index (GPI) is created based on the average standardized response to these questions. Specifically, the responses to each question are indexed such that a higher value corresponds to stronger gender progressivity. The responses are then standardized to have mean zero and standard deviation one, defining GPI as the average standardized response. The data is collapsed to the state-decade level, a total of 255 cells. Some of these cells have missing observations: even though the norms questions used were included in GSS in all decades, they were not asked for *every* state in *every* decade. Missing state-decade observations of the GPI are imputed based on the percentile of the state's GPI in the decades where it is observed.

Figure 9 illustrates the spatial variation in gender norms. Dividing states into deciles of the GPI, the figure presents a heatmap in which lighter (darker) colors correspond to more progressive (conservative) norms. States in the South (Bible Belt) and Utah are among the most conservative, while states in New England, the Northern Midwest, and the Pacific region are among the most progressive. Comparing the map of gender norms to the previous map of child penalties indicates that the cross-sectional correlation between the two is relatively weak, but the raw cross-sectional relationship is likely affected by confounders. The existence of time variation in both gender norms and child penalties can be used to address this issue.

Figure 10 illustrates the time variation in gender norms and child penalties. It compares the time series of the GPI (red series) to the time series of child penalties in employment and earnings outcomes (black series). The evolution in gender progressivity is an almost perfect mirror image of the evolution in child penalties. The large fall in child penalties between the 1970s and 1990s coincides with a sharp rise in gender progressivity over the same period. The stagnation in child penalties following the 1990s is associated with a stagnation in gender progressivity. The recent fall in child penalties, mainly in the earnings penalty, aligns with a recent rise in gender progressivity. The time series evidence is consistent with a strong effect of gender norms, but inconclusive by itself due to the potentially confounding effect of other time-varying factors.

Appendix Figure A.17 shows the time series of the GPI in each state separately. Gender progressivity has increased in every state, but there is substantial variation in the rate and timing of these increases. This is useful for developing a more credible empirical design that leverages both time and state variation to study the effect of gender norms on child penalties. The results from such a design are presented in Figure 11. This figure presents binscatters of child penalties vs gender progressivity across states and time, controlling for potential confounders. Specifically, the analysis is based on the following regression:

$$\text{Child Penalty}_{st} = \beta \cdot GPI_{st} + \gamma_s + \boldsymbol{\delta} \cdot \boldsymbol{X}_{st} + \nu_{st}. \tag{14}$$

That is, the child penalty is regressed on gender progressivity in state s and decade t, controlling for state fixed effects γ_s and time-varying demographic variables $\boldsymbol{X}_{st}$. The inclusion of state fixed effects absorbs all time-invariant differences across states, such as permanent differences in labor market structure and urbanization. The inclusion of demographic controls absorbs time-varying differences across states. These controls include the demographics analyzed in the previous section: education, marriage, and race.[35]

Having estimated equation (14), child penalties are residualized using the estimated effect

[35]Specifically, the controls are specified as follows. Education: the fraction of women with a high school degree or less and the fraction of women with a college degree or more. Marriage: the fraction of women who are single (never married, separated, divorced, or widowed). Race: the fraction of black women and the fraction of white women.

of the controls, $\hat{\gamma}_s + \hat{\boldsymbol{\delta}} \cdot \boldsymbol{X}_{st}$. The residualized child penalties are plotted against the GPI in a binscatter, dividing the observations of GPI into ten deciles.[36] Figure 11 presents results for each of the three labor market outcomes. The left panels include only state fixed effects, while the right panels include both state fixed effects and time-varying demographic controls. There is a strong and almost perfectly linear relationship between child penalties and gender progressivity. Given the standardization of the GPI variable, the slope coefficients can be interpreted as the effect of increasing gender progressivity by one standard deviation. In the specification with only state fixed effects, an increase in gender progressivity of one standard deviation reduces child penalties by 30.2pp in annual employment, 40.8pp in weekly employment, and 57.9pp in earnings. Adding time-varying controls reduces the effect, but the relationship remains strong. An increase in gender progressivity of one standard deviation reduces child penalties by 17.8pp in annual employment, 23.2pp in weekly employment, and 22.8pp in earnings.

These results suggest that gender norms may have sizable effects on child penalties. The use of granular within-country variation makes a causal interpretation more plausible than for existing cross-country evidence (Kleven *et al.* 2019), but the evidence is still correlational. The variation in elicited gender norms is potentially endogenous, and the choice of controls involves a great deal of uncertainty. Motivated by such concerns, the following sections consider a different approach to estimating the impact of social norms: epidemiological studies of US movers and foreign immigrants.

6.2 Epidemiological Approach: US Movers

This section investigates child penalties among US movers using information on state of birth and state of residence available in ACS data.[37] Movers are defined as US-born individuals who live in a different state than where they were born. The effect of culture is estimated based on the relationship between the child penalty for movers and the child penalty in their state of birth. This builds on the epidemiological approach to studying culture (reviewed by Fernández 2011), but typical applications of the approach focus on immigrants rather than within-country movers. Two recent studies use mover designs to estimate the effect of sexism and norms on female labor market outcomes (Charles, Guryan and Pan 2022; Boelmann, Raute and Schönberg 2023). The

[36] When plotting residualized child penalties against GPI, the average effect of the controls, $\mathbb{E}\left[\hat{\gamma}_s + \hat{\boldsymbol{\delta}} \cdot \boldsymbol{X}_{st}\right]$, is added to the residuals. This ensures that the level of the outcome variable is comparable to the child penalty estimates shown elsewhere.

[37] Information on state of birth is not available in CPS data.

analysis presented here considers the effect on child penalties, relying on granular event studies of childbirth that are feasible due to the pseudo-event study methodology. The approach is first applied to US movers and then to foreign immigrants.

As a first visualization of the results, Figure 12 presents case studies of three states: North Dakota, New Jersey, and Utah. The figure shows event studies of first childbirth for movers and stayers born in each of these states. The idea is to capture variation in child penalty culture using stayers — those born and living in the same state — as the full sample of residents will be contaminated by movers coming from states with different cultural environments. To construct the figure, specification (6) is run separately for women movers and women stayers, interacting the event time dummies with state-of-birth dummies. Because childbirth is always a non-event for men, the sample of men is not split by whether they move or stay. Results are shown for annual employment (top row) and weekly employment (bottom row). The child penalties for movers and stayers are similar in each state, but vary greatly in magnitude across states. North Dakota has small child penalties for both movers and stayers, Utah has large child penalties for both groups, while New Jersey has intermediate child penalties for both. In other words, for the three states shown, the impact of parenthood on women movers is similar to the impact in their states of birth, even though they live somewhere else and are not exposed to the labor market institutions and public policies of those birth states. This is consistent with an effect of childhood culture on child penalties.

Figures A.18-A.19 in the appendix present event studies for movers and stayers for all states. The results from these event studies are summarized in the left panels of Figure 13. These panels provide raw scatter plots of the child penalty for movers against the child penalty for stayers by state of birth. The relationship between mover and stayer penalties is very strong. Movers born in high-penalty states (such as Utah, Idaho, and Nevada) have much larger employment penalties than those born in low-penalty states (such as the Dakotas, Hawaii, and D.C.). For annual employment, the slope coefficient implies that increasing the child penalty in a woman's state of birth by 10pp increases her own child penalty by 7.2pp, although she lives and has children somewhere else. The effect of the birth-state penalty in weekly employment is similar.

While these results are striking, a threat to causal interpretation is that state of birth and state of residence may be correlated. People born in high-penalty states (such as Utah) may be more likely to move to other high-penalty states (such as Idaho), and vice versa. If moves are selected in this way, the estimated effect of state of birth (norms/culture) may be contaminated by effects of state of residence (local labor markets). The right panels of Figure 13 address this issue based on the following specification:

$$CP_s^{movers} = \beta \cdot CP_s^{stayers} + \gamma \cdot \widetilde{CP}_s^{movers} + \nu_s. \tag{15}$$

In this specification, the child penalty for movers born in state s is regressed on the child penalty for stayers and a predicted child penalty, $\widetilde{CP}_s^{movers}$, based on where they reside. The predicted child penalty is calculated as $\widetilde{CP}_s^{movers} = \sum_{s'} \alpha_{ss'}^r CP_{s'}^{stayers}$, where $\alpha_{ss'}^r$ denotes the fraction of movers from state s residing in state s'. This is the average stayer penalty across states, weighted by the actual residence choices of movers from state s. Having run the regression, the mover penalties are residualized using the estimated residence effects, $\hat{\gamma} \cdot \widetilde{CP}_s^{movers}$, and plotted against stayer penalties by state of birth.[38] As shown in Figure 13, the residualized and raw scatter plots are very similar. Controlling for selection on state of residence hardly reduces the slope coefficients and does not increase the adjusted R-squared.

The finding that place of birth has large effects on child penalties does not rule out that place of residence has important effects too. The former proxies for childhood environment (norms), while the latter proxies for adulthood environment (labor markets). To compare the relative importance of the two mechanisms, we conduct an analysis similar to the previous one, but splitting movers by their state of residence rather than their state of birth. Specifically, using specification (15), the child penalty for movers living in state s is regressed on the child penalty for stayers in state s and a predicted child penalty, $\widetilde{CP}_s^{movers}$, based on where they were born. Appendix Figure A.20 shows the results and is constructed in the same way as the preceding figure. As one would expect, place of residence also has sizable effects on child penalties. The effects are somewhat weaker than for place of birth and the R-squared values are smaller, but the findings suggest that both norms and labor markets are important for realized child penalties.

While the analysis of state-of-birth effects controls for selection on state of residence, one may still be concerned about selection in other dimensions. If movers born in low-penalty and high-penalty states differ in other dimensions that impact child penalties, the results cannot necessarily be interpreted as causal. To investigate the relevance of such concerns, Table 3 provides descriptive statistics on movers by state of birth. Specifically, Panel A compares the demographic charac-

[38]When plotting the residualized mover penalties, the average residence effect $\mathbb{E}\left[\hat{\gamma} \cdot \widetilde{CP}_s^{movers}\right]$ is added to the residuals. This makes the levels in the raw and residualized scatter plots comparable.

teristics of mothers who moved from states in the top and bottom quartiles of child penalties. The table shows that these movers are very similar on observables, apart from their residence choices as already addressed. This leaves selection on unobservables as the remaining threat to identification. The absence of selection on observables mitigates concerns about selection on unobservables (Altonji, Elder and Taber 2005), but such concerns cannot be ruled out entirely.

6.3 Epidemiological Approach: Foreign Immigrants

This section shifts the focus from US-born movers to foreign-born immigrants, using information on country of birth available in ACS data and in CPS data since 1994. The effect of culture is estimated based on the relationship between child penalties for immigrants and child penalties in their countries of birth. This is closer in spirit to typical epidemiological studies, which focus on immigrants or their descendants.

An advantage of studying immigrants from abroad rather than movers within the US is that child penalties display greater variation globally than within the US. Building on the pseudo-event study approach developed here, Kleven, Landais and Leite-Mariante (2024) estimate child penalties in employment for 134 countries. Child penalties exist in almost every country, but their magnitudes vary enormously. For example, employment penalties are small in countries such as China, Haiti, Nigeria, and Portugal, but very large in countries such as Bangladesh, Czech Republic, Jordan, and Mexico. The large variation in child penalties around the world gives large variation in the childhood culture of immigrants.

The analysis divides US immigrants by country of birth (source country). To obtain clean estimates for as many source countries as possible, information on weekly employment (worked last week) and annual employment (worked last year) is pooled. For major source countries where event studies of weekly and annual employment can be conducted separately, the results for pooled employment are very similar (but more precisely estimated). Using pooled employment, the analysis includes 81 source countries where event studies of US immigrants are feasible and where Kleven, Landais and Leite-Mariante (2024) provide estimates of source-country child penalties.[39] The source-country penalties vary from 0% to 64% in the estimation sample.

Figure 14 presents case studies of US immigrants from specific countries. The case studies in-

[39]The 81 countries include Scandinavia (Denmark, Norway, and Sweden) as well as the Czech and Slovak Republics as single units. Besides increasing statistical precision for these smaller source countries, the aggregation of the Czech and Slovak Republics allows for the inclusion of immigrants from the former Czechoslovakia, dissolved as an independent country on December 31, 1992.

clude countries on three different continents — Asia, Latin America, and Africa — and they span a wide range of economic, political, and cultural institutions. Each panel shows an event study of first childbirth for US immigrants born in a given country, and it displays the child penalty for both the immigrants (based on the event study shown) and for people in their country of birth (based on Kleven, Landais and Leite-Mariante 2024). Each row considers a given continent, and within each row, the event studies are sequenced according to the child penalty in country of birth. The relationship between immigrant penalties and birth-country penalties is very strong. For Asian immigrants, as we move from the lowest to the highest birth-country penalty (from Vietnam to Jordan), the child penalty increases from 9% to 69%. For Latin American immigrants, as we move from the lowest to the highest birth-country penalty (from Haiti to Mexico), the child penalty increases from 8% to 45%. And for African immigrants, the pattern is similar: moving from the lowest to the highest birth-country penalty (from Nigeria to Morocco) increases the child penalty from 19% to 50%. Appendix Figure A.21 provides event studies for all 81 source countries in the sample.

Figure 15 pools immigrants from different countries by decile of the child penalty in country of birth. The figure shows event studies for immigrants from the bottom and top deciles, respectively. Again, the findings are striking: the child penalty for US immigrants equals 14% in the bottom decile (where the average birth-country penalty is 3%) and 42% in the top decile (where the average birth-country penalty is 49%).[40] Figure 16 extends these results to the full distribution of birth-country penalties. It plots immigrant penalties by decile of birth-country penalties.[41] Panel A is based on raw child penalty estimates. The relationship between immigrant and birth-country penalties is positive and strong: the slope coefficient of 0.521 implies that, as the employment penalty in a woman's country of birth increases by 10pp, her employment penalty in the US increases by 5.1pp. Because women living in the US are not directly affected by the incentives and institutions of their birth countries, this evidence is most naturally interpreted as an effect of childhood culture on preferences.

As discussed above, epidemiological studies raise concerns about the selection of movers or migrants from different places. The analysis of domestic movers presented results to assuage such concerns. To investigate the selection of foreign immigrants, Panel B of Table 3 compares the

[40] For comparability with the estimated immigrant penalties, the birth-country penalties displayed are weighted averages, where the weight on each country equals its within-decile share of US immigrants in the estimation sample.

[41] Appendix Figure A.22 provides country-level scatter plots of immigrant penalties vs birth-country penalties. These plots show all the country-level penalties used to construct the decile-level penalties presented in Figure 16.

demographic characteristics of mothers who immigrated from countries in the top and bottom quartiles of child penalties. We see selection on two observables: education and race. Mothers from high-penalty countries have less education and a different racial composition (less black, more white) than mothers from low-penalty countries. All other observables are quite similar between the two groups. Importantly, the fact that immigrants are selected on education and race is only a threat to identification if those variables affect child penalties. The heterogeneity analysis in section 5.3 showed that child penalties are unrelated to education, alleviating concerns about selection in this dimension.[42] On the other hand, the heterogeneity analysis also showed that child penalties *are* related to race, and this may act as a confound here.

To address selection on observables, Panel B of Figure 16 controls for differences in education, marriage, race, fertility, age at first birth, and US state of residence (low-penalty vs high-penalty states) across immigrant mothers from different countries.[43] The graph is constructed by regressing child penalties for immigrants on child penalties in birth countries and demographic controls. The immigrant penalties are then residualized using the estimated controls and plotted against birth-country penalties. When plotting the residualized immigrant penalties, the average effect of the estimated controls is added to the residuals to make the levels in Panel B comparable to those in Panel A. The resulting binscatter shows that controlling for observables does not weaken the results. The relationship between immigrant and birth-country penalties is more stable and linear in this specification, and the slope coefficient is about the same. If anything, adjusting for observable differences between immigrants from different countries makes the findings more convincing.

Taken together, the epidemiological studies of foreign immigrants and domestic movers — along with the correlational analysis of elicited gender attitudes — suggest that gender norms and culture are important for explaining child penalties. Given child penalties account for most of the remaining gender inequality in developed countries, this suggests that any additional gender convergence will be hard to achieve without a change in gender norms.

Cultural Assimilation: How persistent are cultural norms? Do immigrants retain their ancestral culture over time or do they assimilate to their surrounding culture? Appendix Figure A.24 provides evidence on cultural assimilation by comparing first-generation and later-generation

[42] Appendix Figure A.23 reproduces this finding for the sample of immigrants.

[43] These control variables are defined as shown in Table 3.

immigrants. First-generation immigrants are defined as foreign-born US residents (those studied above), while later-generation immigrants are defined as US-born residents who report foreign ancestry. The analysis uses information on country of birth (available in ACS data and in CPS data from 1994) and country of ancestry (available in ACS data). Immigrants are divided into quartiles of the child penalty in their country of origin, running the event study specification (6) separately for first-generation and later-generation immigrants within each quartile. The figure shows child penalties for first- and later-generation immigrants in the bottom quartile (Panel A) and in the top quartile (Panel B). The evidence is consistent with strong assimilation effects. Immigrants from low-penalty and high-penalty countries have very different child penalties in the first generation, but they have identical child penalties in later generations. While the convergence is therefore complete, it could take many generations to materialize. The sample of later-generation immigrants includes all descendants with a known country of ancestry, regardless of the time at which their ancestors arrived.

7 Conclusion

A recent literature shows that child penalties account for most of the remaining gender inequality in developed countries (Kleven, Landais and Søgaard 2019; Kleven *et al.* 2019). In other words, eliminating gender inequality is virtually synonymous with eliminating child penalties. Understanding the mechanisms that drive child penalties is therefore very important. This paper contributes methodologically and empirically to this question.

Methodologically, the paper develops a pseudo-event study approach to estimate child penalties using only cross-sectional data. A potential outcomes framework formalizes the assumptions for causal identification in pseudo-event studies.[44] The approach can be validated against a true event study approach using panel data. The two approaches yield virtually identical results, but the cross-sectional approach is much more precise and allows for studying child penalties at granular levels. Furthermore, because of its minimal data requirements, the approach allows for estimating child penalties across most countries of the world and over the long run of history (Kleven, Landais and Leite-Mariante 2024).

Empirically, the paper investigates the variation in child penalties across time, geography, and

[44]The framework also formalizes the assumptions for causal identification in standard event studies based on panel data. Such a framework has been lacking in the child penalty literature.

demographic/cultural groups. There is large variation in these dimensions. The evidence on the effect of social norms is particularly striking. Epidemiological studies of child penalties among domestic movers and foreign immigrants — along with more suggestive evidence using elicited gender norms — show that gender norms are critical for explaining child penalties.

References

ACS. 2000-2019. *American Community Survey 2000-2019.* United States Census Bureau. `https://usa.ipums.org/usa/` (accessed February 13, 2022). 1, 5

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics*. Vol. 3, , ed. O. Ashenfelter and D. Card, Chapter 48. Elsevier: Amsterdam. 4

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113(1): 151–184. 31

Andresen, Martin E., and Emily Nix. 2022. "What Causes the Child Penalty? Evidence from Adopting and Same Sex Couples." *Journal of Labor Economics*, 40(4): 971–1004. 4

Angelov, Nikolay, Per Johansson, and Erica Lindahl. 2016. "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics*, 34(3): 545–579. 4

Bertrand, Marianne. 2011. "New Perspectives on Gender." In *Handbook of Labor Economics*. Vol. 4b, , ed. O. Ashenfelter and D. Card, Chapter 17. Elsevier: Amsterdam. 4

Bertrand, Marianne. 2020. "Gender in the Twenty-First Century." *AEA Papers and Proceedings*, 110: 1–24. 5

Blau, Francine D., and Lawrence M. Kahn. 2006. "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." *Industrial and Labor Relations Review*, 60(1): 45–66. 2, 20

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature*, 55(3): 789–865. 2, 20

Blau, Francine D., Lawrence M. Kahn, and Kerry L. Papps. 2011. "Gender, Source Country Characteristics and Labor Market Assimilation Among Immigrants." *The Review of Economics and Statistics*, 93(1): 43–58. 5

Boelmann, Barbara, Anna Raute, and Uta Schönberg. 2023. "Wind of Change? Cultural Determinants of Maternal Labor Supply." Working Paper. 5, 28

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event Study Designs: Robust and Efficient Estimation." *Review of Economic Studies*. Forthcoming. 4, 12

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2): 200–230. 4, 12

Charles, Kerwin K., Jonathan Guryan, and Jessica Pan. 2022. "The Effects of Sexism on American Women: The Role of Norms vs. Discrimination." *The Journal of Human Resources*. Forthcoming. 5, 28

Cortés, Patricia, and Jessica Pan. 2023. "Children and the Remaining Gender Gaps in the Labor Market." *Journal of Economic Literature*, 61(4): 1359–1409. 1, 2, 4

CPS. 1968-2020. *Current Population Survey 1968-2020.* United States Census Bureau and Bureau of Labor Statistics. `https://cps.ipums.org/cps/` (accessed June 30, 2022). 1, 5

Deaton, Angus. 1985. "Panel Data from Time Series of Cross-Sections." *Journal of Econometrics*, 30(1): 109–126. 4

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996. 4, 12

Fernández, Raquel. 2011. "Does Culture Matter?" In *Handbook of Social Economics*. Vol. 1, , ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, Chapter 11. Elsevier: Amsterdam. 3, 28

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. "Mothers and Sons: Preference Formation and Female Labor Force Dynamics." *Quarterly Journal of Economics*, 119(4): 1249–1299. 4

Fernández, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics*, 1(1): 146–177. 5

Fortin, Nicole. 2005. "Gender Role Attitudes and the Labour-Market Outcomes of Women Across OECD Countries." *Oxford Review of Economic Policy*, 21(3): 416–438. 5

Gallen, Yana, Juanna S. Joensen, Eva R. Johansen, and Gregory F. Veramendi. 2023. "The Labor Market Returns to Delaying Pregnancy." Working Paper. 4

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119. 2, 23

Goldin, Claudia, and Lawrence F. Katz. 2016. "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation." *Journal of Labor Economics*, 34(3): 705–746. 2, 23

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277. 4, 12

GSS. 1972-2018. *General Social Surveys 1972-2018*. NORC at the University of Chicago. `https://gss.norc.org/us/en/gss/get-the-data.html` (accessed March 17, 2019). 3, 26

Kleven, Henrik. 2021. "Lecture 3: Public Policy and Child Penalties." Zeuthen Lecture Series, September 2021. 25

Kleven, Henrik. 2024. "The EITC and the Extensive Margin: A Reappraisal." *Journal of Public Economics*, 236: 1–28. 5

Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante. 2024. "The Child Penalty Atlas." *Review of Economic Studies*. Forthcoming. 2, 3, 4, 21, 31, 32, 34, 41, 54, 55, 79, 80, 81, 83

Kleven, Henrik, Camille Landais, and Jakob E. Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics*, 11(4): 181–209. 1, 4, 8, 11, 13, 23, 34

Kleven, Henrik, Camille Landais, and Jakob E. Søgaard. 2021. "Does Biology Drive Child Penalties? Evidence from Biological and Adoptive Families." *American Economic Review: Insights*, 3(2): 183–198. 1, 3, 4, 24

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2019. "Child Penalties Across Countries: Evidence and Explanations." *AEA Papers and Proceedings*, 109: 122–126. 1, 2, 4, 5, 13, 21, 26, 28, 34

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2024. "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation." *American Economic Journal: Economic Policy*, 16: 110–149. 1, 4

Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington. 2018. "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?" NBER Working Paper No. 24740. 2, 20

Lundborg, Petter, Erik Plug, and Astrid W. Rasmussen. 2017. "Can Women Have Children and a Career? IV Evidence from IVF Treatments." *American Economic Review*, 107(6): 1611–1637. 4

Melentyeva, Valentina, and Lukas Riedel. 2023. "Child Penalty Estimation and Mothers' Age at First Birth." Working Paper. 12

Moriconi, Simone, and Núria Rodríguez-Planas. 2021. "Gender Norms and the Motherhood Employment Gap." Working Paper. 5

NLSY. 1979-2018. *National Longitudinal Survey of Youth 1979-2018.* United States Department of Labor, produced and distributed by the Center for Human Resource Research (CHRR), Ohio State University. https://www.nlsinfo.org/investigator (accessed July 31, 2022). 2, 6

PSID. 1968-2019. *Panel Study of Income Dynamics 1968-2019.* University of Michigan Institute for Social Research. https://psidonline.isr.umich.edu/ (accessed April 6, 2022). 2, 6

Rabaté, Simon, and Sara Rellstab. 2022. "What Determines the Child Penalty in the Netherlands? The Role of Policy and Norms." *De Economist*, 170(2): 195–229. 5

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics*, 235(2): 2218–2244. 4, 6, 12

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199. 4, 12

Table 1: Descriptive Statistics in the Cross-Section

	Men			Women		
	Child	No Child	**Difference**	Child	No Child	**Difference**
Annual Employment Rate	0.89	0.79	**0.10**	0.71	0.80	**-0.09**
Weekly Employment Rate	0.91	0.75	**0.15**	0.68	0.75	**-0.07**
Earnings	53,254	28,650	**24,604**	23,796	24,943	**-1,147**
Fraction High School or Below	0.43	0.44	**-0.01**	0.41	0.32	**0.09**
Fraction College	0.30	0.25	**0.05**	0.28	0.34	**-0.06**
Fraction Married	0.87	0.25	**0.62**	0.72	0.34	**0.39**
Fraction Black	0.07	0.11	**-0.04**	0.11	0.11	**0.00**
Fraction White	0.72	0.67	**0.04**	0.67	0.70	**-0.03**
Fraction Hispanic	0.14	0.13	**0.01**	0.15	0.11	**0.04**
Age	38.63	32.55	**6.08**	37.28	32.90	**4.38**
Cohort	1967.00	1974.43	**-7.43**	1968.44	1973.92	**-5.48**
Number of Observations	9,901,305	11,468,329		13,247,471	9,085,312	

Notes: This table compares labor market and demographic outcomes for men and women with and without children in cross-sectional data. The sample includes all individuals aged 20-50 in all years of the pooled CPS and ACS data.

Table 2: Descriptive Statistics in the Pseudo-Panel

	Matched Men			Matched Women		
	$\tau = 0$	$\tau = -1$	**Difference**	$\tau = 0$	$\tau = -1$	**Difference**
Annual Employment Rate	0.92	0.91	**0.01**	0.72	0.87	**-0.15**
Weekly Employment Rate	0.93	0.90	**0.03**	0.69	0.83	**-0.14**
Earnings	55,136	49,102	**6,034**	29,846	36,820	**-6,974**
Fraction High School or Below	0.26	0.26	**0.00**	0.17	0.17	**0.00**
Fraction College	0.47	0.47	**-0.00**	0.57	0.57	**-0.00**
Fraction Married	0.88	0.88	**-0.00**	0.85	0.85	**-0.00**
Fraction Black	0.04	0.04	**0.00**	0.05	0.05	**0.00**
Fraction White	0.80	0.80	**-0.00**	0.77	0.77	**-0.00**
Fraction Hispanic	0.10	0.10	**0.00**	0.09	0.09	**0.00**
Age at First Birth	31.79	31.79	**0.00**	30.60	30.60	**0.00**
Age	31.79	30.79	**1.00**	30.60	29.60	**1.00**
Cohort	1974.56	1974.56	**0.00**	1976.21	1976.21	**0.00**
Number of Observations	246,763	246,763		244,376	244,376	

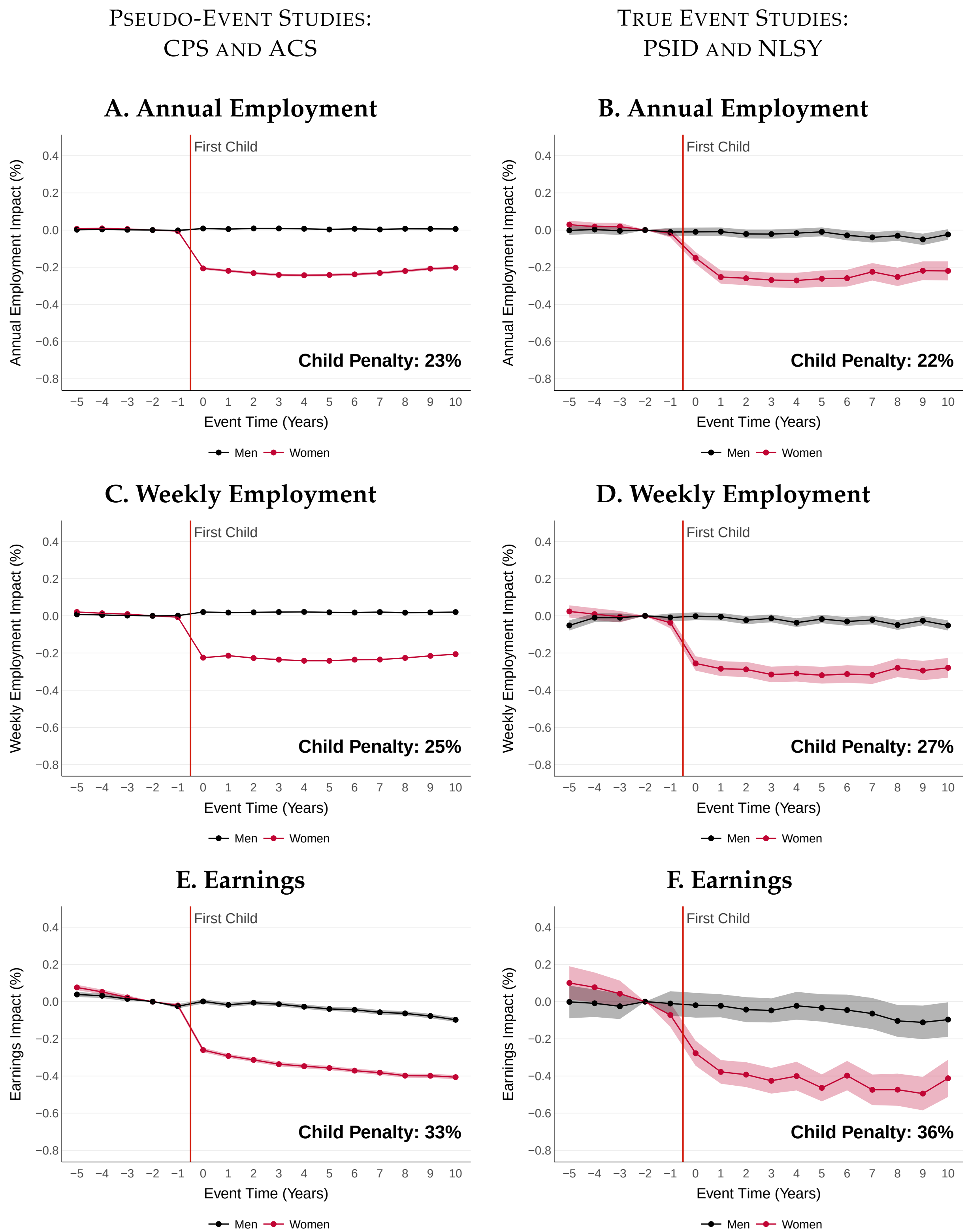
Notes: This table compares labor market and demographic outcomes for matched men and women at event times $\tau = 0$ and $\tau = -1$ in the pseudo-panel. By construction, individuals at event time $\tau = 0$ are exactly one year older and born in the same cohort as those at event time $\tau = -1$. Also by construction, individuals at $\tau = 0$ and $\tau = -1$ match exactly on all demographic characteristics, but not on labor market outcomes. The sample includes all matched parents at $\tau = 0$ (together with their matched non-parents at $\tau = -1$) with an age at first birth between 25-45 in all years of the pooled CPS and ACS data.

TABLE 3: SELECTION OF MOVERS AND IMMIGRANTS BY PLACE OF BIRTH

	A. Movers by State of Birth			B. Immigrants by Country of Birth		
	High-Penalty States	Low-Penalty States	**Difference**	High-Penalty Countries	Low-Penalty Countries	**Difference**
Fraction in High-Penalty States	0.25	0.18	**0.07**	0.19	0.20	**-0.01**
Fraction High School or Below	0.12	0.11	**0.01**	0.47	0.30	**0.17**
Fraction College	0.61	0.63	**-0.02**	0.34	0.49	**-0.16**
Fraction Married	0.84	0.84	**0.00**	0.82	0.86	**-0.04**
Fraction Black	0.04	0.09	**-0.05**	0.02	0.17	**-0.15**
Fraction White	0.91	0.86	**0.05**	0.66	0.15	**0.51**
Fertility	1.78	1.76	**0.02**	1.72	1.67	**0.06**
Age at First Birth	31.39	31.33	**0.06**	30.65	31.18	**-0.53**
Age	37.59	37.60	**0.00**	36.53	36.91	**-0.38**
Cohort	1973.20	1972.98	**0.22**	1972.83	1973.14	**-0.31**
Number of Observations	95,437	77,971		191,017	114,672	

Notes: This table provides evidence on the selection of US movers by state of birth (Panel A) and US immigrants by country of birth (Panel B). Movers are defined as US-born individuals living in a different state than where they were born, while immigrants are foreign-born individuals living in the US. Each group is divided by the child penalty in their place of birth (top vs bottom quartile of child penalties in US states and foreign countries, respectively). The child penalties used to split movers by state of birth are annual employment penalties in the sample of stayers (as presented in Figure A.18), while the child penalties used to split immigrants by country of birth are taken from Kleven, Landais and Leite-Mariante (2024). The table shows demographic characteristics for mothers. The mover sample is based on ACS 2000-2019 (where state of birth is observed). The immigrant sample is based on ACS 2000-2019 and CPS 1994-2020 (where country of birth is observed), including foreign-born individuals from any of the countries shown in Figure A.21.

FIGURE 1: VALIDATION OF PSEUDO-EVENT STUDY APPROACH



Notes: This figure validates the pseudo-event study approach (left panels) against a true event study approach (right panels). The pseudo-event studies are based on pooled CPS and ACS data from 1968-2020, while the true event studies are based on pooled PSID and NLSY data from 1968-2019. Each panel shows an event study for men and women around the birth of their first child at $\tau = 0$. The series show the percentage impact of childbirth on men and women at each event time τ, i.e. $\hat{P}^m_\tau$ and $\hat{P}^w_\tau$ defined in equation (7). Each panel also displays the average child penalty over event times 0-10 defined in equation (8). Three labor market outcomes are shown: annual employment, weekly employment, and earnings. Age at first birth is restricted to be between ages 25-45. The 95% confidence intervals are based on robust standard errors.

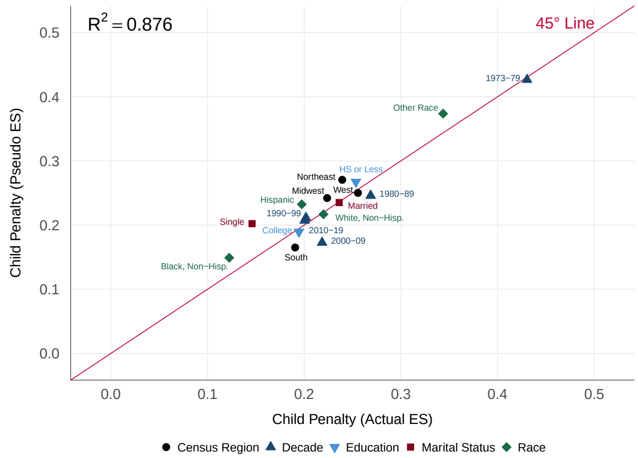
FIGURE 2: WITHIN-PANEL VALIDATION OF PSEUDO-EVENT STUDY APPROACH

PSEUDO-EVENT STUDIES: PSID AND NLSY | TRUE EVENT STUDIES: PSID AND NLSY

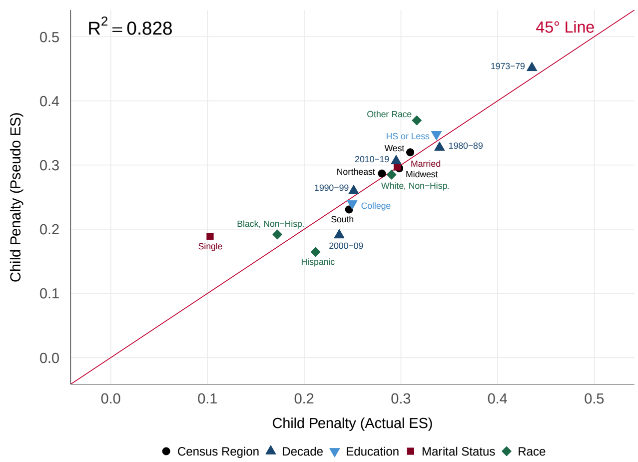
Notes: This figure validates the pseudo-event study approach (left panels) against a true event study approach (right panels), both using pooled PSID and NLSY data from 1968-2019. Each panel shows an event study for men and women around the birth of their first child at $\tau = 0$. The series show the percentage impact of childbirth on men and women at each event time τ, i.e. $\hat{P}^m_\tau$ and $\hat{P}^w_\tau$ defined in equation (7). Each panel also displays the average child penalty over event times 0-10 defined in equation (8). Three labor market outcomes are shown: annual employment, weekly employment, and earnings. Age at first birth is restricted to be between ages 25-45. The 95% confidence intervals are based on robust standard errors.

FIGURE 3: WITHIN-PANEL VALIDATION IN SUBSAMPLES

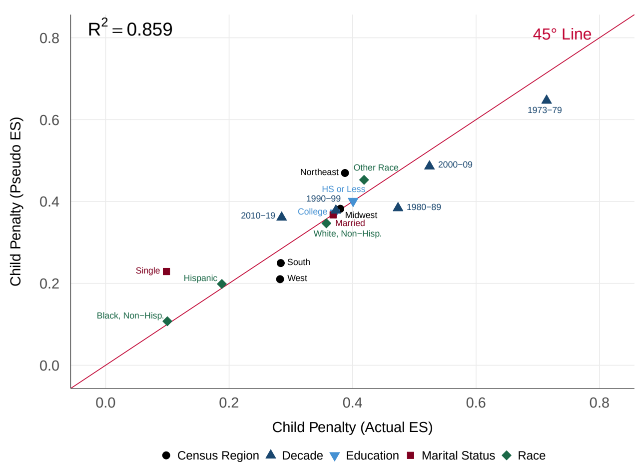
A. Annual Employment



B. Weekly Employment

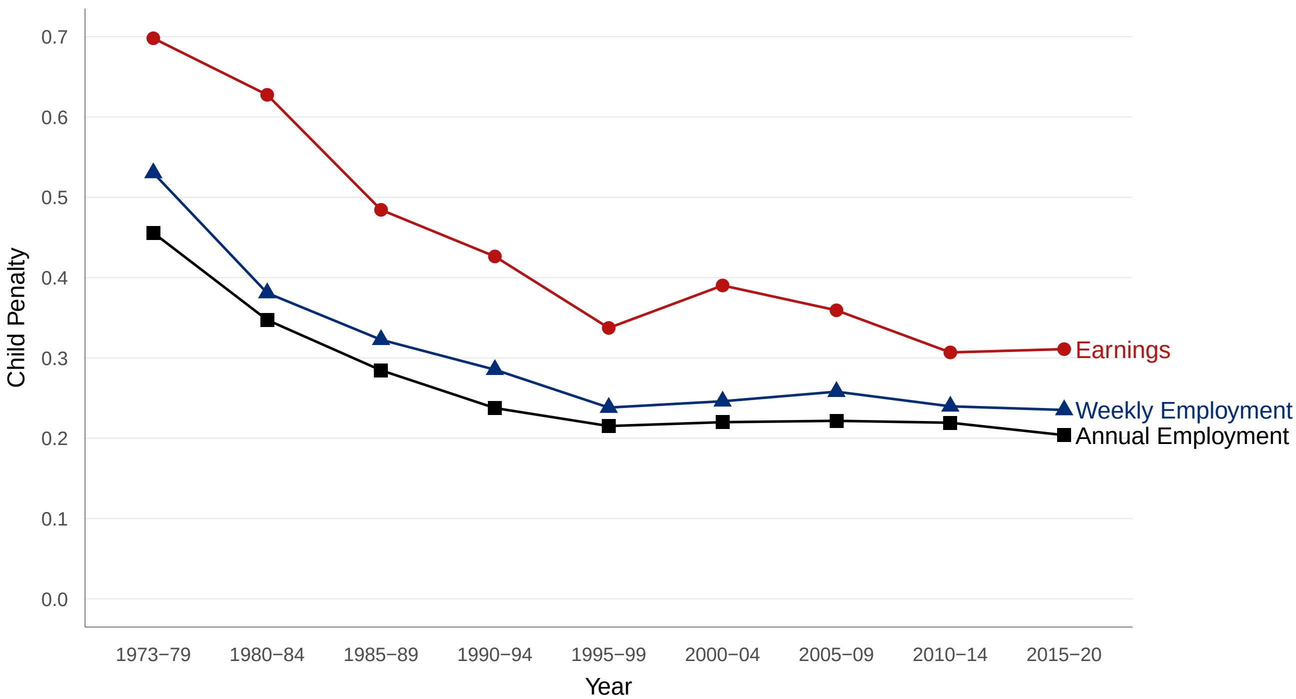


C. Earnings



Notes: This figure validates the pseudo-event study approach in subsamples using pooled PSID/NLSY data. For each labor market outcome, the figure plots child penalties estimated from pseudo-event studies against child penalties estimated from true event studies in subsamples. The sample is split by geography (4 census regions), time (5 decades), education (high school or less vs college), marital status (single vs married), and race (4 categories). For all outcomes and subsamples, the child penalty pairs lie close to 45-degree line. The R-squared from a regression of pseudo-panel estimates on panel estimates lies between 0.83-0.88 across the three outcomes. This suggests that the pseudo-event study approach remains valid in subsamples.

FIGURE 4: CHILD PENALTIES OVER TIME

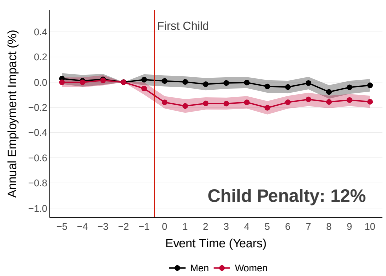


Notes: This figure shows the evolution of child penalties in each of the three labor market outcomes over time. Each series depicts the average child penalty over event times 0-10 (defined in equation 8) in different time intervals. These are estimated by splitting the sample of parents by interview year and running the event study specification (6) separately for each time period. The child penalty series start in 1973, because the first five years of the data (1968-1972) are reserved for obtaining synthetic pre-birth observations for those who had their first child in 1973. The underlying event studies for each time period and labor market outcome are presented in Appendix Figures A.6-A.8.

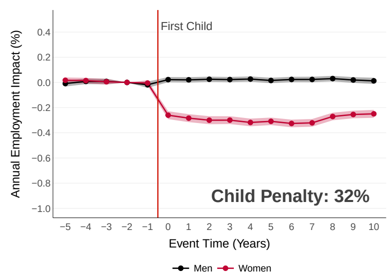
Figure 5: Event Studies of First Childbirth in Selected States

Annual Employment

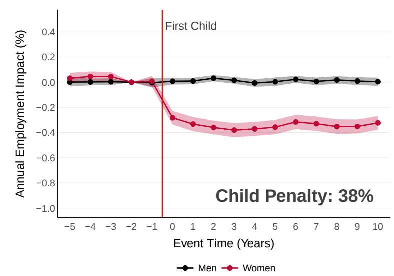
A: North Dakota



B: New Jersey

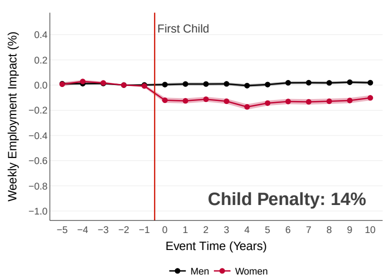


C: Utah

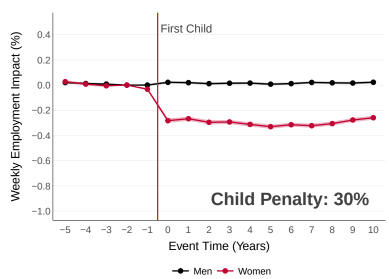


Weekly Employment

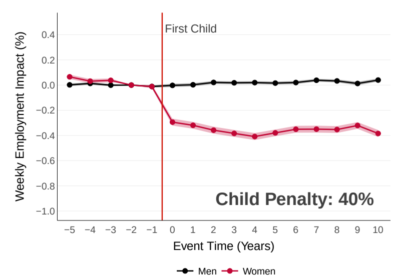
D: North Dakota



E: New Jersey

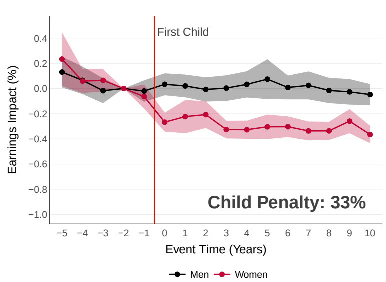


F: Utah

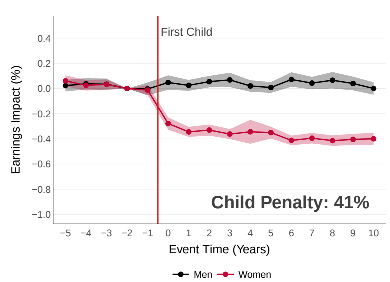


Earnings

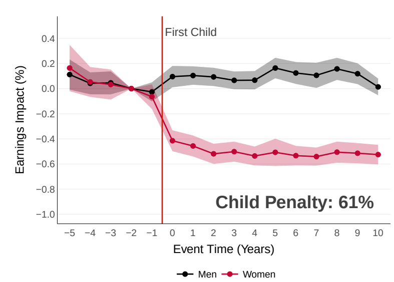
G: North Dakota



H: New Jersey



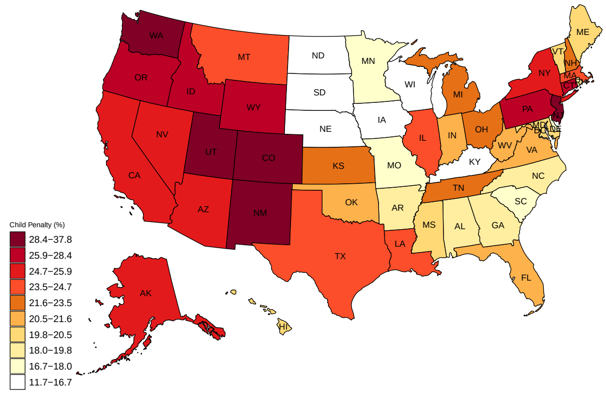
I: Utah



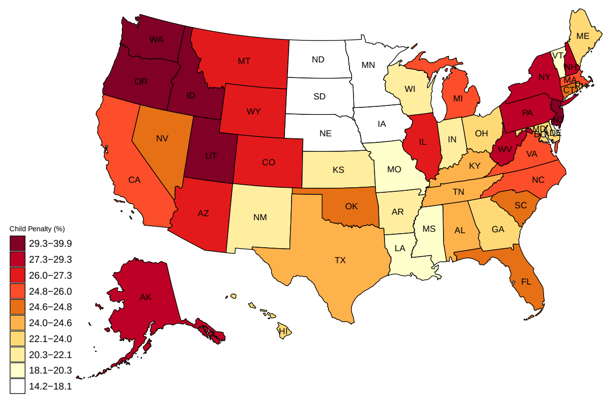
Notes: This figure shows event studies of first childbirth for three US states and each of the three labor market outcomes. State-level event studies are constructed by interacting the event time dummies in equation (6) with state dummies, estimating percentage impacts of childbirth on men and women at each event time ($\hat{P}^m_\tau$ and $\hat{P}^w_\tau$) as well as average child penalties over event times 0-10 separately for each state. In this specification, the lifecycle and time trends in equation (6) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors. Event studies for all 51 states (including the federal district of D.C.) and all three labor market outcomes are provided in Appendix Figures A.11-A.13.

Figure 6: Heatmaps of Child Penalties

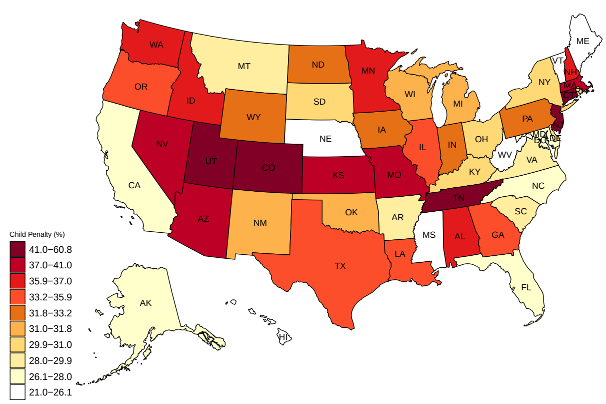
A. Annual Employment



B. Weekly Employment

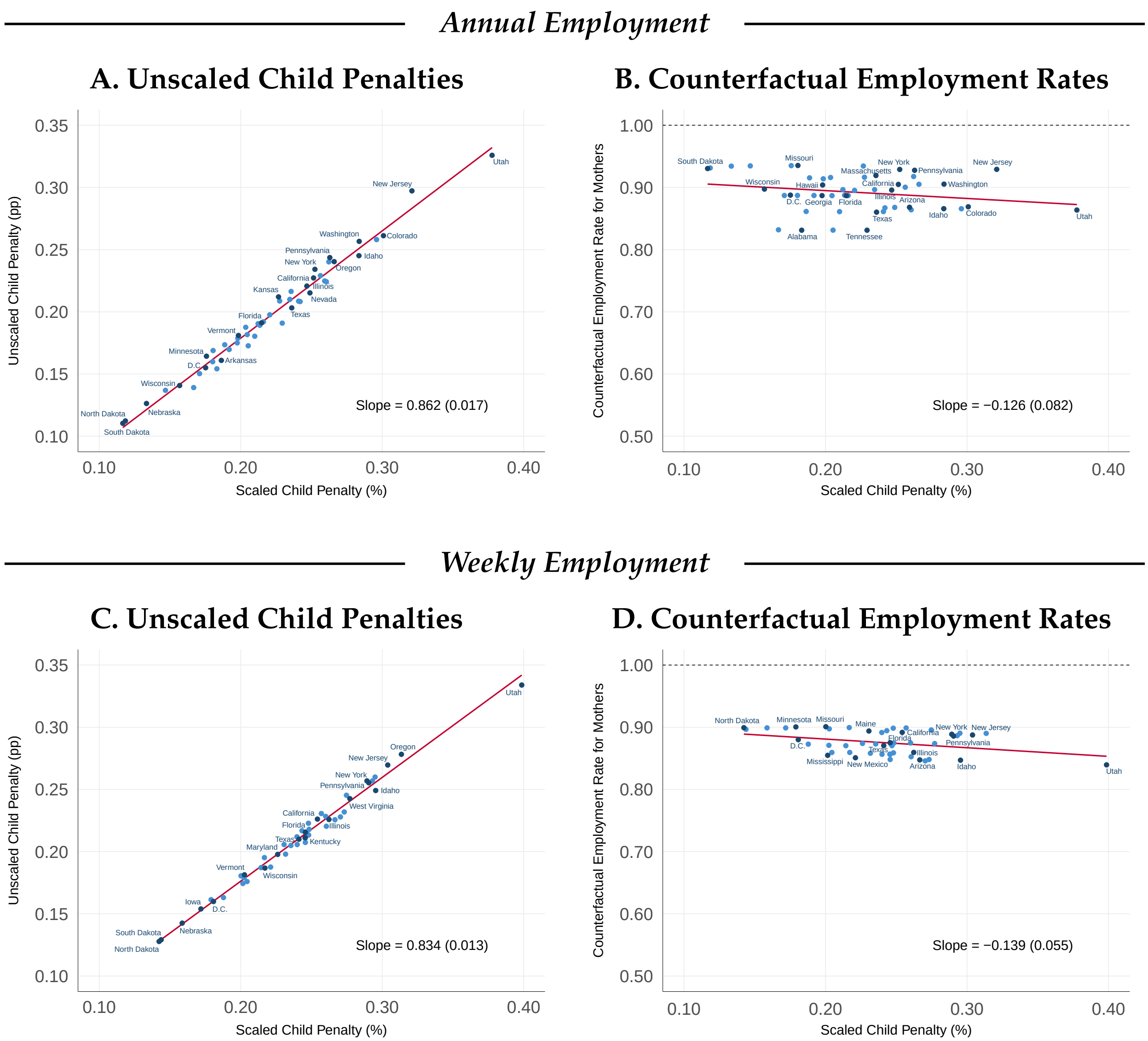


C. Earnings



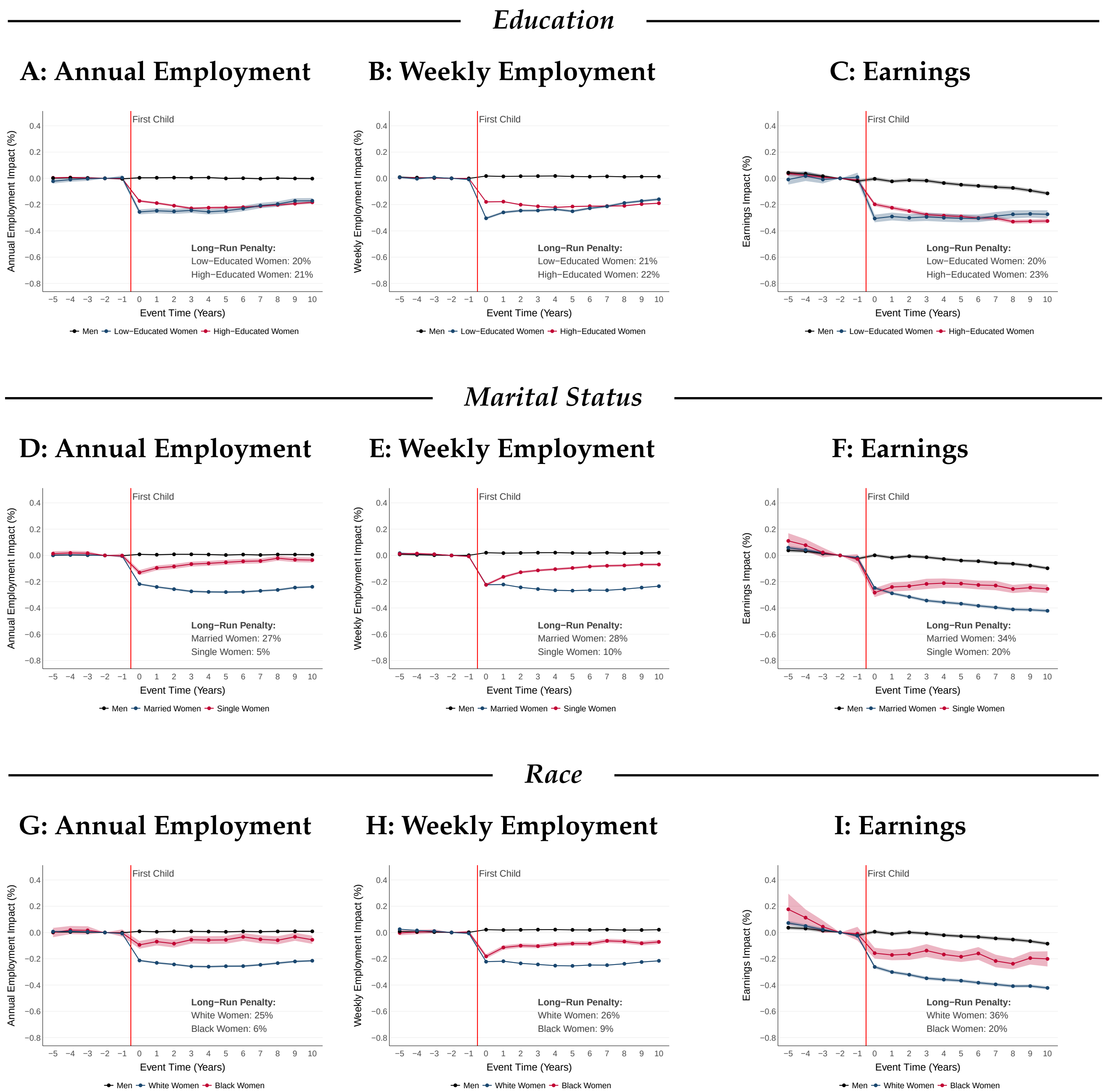
Notes: This figure summarizes the results from the state-level event studies of childbirth (shown in Figures A.11-A.13 of the appendix) in heatmaps. In these maps, states are divided into deciles of the child penalty (as defined in equation 8), with darker colors implying larger child penalties.

FIGURE 7: UNSCALED VS SCALED CHILD PENALTIES ACROSS STATES



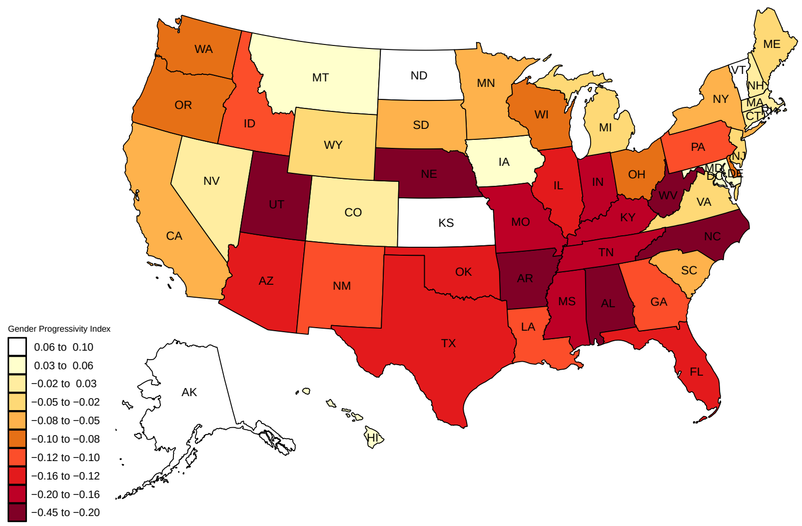
Notes: This figure investigates if the spatial variation in scaled child penalties (effects in percentages) reflects mostly variation in unscaled child penalties (effects in absolute terms) or in the counterfactual levels used for scaling. Results are shown for the two employment outcomes. The left panels plot unscaled child penalties against scaled child penalties across states, while the right panels plot counterfactual employment rates for mothers against scaled child penalties across states. The counterfactual employment rate is calculated as the predicted outcome from equation (6) when omitting the contribution of the event time coefficients. The displayed child penalties and counterfactual employment rates are averages over event times 0-10. The figure shows that the spatial variation is driven almost exclusively by variation in the effect of children in absolute terms. The relationship between scaled and unscaled penalties is almost perfectly linear with a slope close to one, whereas the counterfactual employment rate is almost flat across states.

FIGURE 8: CHILD PENALTIES ACROSS DEMOGRAPHIC GROUPS



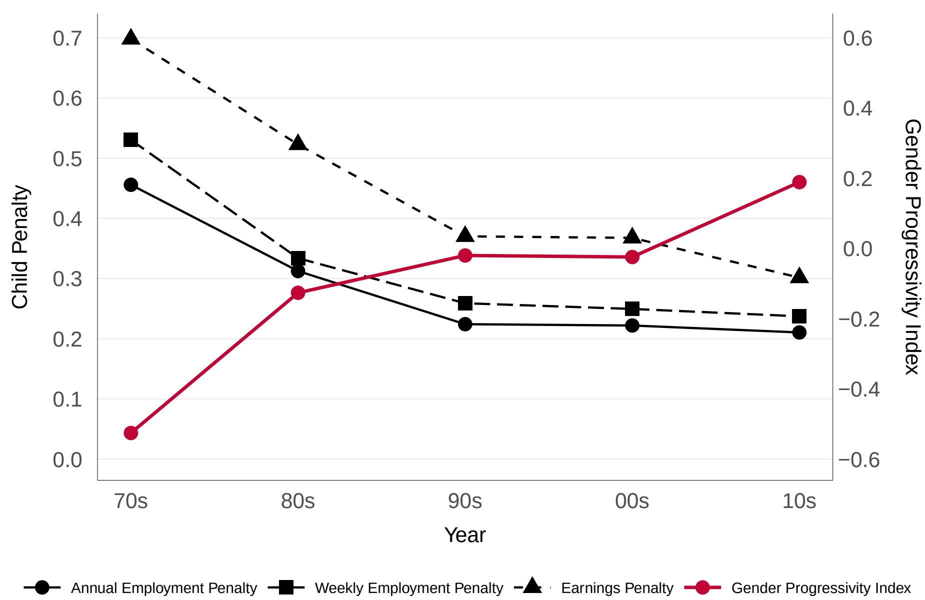
Notes: This figure presents event studies of first childbirth by education, marital status, and race. To construct the figure, the sample of women is split into different demographic groups and specification (6) is estimated separately for each group. The sample of men is not split by demographics as childbirth is always a non-event for them. Low-educated women are defined as those with a high school degree or less, while high-educated women are those with a college degree or more. Single women include all unmarried women (never married, separated, divorced, or widowed). Results are shown for each of the three labor market outcomes, and the long-run child penalty (over event times 5-10) is displayed for each outcome. The 95% confidence intervals are based on robust standard errors.

Figure 9: Heatmap of Gender Norms



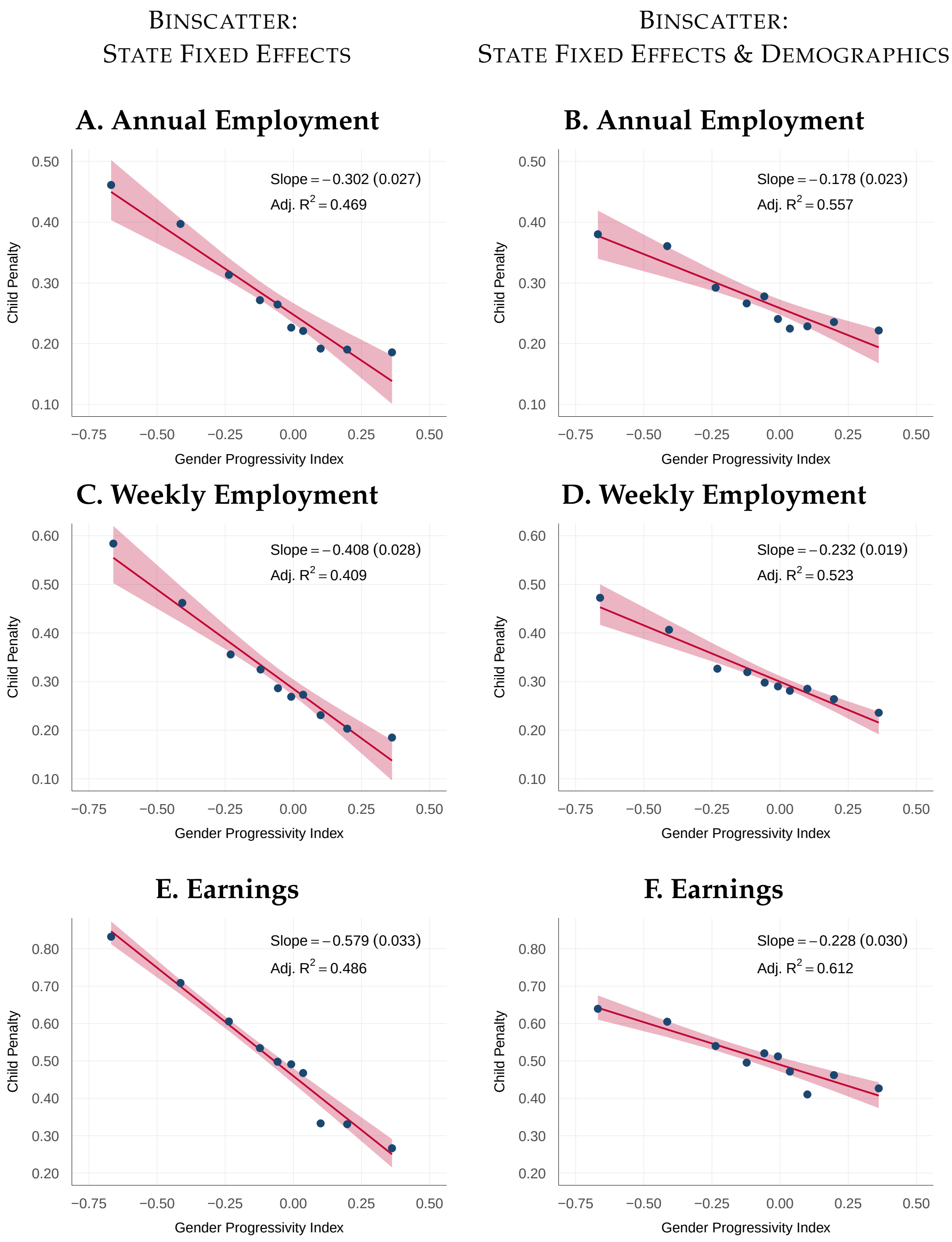
Notes: This figure presents a heatmap of gender norms using GSS data from 1972-2018. States are divided into deciles of a Gender Progressivity Index (GPI). This index is calculated as the average standardized response to GSS questions that elicit attitudes towards gender roles in families with children. A higher value of GPI (lighter colors) corresponds to a more gender progressive norm.

Figure 10: Child Penalties vs Gender Norms Over Time



Notes: This figure plots the evolution of child penalties and gender progressivity over the last 50 years. The construction of the Gender Progressivity Index (GPI) is described in the notes to the preceding figure. The GPI time series is obtained by taking an average of state-level GPIs within each decade, weighting different states according to their share of the US population in 2019.

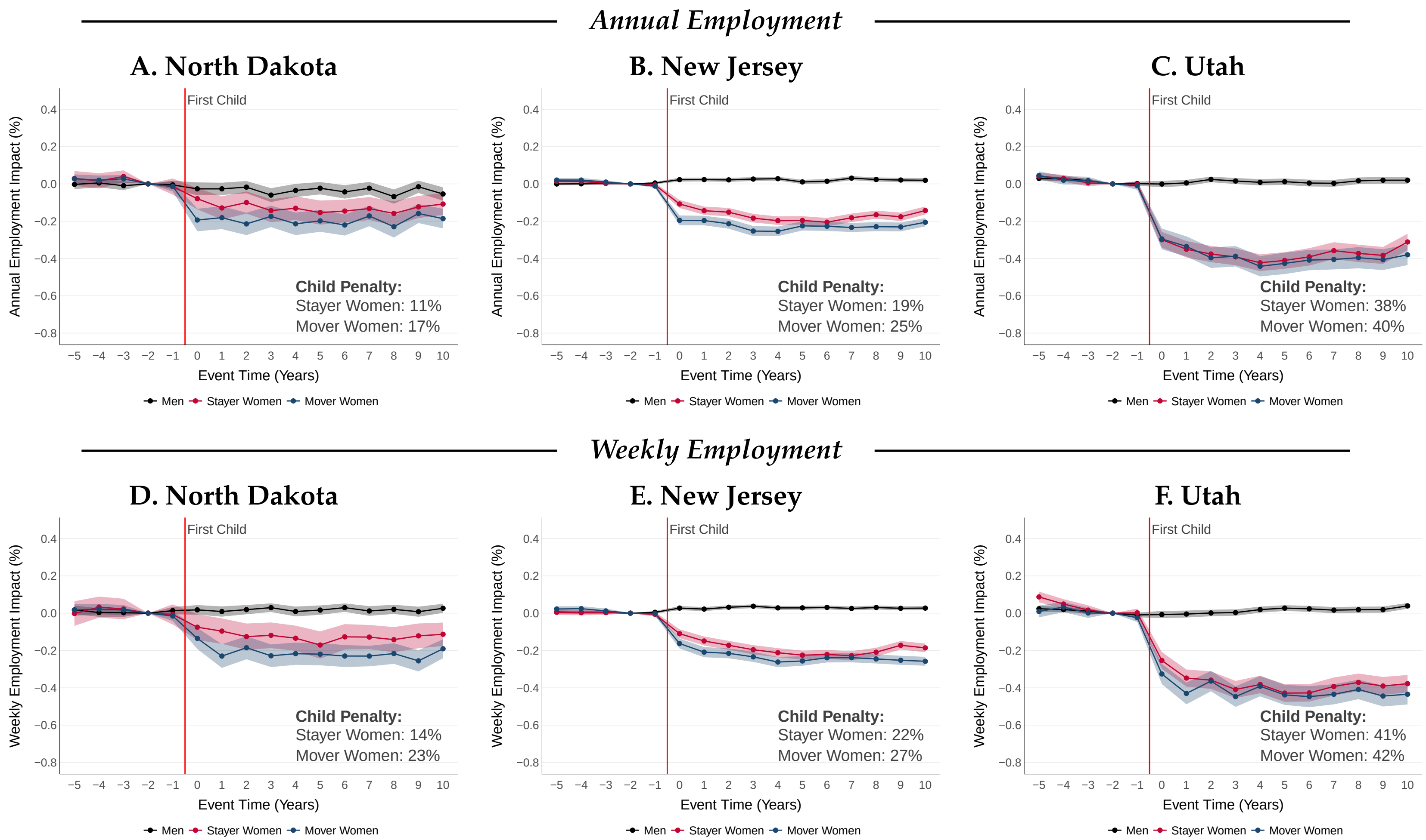
FIGURE 11: CHILD PENALTIES VS GENDER NORMS ACROSS STATES AND TIME



Notes: This figure provides binscatters of child penalties vs gender progressivity across states and time. The analysis is based on equation (14), i.e. regressing the child penalty by state and time on the Gender Progressivity Index (GPI) by state and time, controlling for state fixed effects and time-varying demographics (education, marital status, and race). Each panel plots residualized child penalties (i.e., net of the effect of controls) by decile of the GPI. When plotting residualized child penalties by bin of the GPI, the average effect of the controls is added to the residuals such that the level of the outcome is comparable across panels with different controls. The left panels control only for state fixed effects, while the right panels control both for state fixed effects and time-varying demographics. Given the standardization of GPI, the slope coefficient in each panel can be interpreted as the effect of increasing GPI by one standard deviation.

Figure 12: Epidemiological Study of US Movers

Event Studies of First Childbirth for Movers vs Stayers by State of Birth



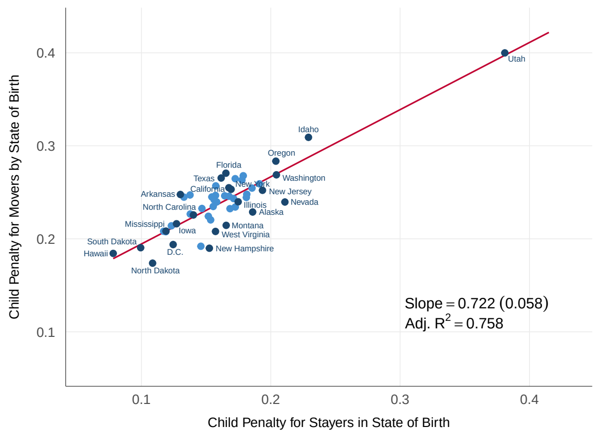
Notes: This figure presents event studies of first childbirth for movers and stayers born in different states. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. To construct the figure, specification (6) is run separately for women movers and women stayers, interacting the event time dummies by state-of-birth dummies. The sample of men is not split by mover/stayer status as childbirth is a non-event for them regardless of status. Each panel displays child penalties over event times 0-10 for mover women and stayer women with a given state of birth (North Dakota, New Jersey, or Utah) and in a given outcome (annual or weekly employment). The 95% confidence intervals are based on robust standard errors. The sample is based on ACS data from 2000-2019, which contains information on both state of residence and state of birth. Event studies for movers and stayers for all states of birth and in both employment outcomes are provided in Appendix Figures A.18-A.19.

Figure 13: Epidemiological Study of US Movers

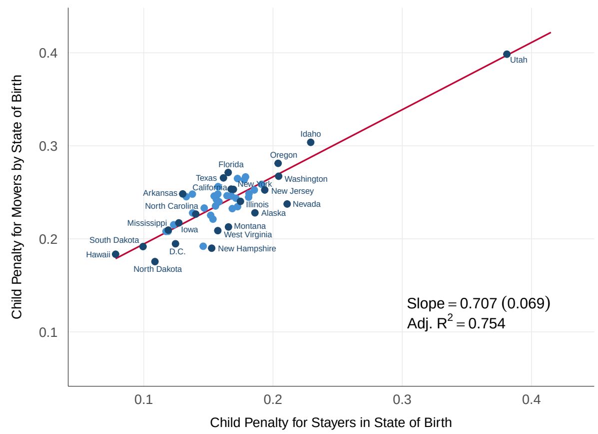
Child Penalties for Movers vs Stayers by State of Birth

Annual Employment

A. Raw Data

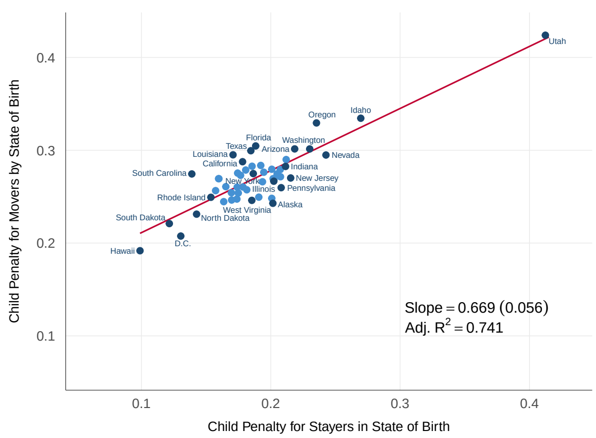


B. Residualize by Effect of State-of-Residence Penalties

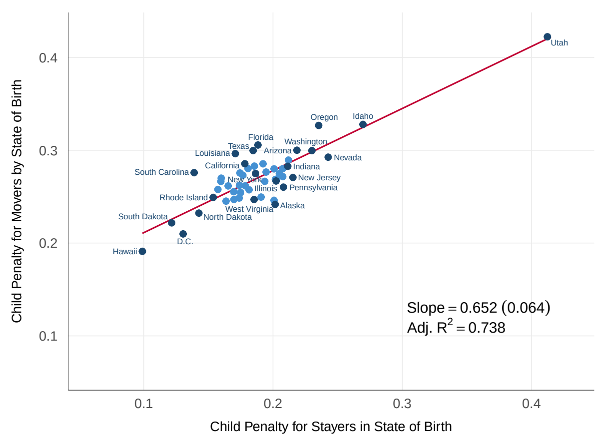


Weekly Employment

C. Raw Data



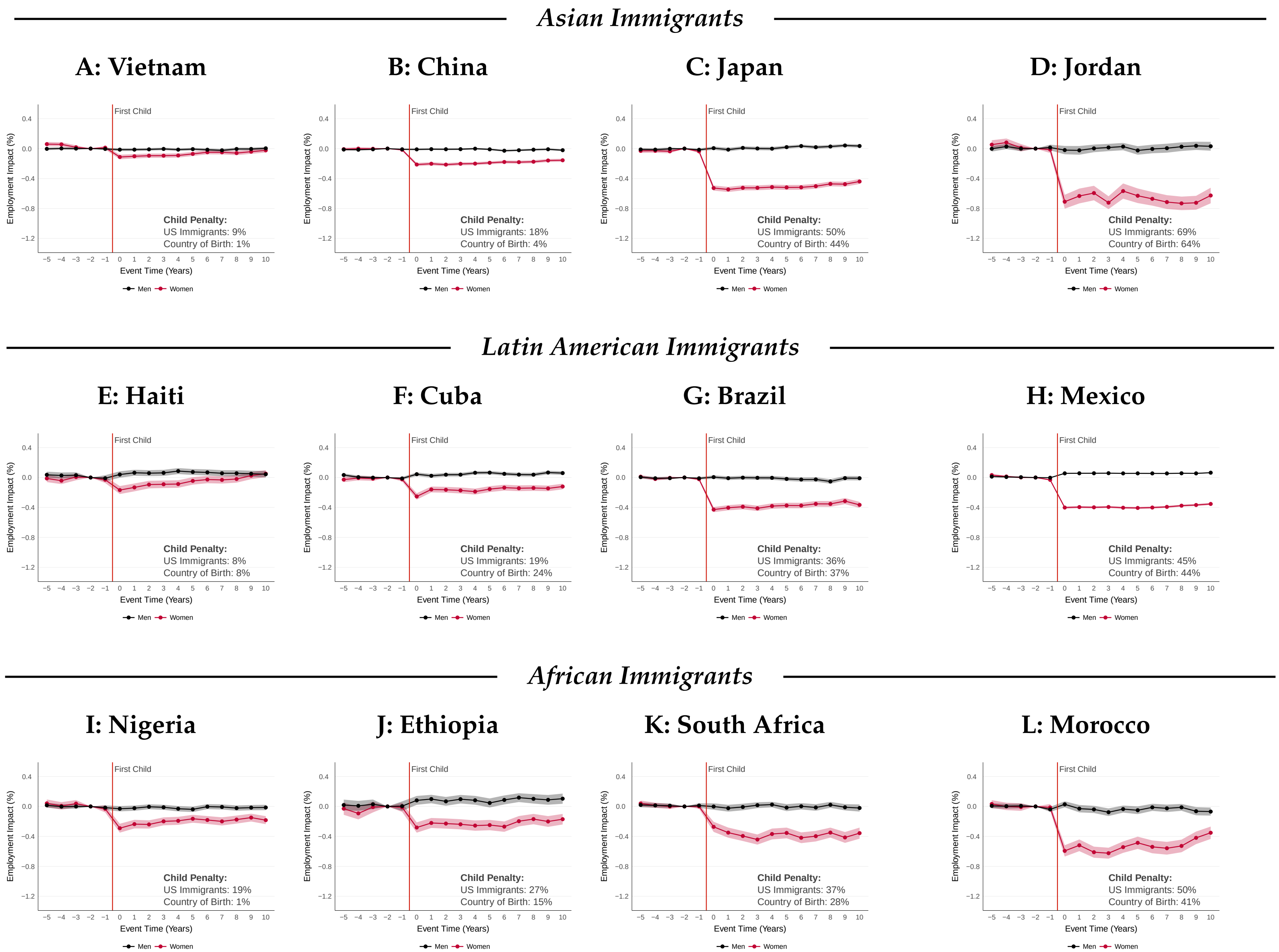
D. Residualize by Effect of State-of-Residence Penalties



Notes: This figure provides scatter plots of the child penalty for movers against the child penalty for stayers by state of birth. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. The left panels show raw child penalties, while the right panels show residualized child penalties using the specification in eq. (15). The residualized plots control for selection on state of residence, which would otherwise contaminate the estimated effects of state of birth (norms/culture) with effects of state of residence (local labor markets). The figure shows that place of birth has very strong effects and that these are robust to controlling for selection on state of residence. The underlying event studies used to estimate the child penalties for movers and stayers in each state and for each outcome are presented in Appendix Figures A.18-A.19. The sample is based on ACS data from 2000-2019, which contains information on both state of residence and state of birth.

Figure 14: Epidemiological Study of Foreign Immigrants

Event Studies of First Childbirth for Immigrants by Country of Birth

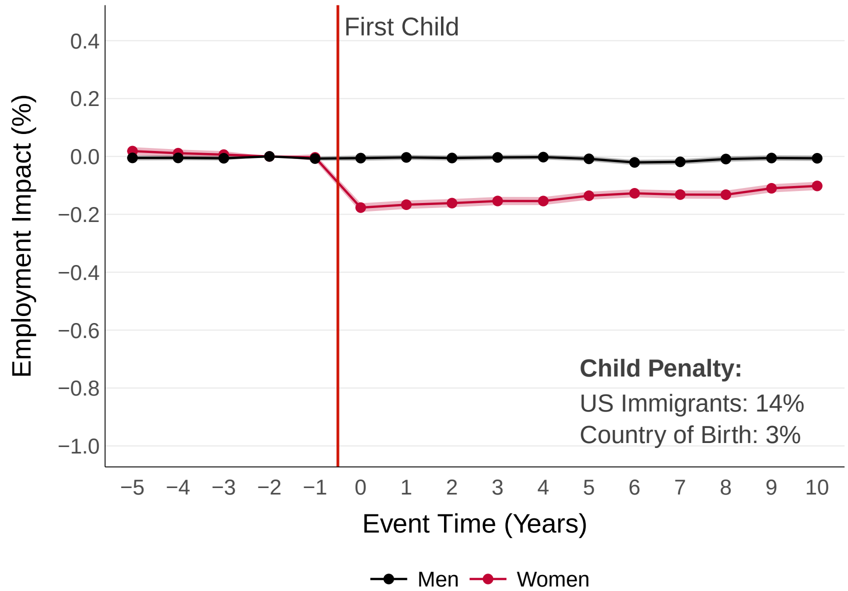


Notes: This figure presents event studies of first childbirth for foreign-born immigrants by country of birth. Results are shown for selected countries in Asia (top row), Latin America (middle row), and Africa (bottom row). The results for all 81 countries in the sample are provided in Appendix Figure A.21. Each panel displays the child penalty for US immigrants (based on the series shown) and the child penalty in country of birth (based on Kleven, Landais and Leite-Mariante 2024), ordering panels by the child penalty in country of birth. The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020. The 95% confidence intervals are based on robust standard errors.

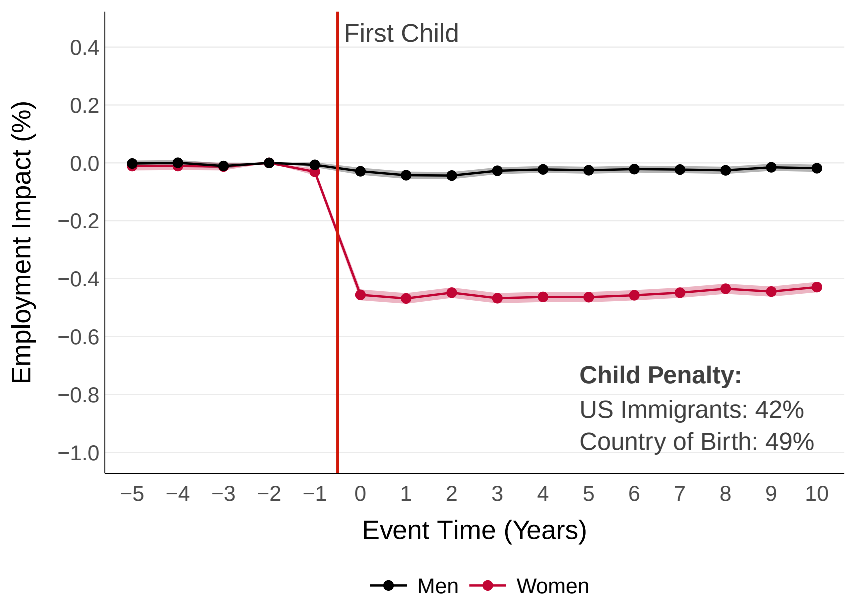
Figure 15: Epidemiological Study of Foreign Immigrants

Event Studies of First Childbirth for Immigrants by Decile of Birth-Country Penalties

A. Bottom Decile of Child Penalty in Country of Birth



B. Top Decile of Child Penalty in Country of Birth

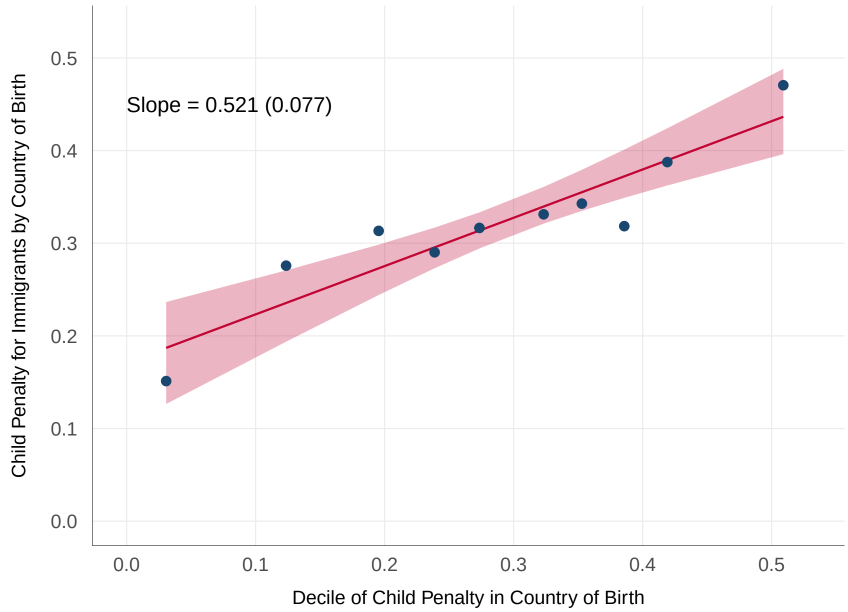


Notes: This figure presents event studies of first childbirth for foreign-born immigrants in the bottom and top deciles of birth-country penalties. Countries are assigned to deciles using the child penalty estimates in Kleven, Landais and Leite-Mariante (2024) for the sample of 81 countries shown in Appendix Figure A.21. The figure is constructed by running the event study specification (6) separately for each decile, graphing the percentage impacts on men and women at each event time τ (as defined in equation 7). Each panel displays the average child penalty for US immigrants (based on the series shown) along with the average child penalty in country of birth. To make the two child penalty estimates comparable, the average birth-country penalty is weighted by each country's share of US immigrants within each decile of the data. The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020. The 95% confidence intervals are based on robust standard errors.

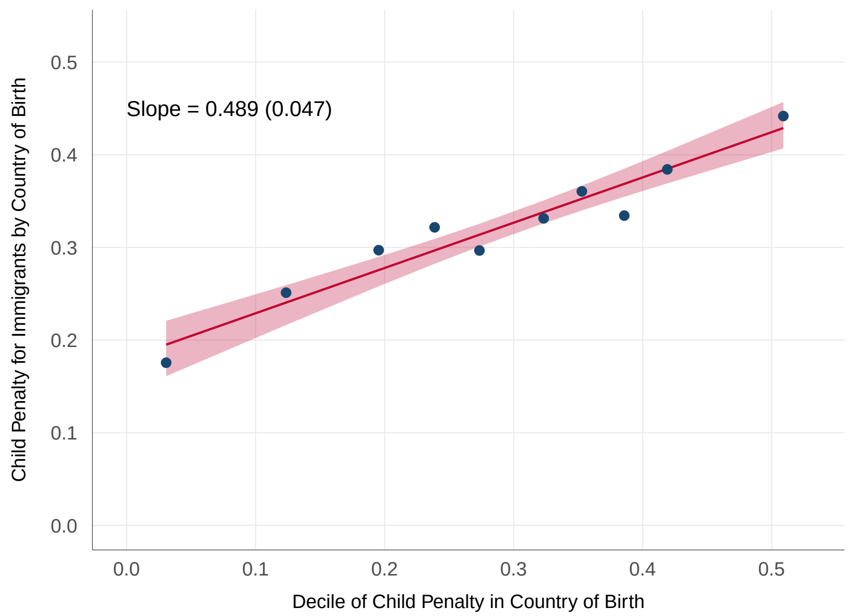
FIGURE 16: EPIDEMIOLOGICAL STUDY OF FOREIGN IMMIGRANTS

CHILD PENALTIES FOR IMMIGRANTS BY DECILE OF BIRTH-COUNTRY PENALTIES

A. Raw Data



B. With Controls



Notes: This figure presents binscatters of child penalties for foreign-born immigrants by decile of the child penalty in country of birth. Panel A shows raw child penalty estimates, while Panel B controls for differences in education, marriage, race, fertility, age at first birth, and US location across immigrant mothers from different countries. To construct Panel B, immigrant penalties are regressed on birth-country penalties and demographic controls, residualizing the immigrant penalties by the estimated effect of the controls for each country. The average effect of the controls across all countries is added to the residualized outcome to make the levels in Panel A and B comparable. The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020.

Online Appendix:

"The Geography of Child Penalties and Gender Norms:
A Pseudo-Event Study Approach"

Henrik Kleven
Princeton University and NBER

August 2025

A Supplementary Exhibits

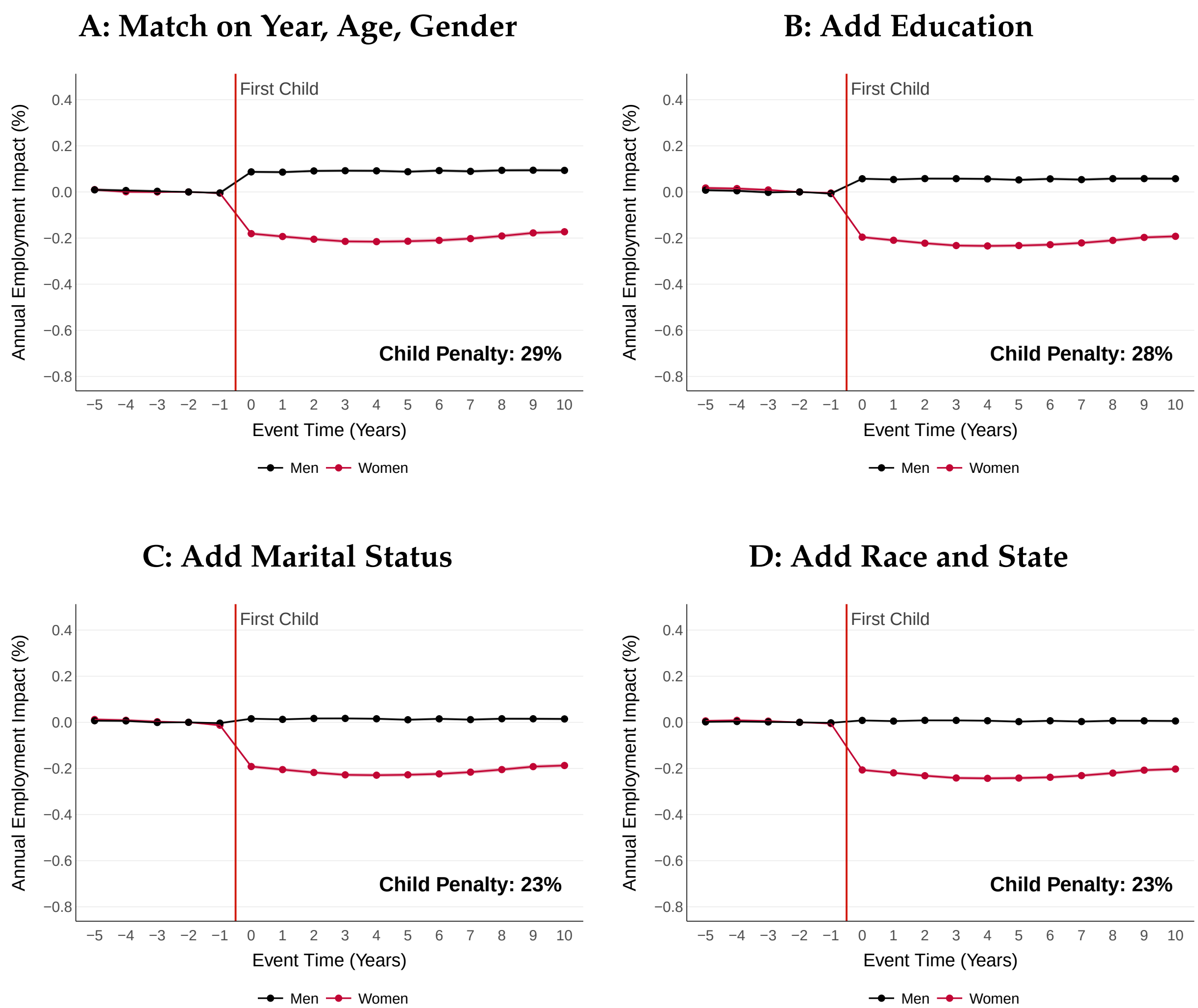
TABLE A.1: TEST OF ASSUMPTION 3 IN PANEL DATA
BALANCE BETWEEN ACTUAL AND SYNTHETIC OBSERVATIONS OF NEGATIVE EVENT TIMES

	Panel: Observed Values	**Pseudo-Panel:** Synthetic Values	**Difference**
Panel A: Men			
Annual Employment Rate			
Event Time τ = -2	0.929 (0.007)	0.917 (0.003)	-0.013 (0.008)
Event Times τ < 0	0.927 (0.004)	0.914 (0.002)	-0.013 (0.004)
Weekly Employment Rate			
Event Time τ = -2	0.923 (0.007)	0.915 (0.003)	-0.009 (0.008)
Event Times τ < 0	0.909 (0.004)	0.914 (0.002)	0.005 (0.005)
Earnings (Normalized)			
Event Time τ = -2	0.681 (0.019)	0.670 (0.010)	-0.011 (0.021)
Event Times τ < 0	0.675 (0.012)	0.678 (0.009)	0.003 (0.016)
Panel B: Women			
Annual Employment Rate			
Event Time τ = -2	0.939 (0.008)	0.940 (0.004)	0.000 (0.010)
Event Times τ < 0	0.949 (0.004)	0.942 (0.003)	-0.007 (0.005)
Weekly Employment Rate			
Event Time τ = -2	0.898 (0.010)	0.881 (0.005)	-0.017 (0.011)
Event Times τ < 0	0.896 (0.005)	0.882 (0.003)	-0.014 (0.006)
Earnings (Normalized)			
Event Time τ = -2	0.728 (0.023)	0.699 (0.011)	-0.029 (0.024)
Event Times τ < 0	0.742 (0.013)	0.704 (0.009)	-0.038 (0.015)
Panel C: Men Minus Women			
Annual Employment Rate			
Event Time τ = -2	-0.010 (0.010)	-0.023 (0.005)	-0.013 (0.012)
Event Times τ < 0	-0.022 (0.005)	-0.028 (0.003)	-0.006 (0.006)
Weekly Employment Rate			
Event Time τ = -2	0.025 (0.013)	0.034 (0.005)	0.008 (0.014)
Event Times τ < 0	0.013 (0.006)	0.032 (0.003)	0.019 (0.007)
Earnings (Normalized)			
Event Time τ = -2	-0.047 (0.028)	-0.029 (0.015)	0.018 (0.032)
Event Times τ < 0	-0.067 (0.018)	-0.026 (0.012)	0.041 (0.021)

Notes: This table tests Assumption 3 using pooled PSID/NLSY data. This assumption requires balance between actual and synthetic observations at negative event times, where the synthetic observations are obtained from the matching algorithm described in section 3.1. The table reports mean outcomes in the panel and pseudo-panel, as well as the differences between the two. Outcomes are shown for men (Panel A), women (Panel B), and men relative to women (Panel C). Each panel includes three outcomes: annual employment, weekly employment, and earnings (normalized by average earnings across all years, for men and women separately). Results are shown for event time $\tau = -2$ (omitted base year) and for the average of all negative event times. The findings support the empirical approach: the differences between the panel and pseudo-panel are small and mostly statistically insignificant. Standard errors are based on bootstrapping with replacement.

FIGURE A.1: PSEUDO-EVENT STUDIES WITH DIFFERENT MATCHING VARIABLES

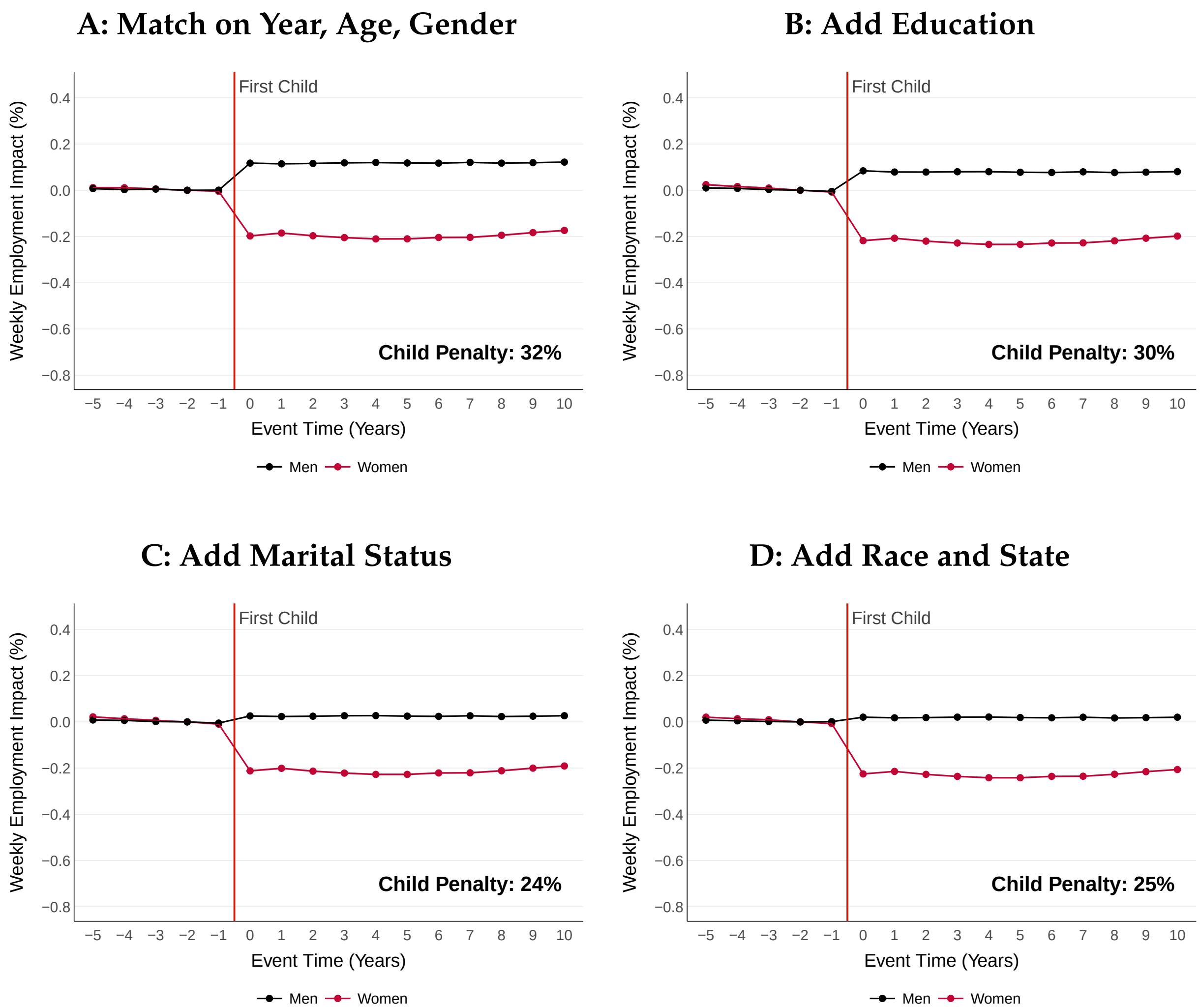
ANNUAL EMPLOYMENT



Notes: This figure presents pseudo-event studies of first childbirth for annual employment based on increasingly granular matching specifications. Panel A matches only on year, age, and gender, Panel B adds education, Panel C adds marital status, and Panel D adds race and state of residence. Panel D corresponds to the baseline specification presented in Figure 1. The more parsimonious specifications in Panels A-C are associated with selection bias, evidenced by the positive jumps for men between event times $\tau = -1$ and $\tau = 0$ as well as the discrepancy between these specifications and the true event study in Figure 1. The baseline specification in Panel D eliminates these selection problems.

FIGURE A.2: PSEUDO-EVENT STUDIES WITH DIFFERENT MATCHING VARIABLES

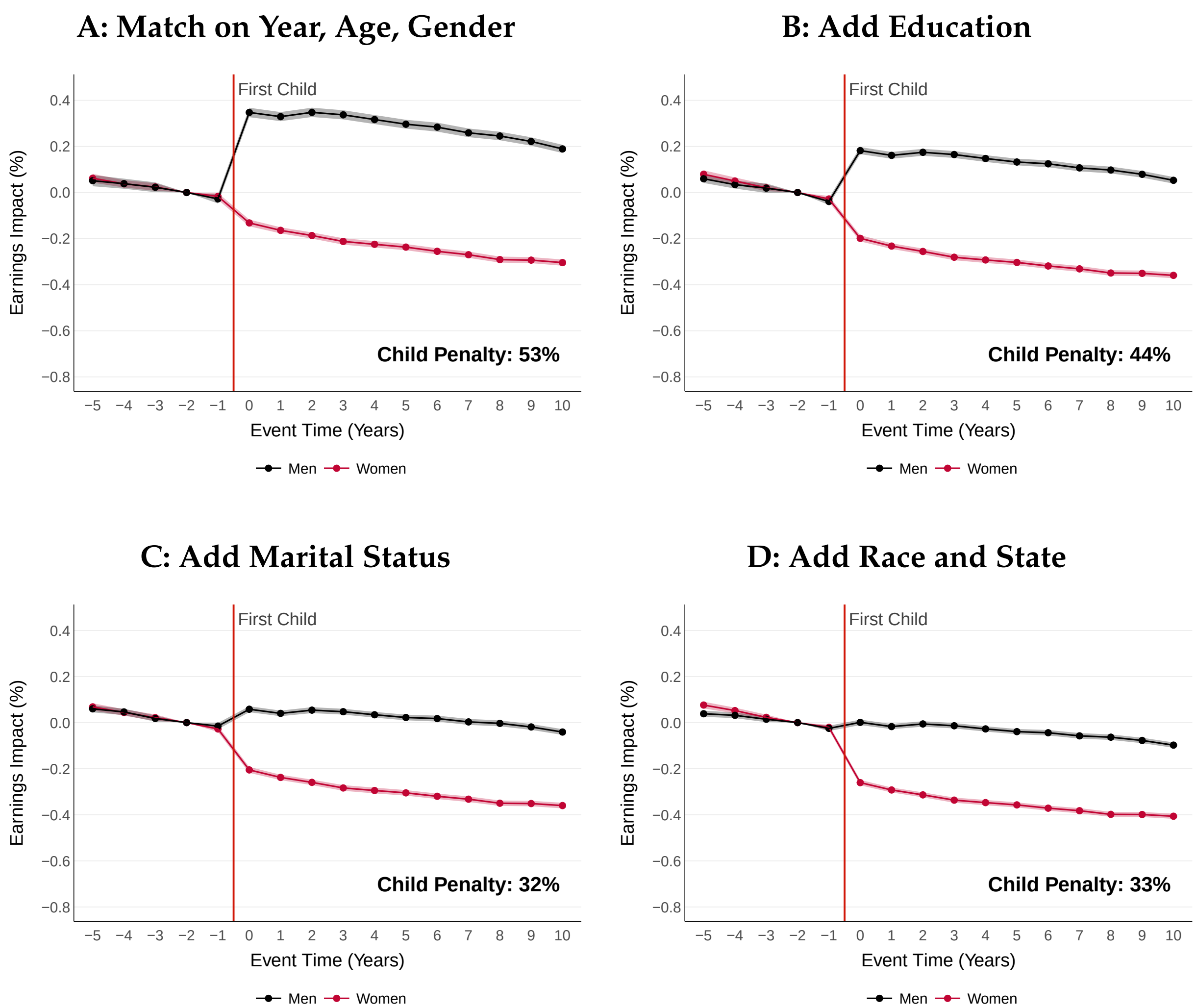
WEEKLY EMPLOYMENT



Notes: This figure presents pseudo-event studies of first childbirth for weekly employment based on increasingly granular matching specifications. Panel A matches only on year, age, and gender, Panel B adds education, Panel C adds marital status, and Panel D adds race and state of residence. Panel D corresponds to the baseline specification presented in Figure 1. The more parsimonious specifications in Panels A-C are associated with selection bias, evidenced by the positive jumps for men between event times $\tau = -1$ and $\tau = 0$ as well as the discrepancy between these specifications and the true event study in Figure 1. The baseline specification in Panel D eliminates these selection problems.

Figure A.3: Pseudo-Event Studies With Different Matching Variables

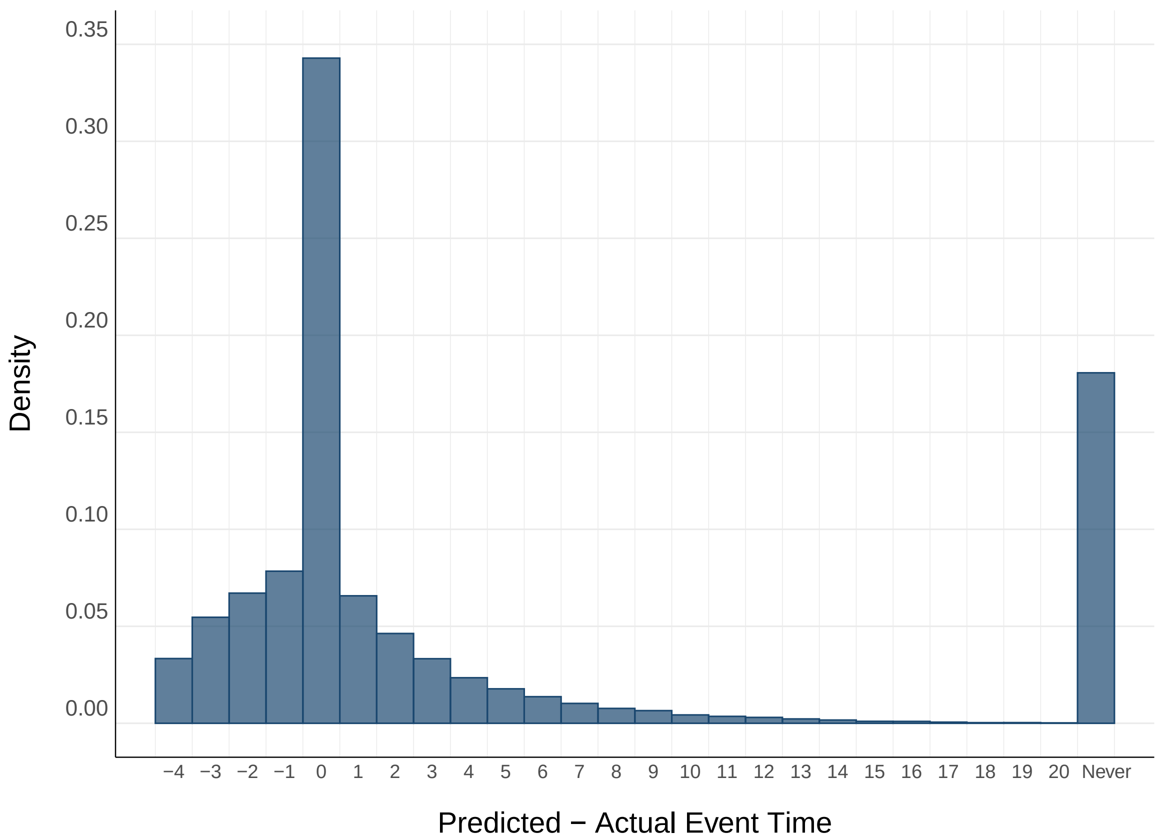
Earnings



Notes: This figure presents pseudo-event studies of first childbirth for earnings based on increasingly granular matching specifications. Panel A matches only on year, age, and gender, Panel B adds education, Panel C adds marital status, and Panel D adds race and state of residence. Panel D corresponds to the baseline specification presented in Figure 1. The more parsimonious specifications in Panels A-C are associated with selection bias, evidenced by the positive jumps for men between event times $\tau = -1$ and $\tau = 0$ as well as the discrepancy between these specifications and the true event study in Figure 1. The baseline specification in Panel D eliminates these selection problems.

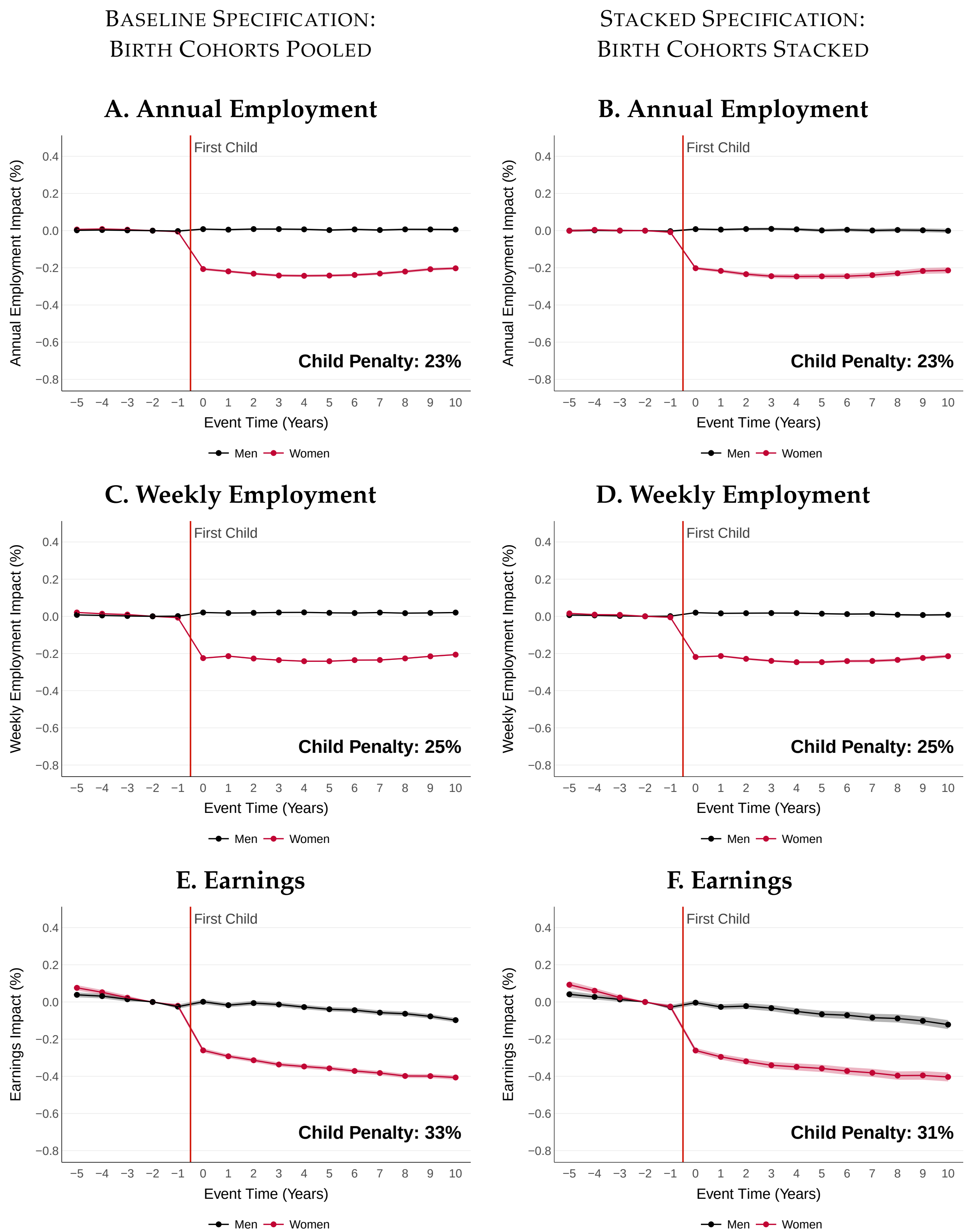
FIGURE A.4: QUALITY OF FERTILITY PREDICTION IN PSEUDO-EVENT STUDY APPROACH

PREDICTED VS ACTUAL EVENT TIMES AMONG CHILDLESS PEOPLE



Notes: This figure shows the distribution of within-person differences in predicted and actual event times among those observed without children. The distribution is based on panel data from PSID and NLSY between 1968-2019, sampling individuals observed after age 45 for whom completed fertility can be measured. Predicted event times for childless individuals are based on the matching specification used in the pseudo-event study approach (these event times vary from -5 to -1), while the actual event times for the same individuals are directly observed in the panel data. Event time is perfectly predicted for 34% of the data and with an error of less than four years for 74% of the data. The bin labeled “never” includes matched individuals (assigned to event times between -5 and -1) who never have children.

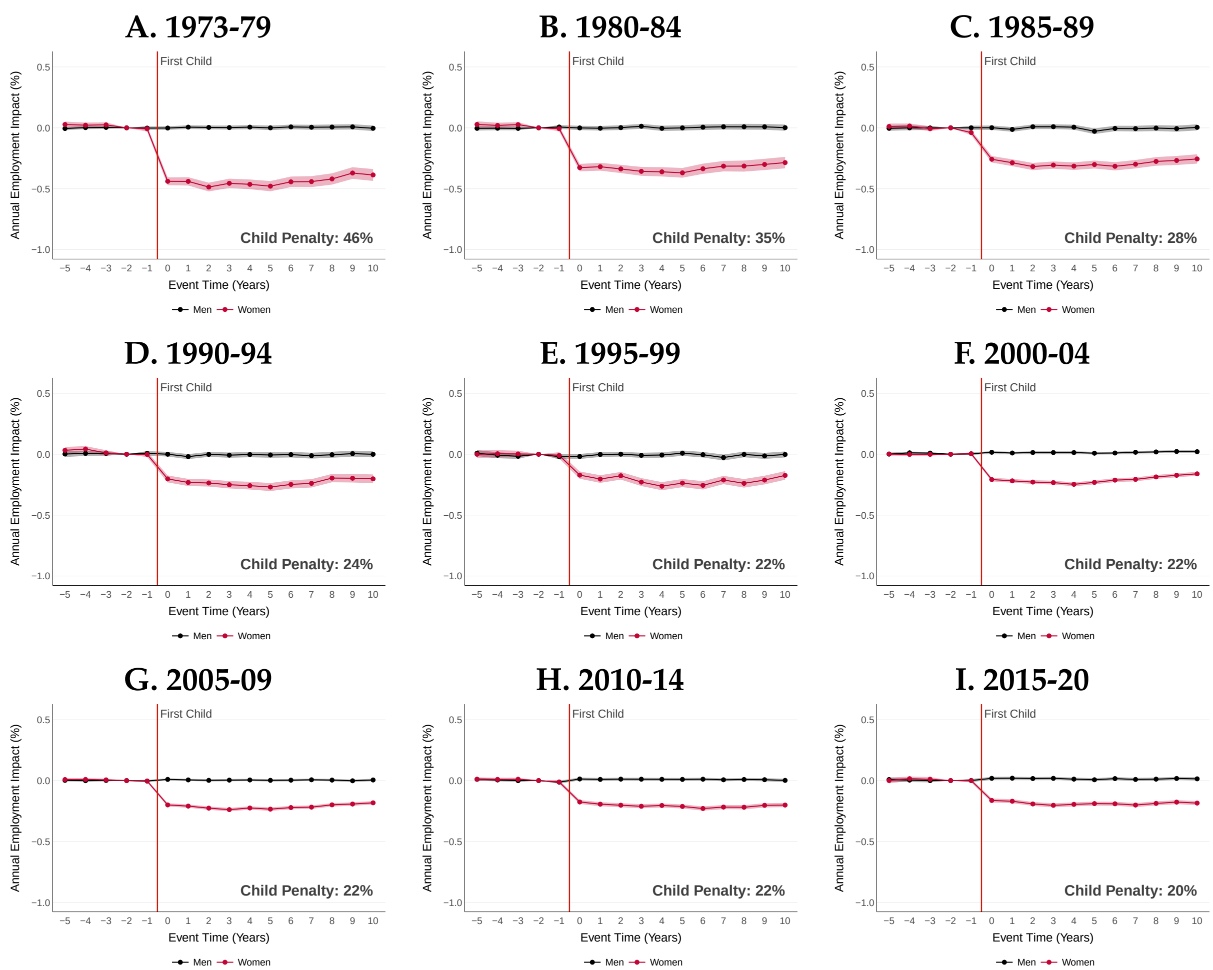
Figure A.5: Robustness to Heterogeneous Treatment Effects



Notes: This figure investigates the possibility of bias from treatment-effect heterogeneity by comparing results from the baseline event study specification (pooling all birth cohorts) to results from a stacked event study specification (stacking different birth cohorts). Specifically, the stacked specification allows for heterogeneous effects by age at first birth (as specified in equation 9) and calculates a weighted average treatment effect using the sample shares of each cohort (as specified in equation 10). The baseline and stacked specifications produce almost identical results in all three labor market outcomes, indicating that heterogeneous treatment effects do not create bias in this context.

Figure A.6: Event Studies of First Childbirth Over Time

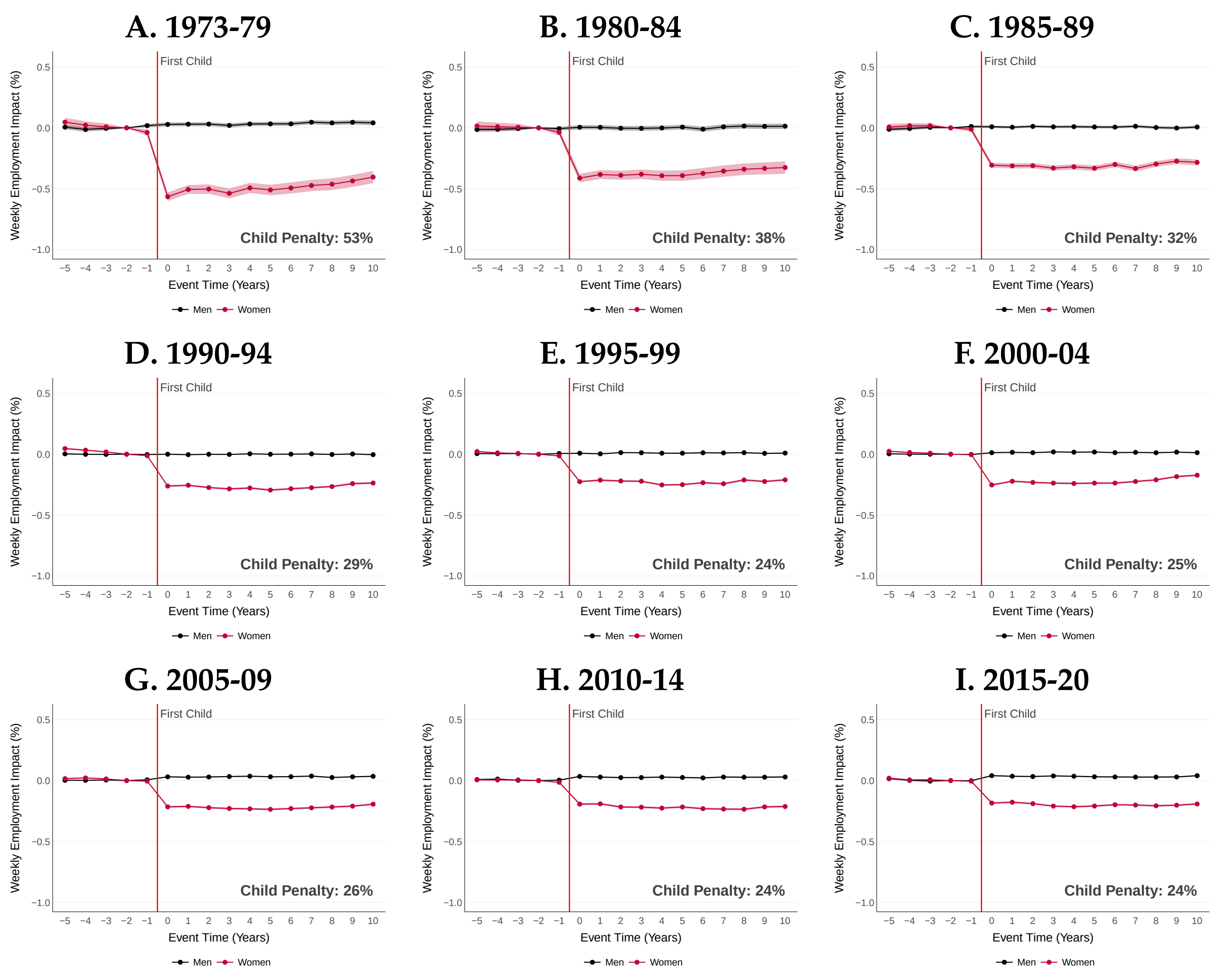
Annual Employment



Notes: This figure shows event studies of first childbirth for annual employment in different time periods. The sample of parents is split by interview year and the event study specification (6) is run separately for each time period. The event studies start in 1973, because the first five years of the data (1968-1972) are reserved for obtaining synthetic pre-birth observations for those who had their first child in 1973. Each panel displays the average child penalty over event times 0-10 (defined in equation 8) for the time period in question. The 95% confidence intervals are based on robust standard errors.

FIGURE A.7: EVENT STUDIES OF FIRST CHILDBIRTH OVER TIME

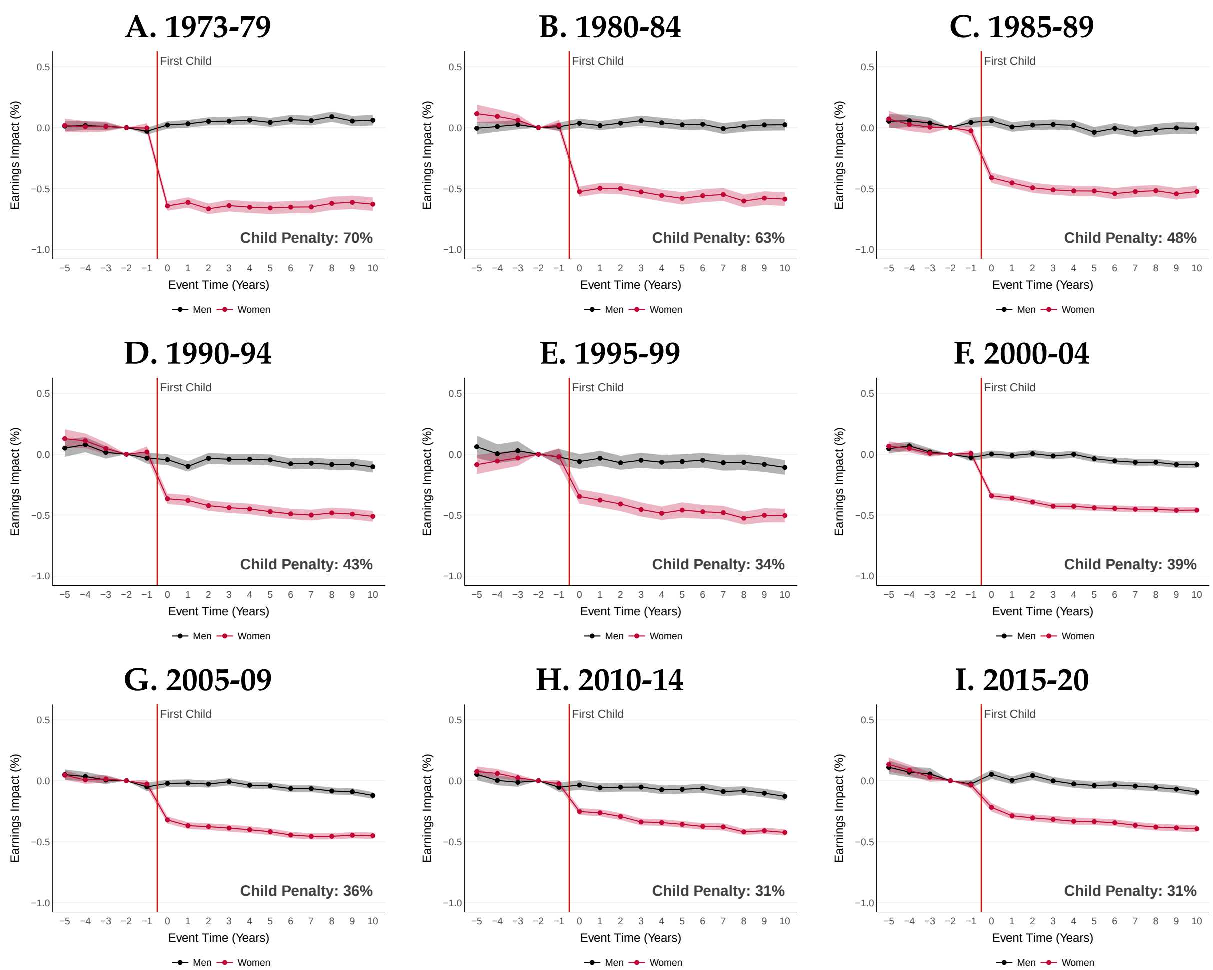
WEEKLY EMPLOYMENT



Notes: This figure shows event studies of first childbirth for weekly employment in different time periods. The sample of parents is split by interview year and the event study specification (6) is run separately for each time period. The event studies start in 1973, because the first five years of the data (1968-1972) are reserved for obtaining synthetic pre-birth observations for those who had their first child in 1973. Each panel displays the average child penalty over event times 0-10 (defined in equation 8) for the time period in question. The 95% confidence intervals are based on robust standard errors.

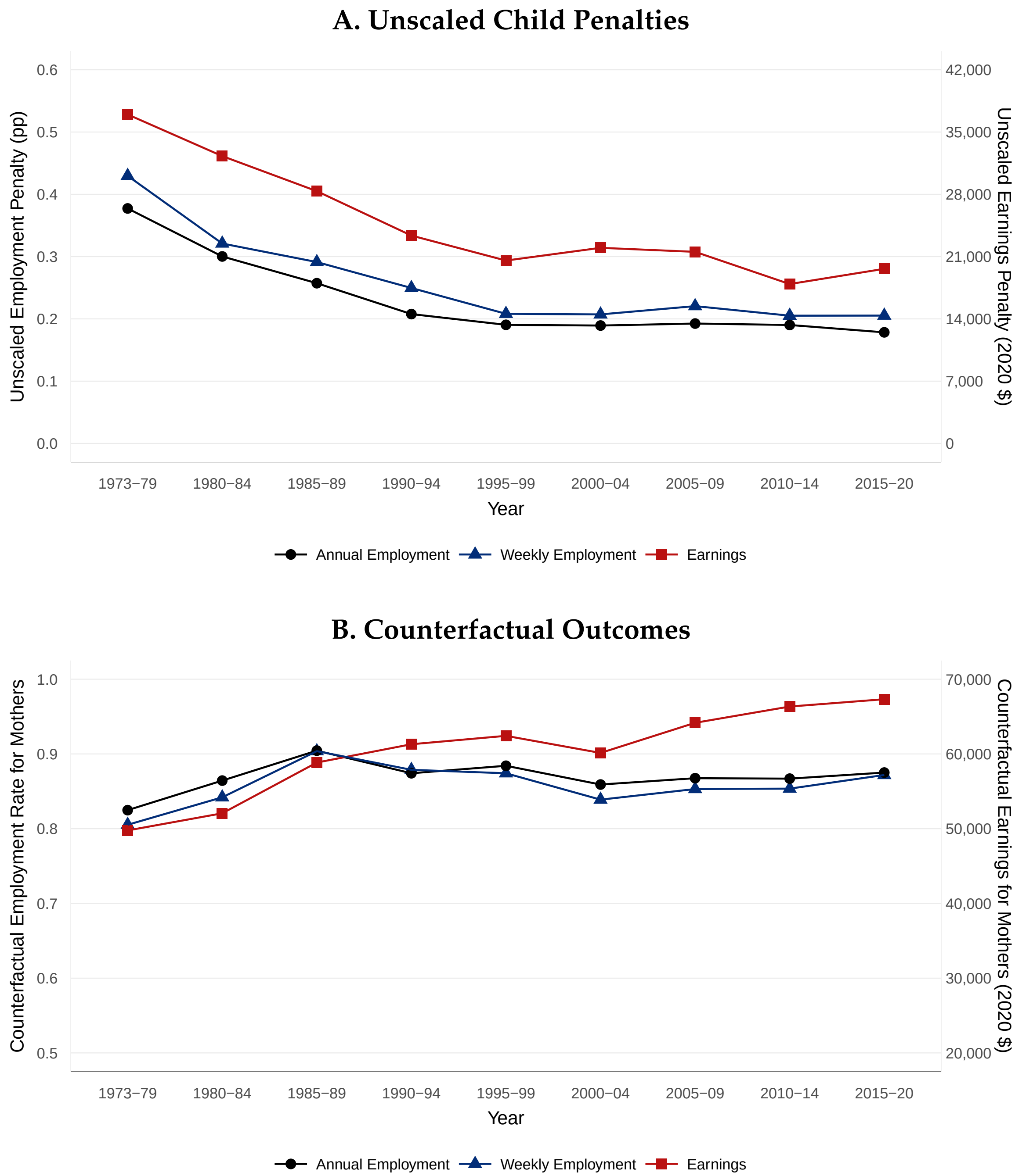
Figure A.8: Event Studies of First Childbirth Over Time

Earnings



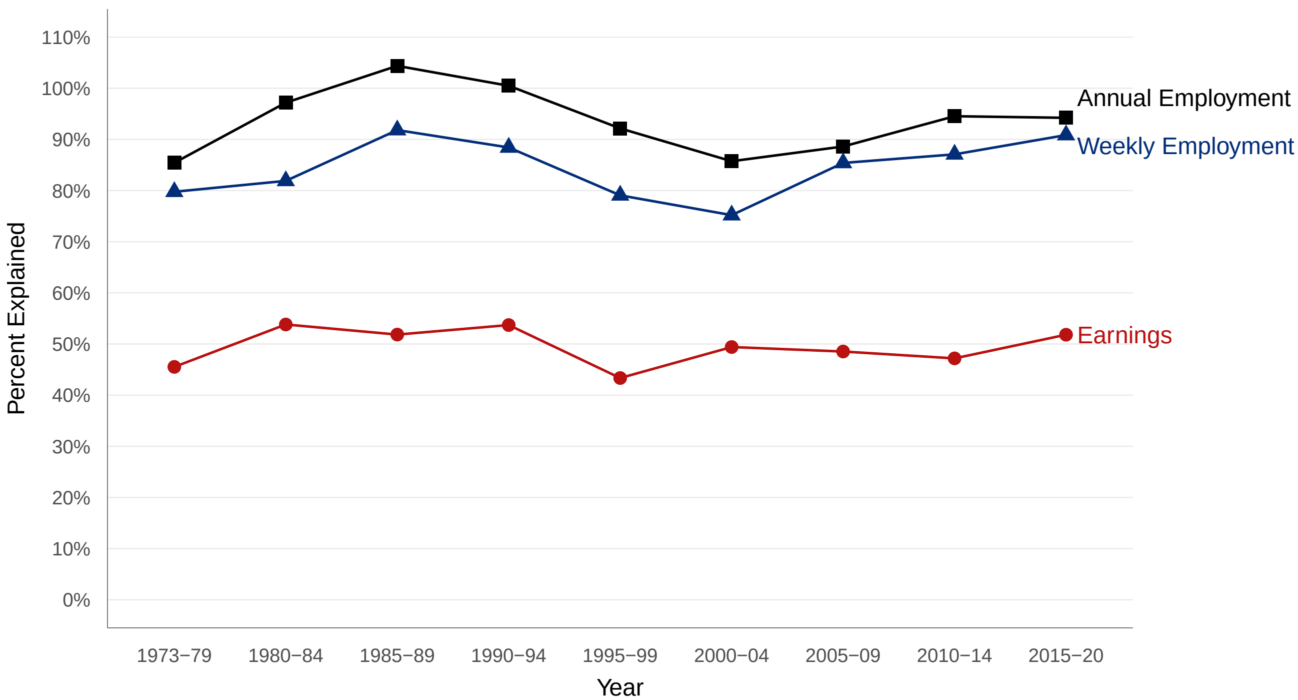
Notes: This figure shows event studies of first childbirth for earnings in different time periods. The sample of parents is split by interview year and the event study specification (6) is run separately for each time period. The event studies start in 1973, because the first five years of the data (1968-1972) are reserved for obtaining synthetic pre-birth observations for those who had their first child in 1973. Each panel displays the average child penalty over event times 0-10 (defined in equation 8) for the time period in question. The 95% confidence intervals are based on robust standard errors.

FIGURE A.9: UNSCALED CHILD PENALTIES AND COUNTERFACTUALS OVER TIME



Notes: This figure investigates if the time series of scaled child penalties (effects in percentage terms) shown in Figure 4 are driven primarily by changes in unscaled child penalties (effects in absolute terms) or by changes in the scaling factor (the level of the counterfactual outcome). Panel A shows the evolution of unscaled child penalties (employment effects in percentage points and earnings effects in dollars), while Panel B hows the evolution of the counterfactual levels used for scaling. The earnings estimates have been adjusted to 2020 dollars using nominal earnings growth in the full sample of working-age men and women in CPS data. The figure shows that the evolution of unscaled penalties is similar to the evolution of scaled penalties: a decline until the mid-1990s and then stagnation. The counterfactual outcomes have remained relatively constant for employment, while they have increased gradually for earnings. Hence, changes in the baseline hardly matter for the evolution of scaled employment penalties, while they play some role for the evolution of scaled earnings penalties.

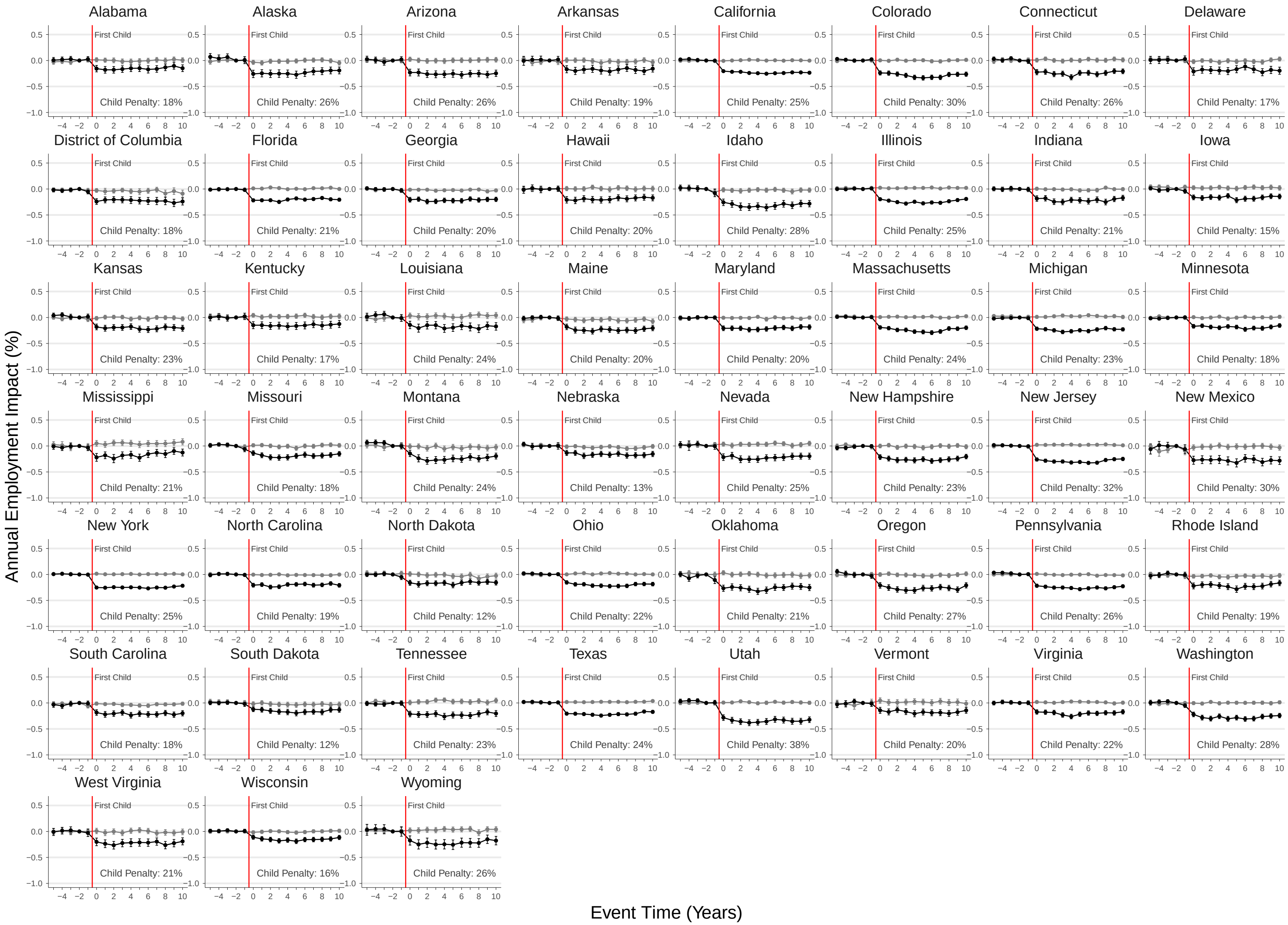
FIGURE A.10: FRACTION OF RAW GENDER GAP EXPLAINED BY CHILD PENALTIES



Notes: This figure shows the fraction of the raw gender gap for parents explained by child penalties over time. Results are shown for each of the three labor market outcomes: annual employment, weekly employment, and earnings. The raw gender gap is defined as the percentage difference between men and women with children, and the child penalty estimates are shown in Figure 4.

Figure A.11: Event Studies of First Childbirth Across States

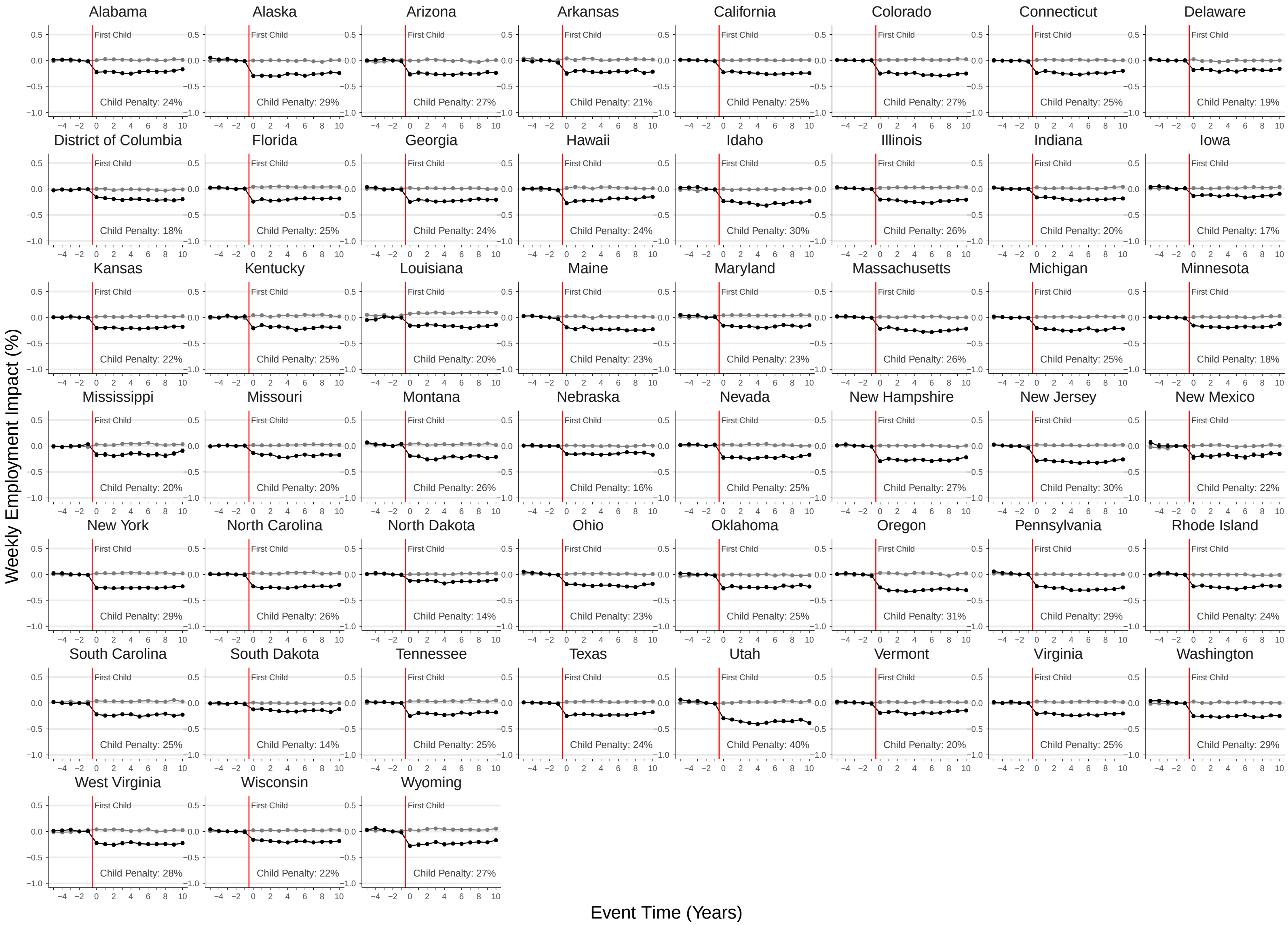
Annual Employment



Notes: This figure shows event studies of first childbirth in annual employment for each of the 51 US states (including the federal district of D.C.). State-level event studies are constructed by interacting the event time dummies in equation (6) with state dummies, estimating percentage impacts of childbirth on men and women at each event time ($\hat{P}^m_\tau$ and $\hat{P}^w_\tau$) as well as average child penalties over event times 0-10 separately for each state. Men are shown in gray and women are shown in black. In this specification, the lifecycle and time trends in equation (6) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

FIGURE A.12: EVENT STUDIES OF FIRST CHILDBIRTH ACROSS STATES

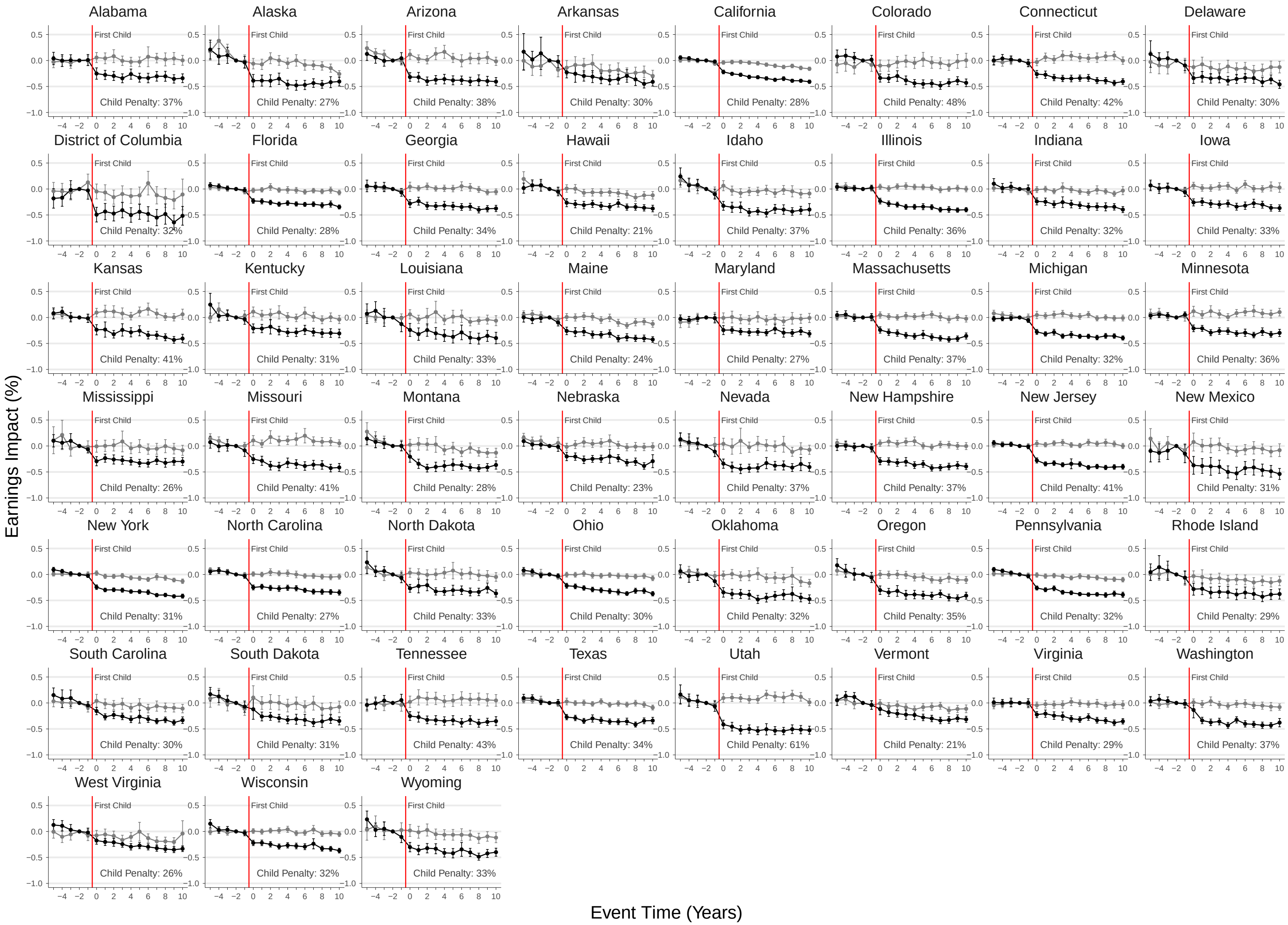
WEEKLY EMPLOYMENT



Notes: This figure shows event studies of first childbirth in weekly employment for each of the 51 US states (including the federal district of D.C.). State-level event studies are constructed by interacting the event time dummies in equation (6) with state dummies, estimating percentage impacts of childbirth on men and women at each event time ($\hat{P}^m_\tau$ and $\hat{P}^w_\tau$) as well as average child penalties over event times 0-10 separately for each state. Men are shown in gray and women are shown in black. In this specification, the lifecycle and time trends in equation (6) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

Figure A.13: Event Studies of First Childbirth Across States

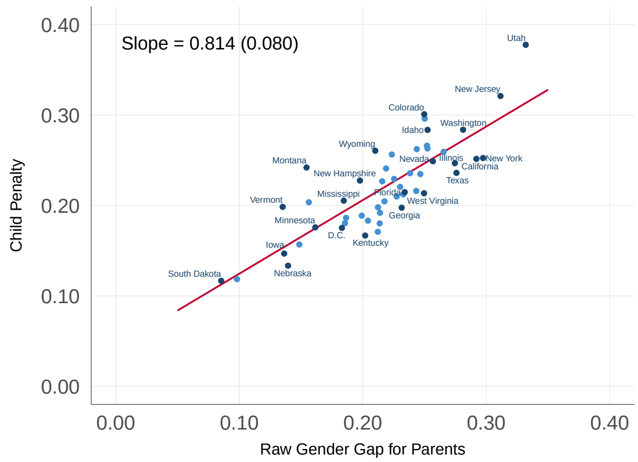
Earnings



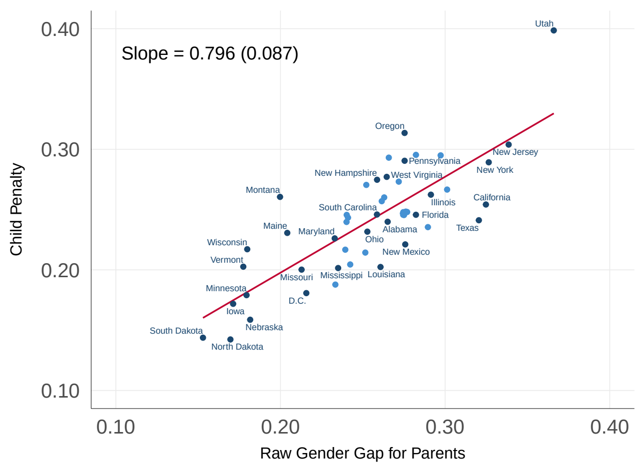
Notes: This figure shows event studies of first childbirth in earnings for each of the 51 US states (including the federal district of D.C.). State-level event studies are constructed by interacting the event time dummies in equation (6) with state dummies, estimating percentage impacts of childbirth on men and women at each event time ($\hat{P}^m_\tau$ and $\hat{P}^w_\tau$) as well as average child penalties over event times 0-10 separately for each state. Men are shown in gray and women are shown in black. In this specification, the lifecycle and time trends in equation (6) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

Figure A.14: Child Penalties vs Raw Gender Gaps Across States

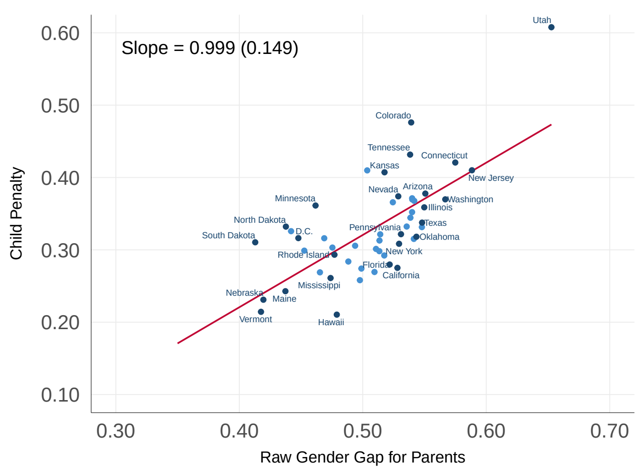
A. Annual Employment



B. Weekly Employment



C. Earnings

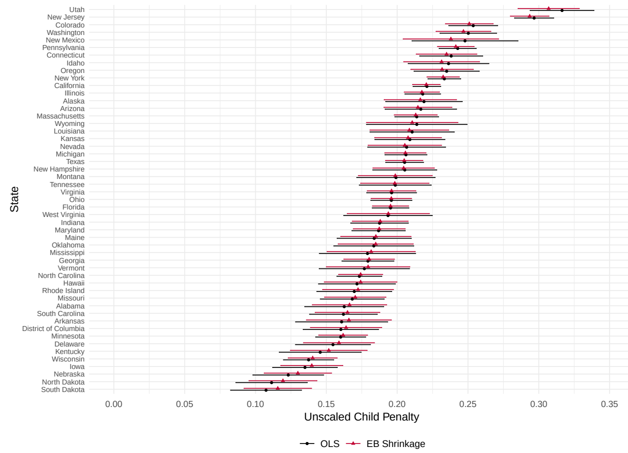


Notes: This figure provides scatter plots of child penalties against raw gender gaps for parents across states. Results are shown for each of the three labor market outcomes: annual employment, weekly employment, and earnings. The raw gender gap is defined as the percentage difference between men and women with children, and the child penalty estimates for each outcome and state are shown in Figures A.11-A.13.

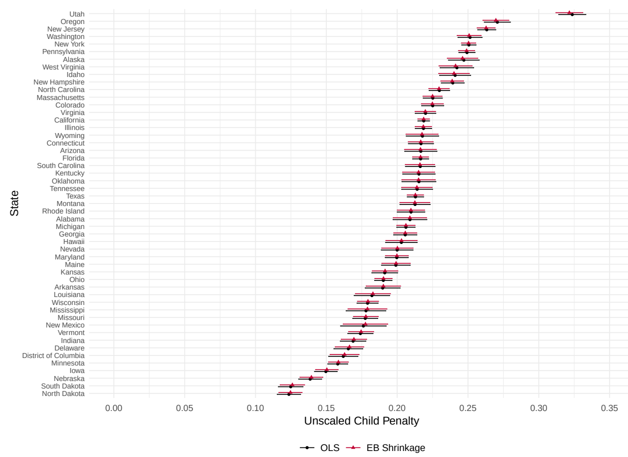
Figure A.15: EB vs OLS Estimates of Child Penalties

Point Estimates and Confidence Intervals by State

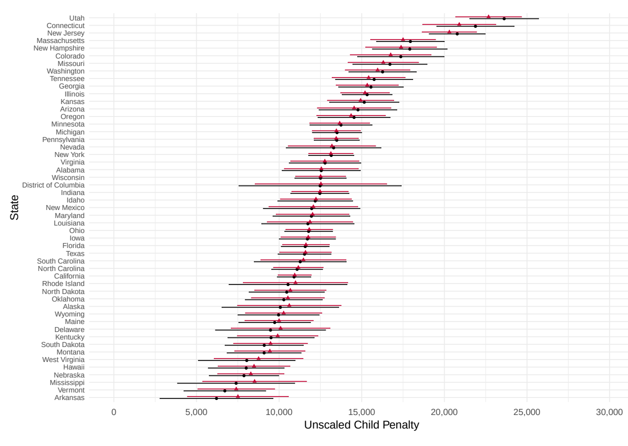
A. Annual Employment



B. Weekly Employment



C. Earnings

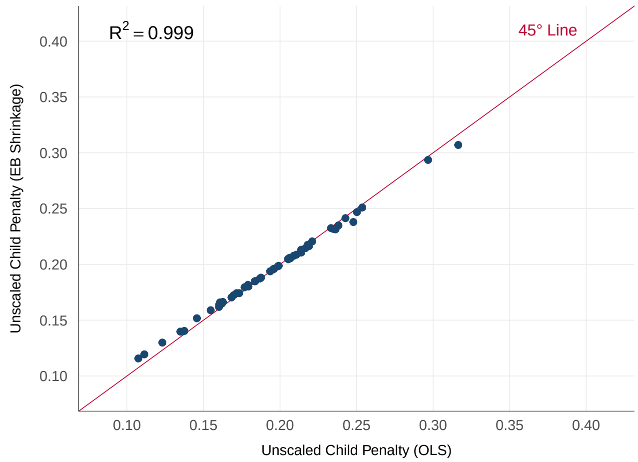


Notes: This figure compares OLS and Empirical Bayes (EB) estimates of unscaled child penalties for each state and each labor market outcome. The EB estimates are based on the linear-shrinkage formula in equation (13). The EB shrinkage adjustment hardly changes the estimates. The reason is the high statistical precision of the pseudo-event study approach: the imprecision of the state-level OLS estimates is very small compared to the variation in OLS estimates across states.

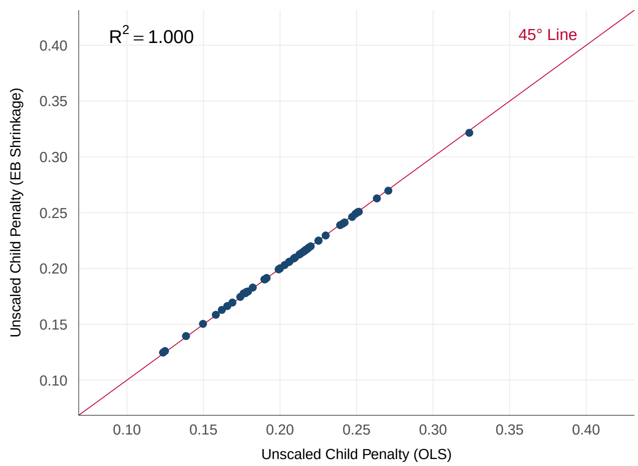
FIGURE A.16: EB VS OLS ESTIMATES OF CHILD PENALTIES

SCATTERPLOTS

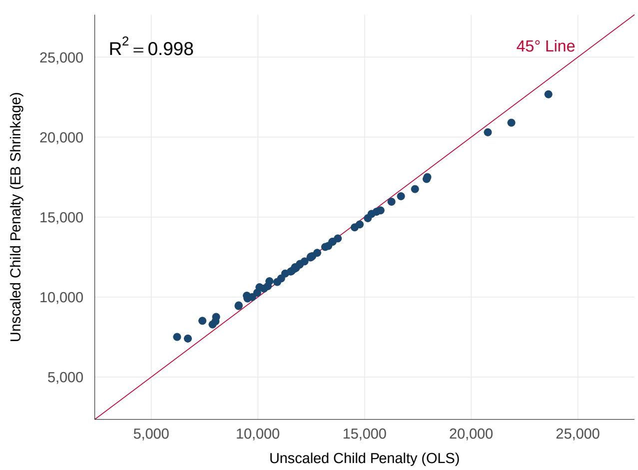
A. Annual Employment



B. Weekly Employment

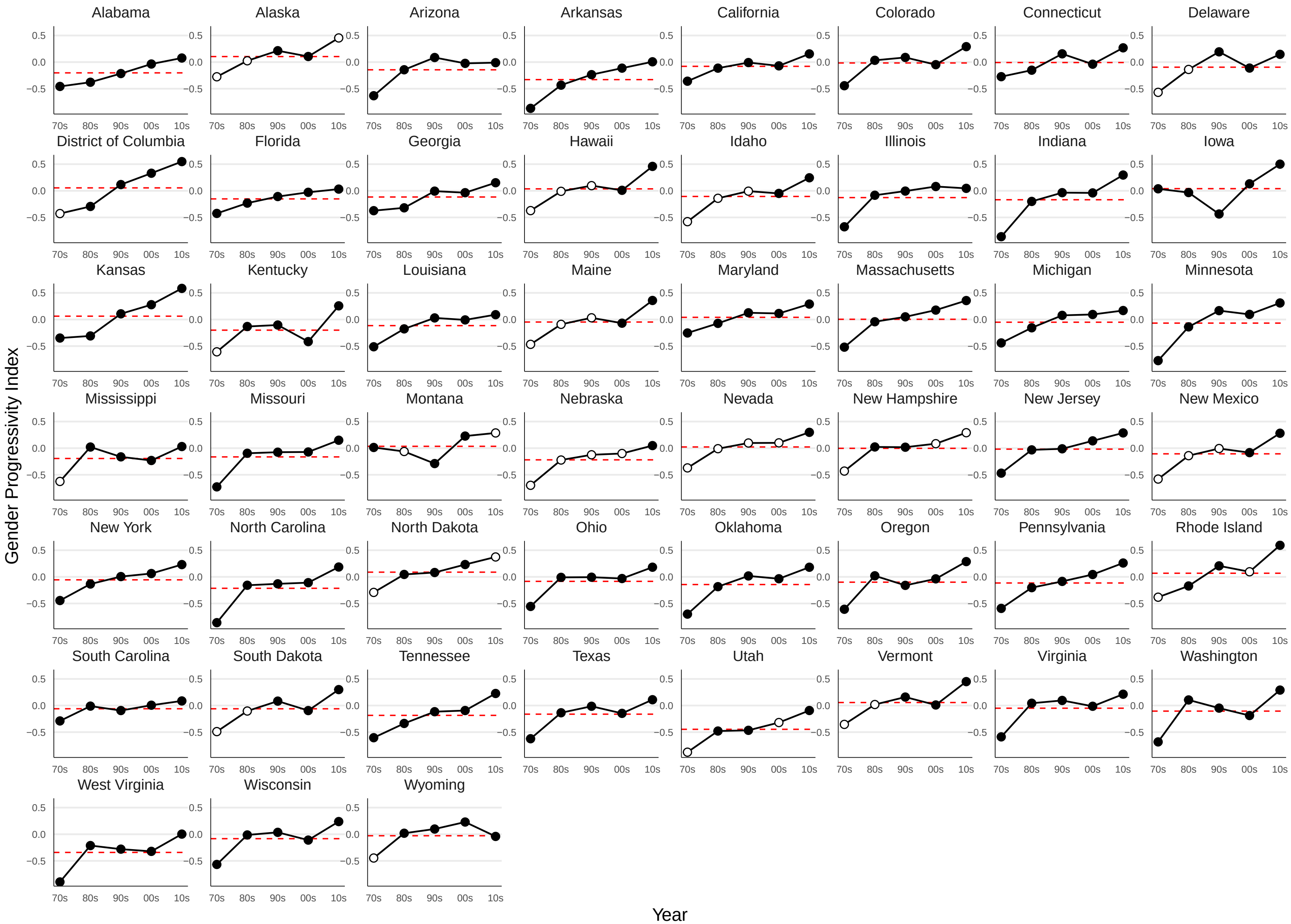


C. Earnings



Notes: This figure plots Empirical Bayes (EB) estimates of child penalties against OLS estimates of child penalties across states for each labor market outcome. The EB estimates are based on the linear-shrinkage formula in equation (13). The EB-OLS pairs align almost perfectly with the 45-degree line for all three outcomes. The reason is the high statistical precision of the pseudo-event study approach: the imprecision of the state-level OLS estimates is very small compared to the variation in OLS estimates across states.

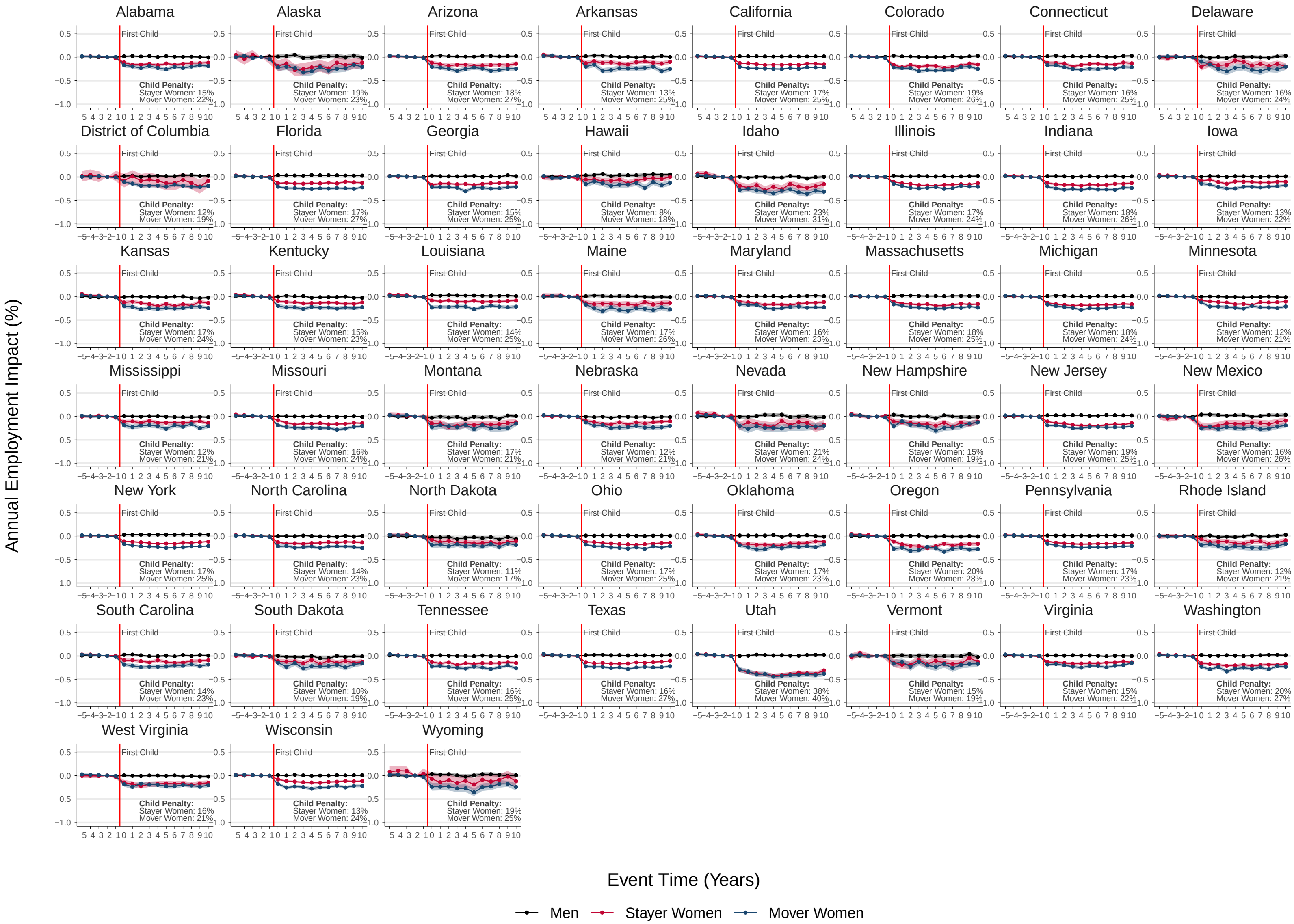
FIGURE A.17: GENDER PROGRESSIVITY INDEX BY STATE AND TIME



Notes: This figure presents time series of the Gender Progressivity Index (GPI) in each state over the last five decades. Using GSS data from 1972-2018, the index is calculated as the average standardized response to questions that elicit attitudes towards gender roles in families with children. The standardization ensures that the index has mean zero and standard deviation one. Three gender norms questions available in all five decades of GSS data are included in the construction of the index. Because these questions were not asked in every state in every decade, some state-decade observations are missing. Missing state-decade observations have been imputed based on the percentile of the state's GPI in the decades where it is observed. Actual state-decade observations are indicated by filled dots and imputed observations are indicated by empty dots.

Figure A.18: Event Studies of First Childbirth for Movers vs Stayers by State of Birth

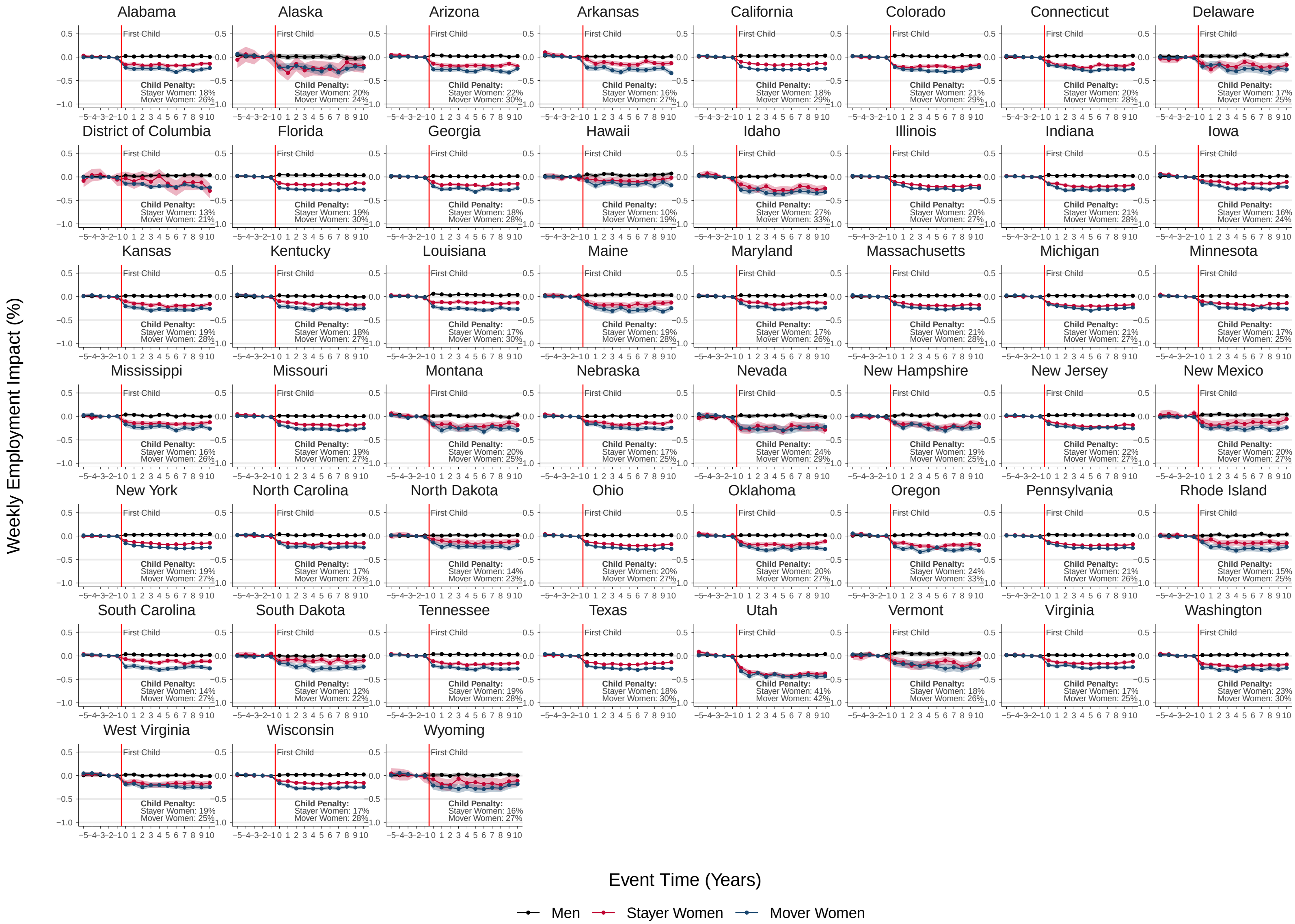
Annual Employment



Notes: This figure presents event studies of first childbirth for movers and stayers born in different states. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. To construct the figure, specification (6) is run separately for women movers and women stayers, interacting the event time dummies by state-of-birth dummies. The sample of men is not split by mover/stayer status as childbirth is a non-event for them regardless of status. The outcome is annual employment. Each panel displays child penalties over event times 0-10 for mover women and stayer women with a given state of birth. The 95% confidence intervals are based on robust standard errors. The sample is based on ACS data from 2000-2019, which contains information on both state of residence and state of birth.

Figure A.19: Event Studies of First Childbirth for Movers vs Stayers by State of Birth

Weekly Employment



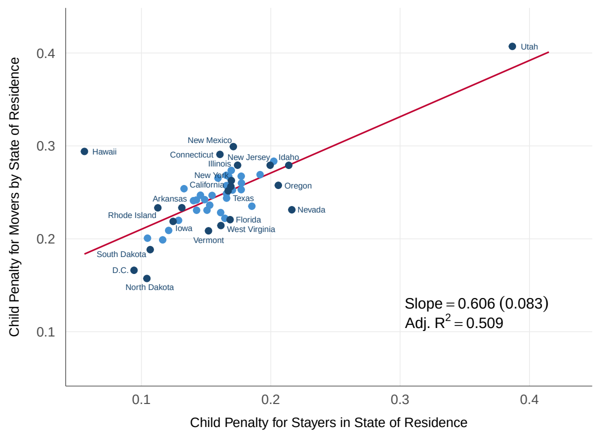
Notes: This figure presents event studies of first childbirth for movers and stayers born in different states. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. To construct the figure, specification (6) is run separately for women movers and women stayers, interacting the event time dummies by state-of-birth dummies. The sample of men is not split by mover/stayer status as childbirth is a non-event for them regardless of status. The outcome is weekly employment. Each panel displays child penalties over event times 0-10 for mover women and stayer women with a given state of birth. The 95% confidence intervals are based on robust standard errors. The sample is based on ACS data from 2000-2019, which contains information on both state of residence and state of birth.

FIGURE A.20: EPIDEMIOLOGICAL STUDY OF US MOVERS

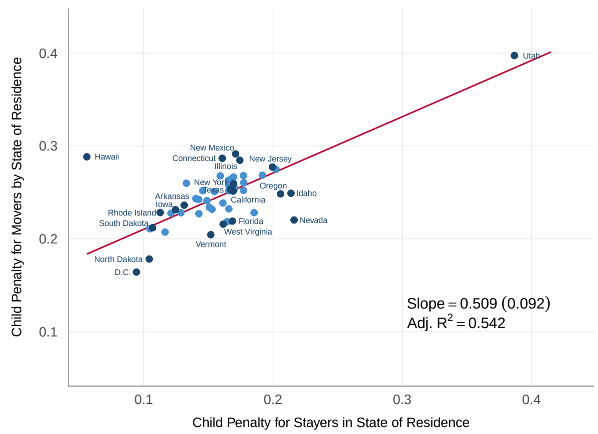
CHILD PENALTIES FOR MOVERS VS STAYERS BY STATE OF RESIDENCE

Annual Employment

A. Raw Data

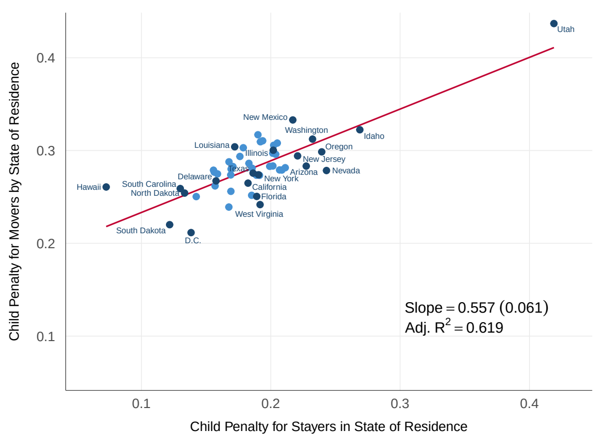


B. Residualize by Effect of State-of-Birth Penalties

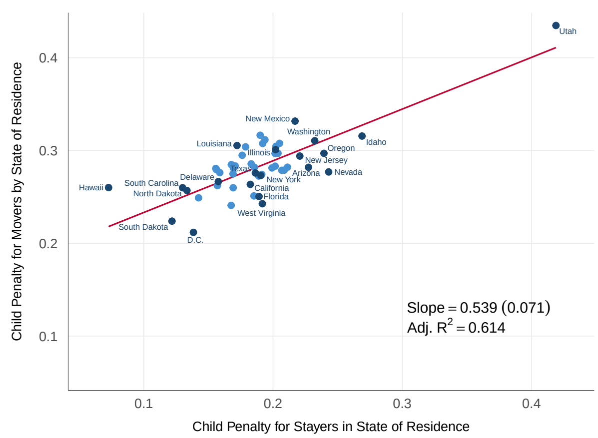


Weekly Employment

C. Raw Data



D. Residualize by Effect of State-of-Birth Penalties



Notes: This figure is symmetric to Figure 13, but focuses on the effect of residence state rather than the effect of birth state. Specifically, the figure provides scatter plots of the child penalty for movers against the child penalty for stayers by state of residence. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. The left panels show raw child penalties, while the right panels show residualized child penalties using the specification in eq. (15). The residualized plots control for selection on state of birth, which would otherwise contaminate the estimated effects of state of residence (local labor markets) with effects of state of birth (norms/culture). The figure shows that place of residence has sizable effects, although not quite as strong as the effects of place of birth. The sample is based on ACS data from 2000-2019, which contains information on both state of residence and state of birth.

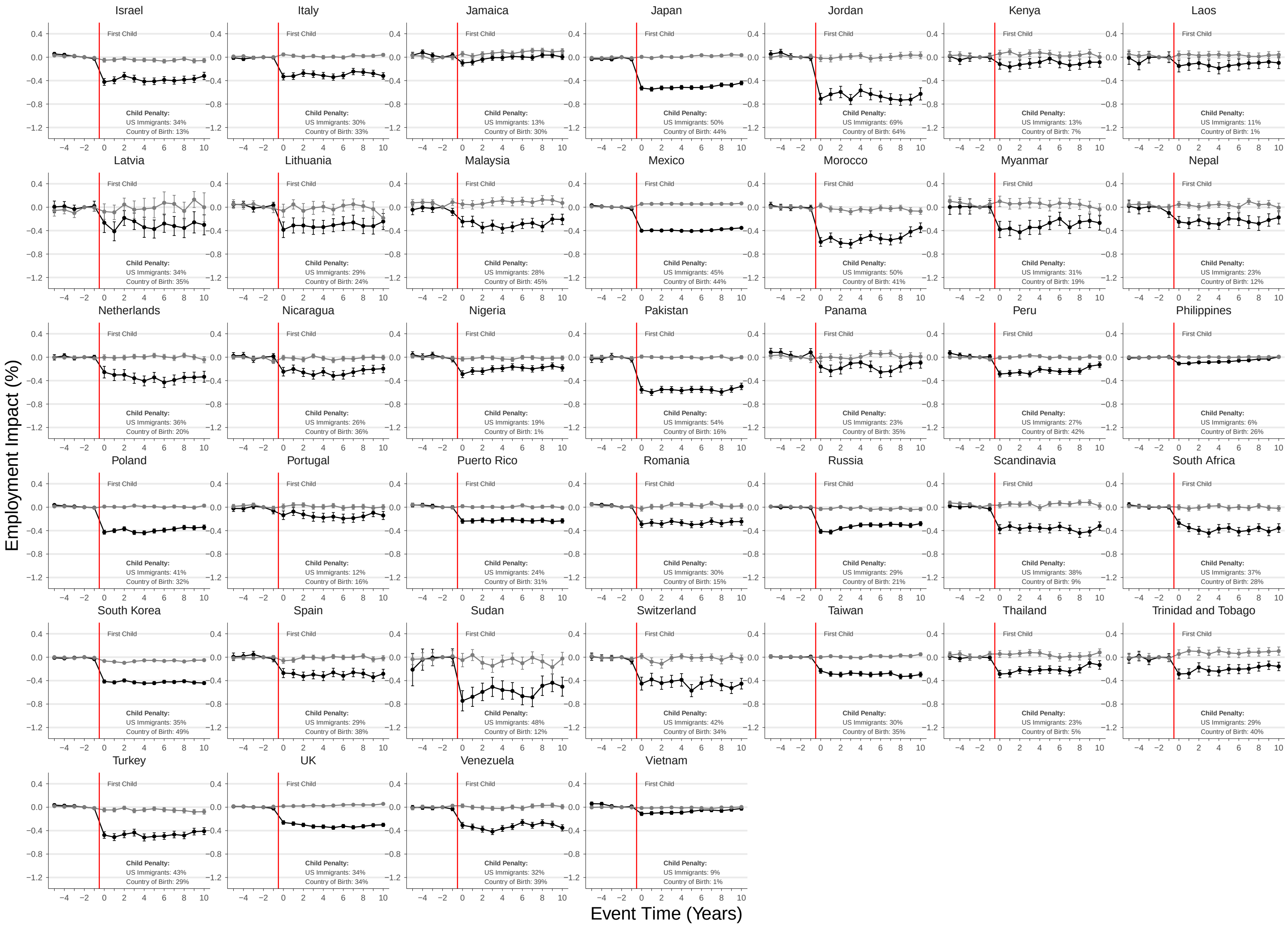
Figure A.21: Epidemiological Study of Foreign Immigrants

Event Studies of First Childbirth for Immigrants by Country of Birth

Notes: This figure presents event studies of first childbirth for foreign-born immigrants by country of birth. Each panel displays the child penalty for US immigrants (based on the series shown) and the child penalty in their country of birth (based on Kleven, Landais and Leite-Mariante 2024). The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020. The 95% confidence intervals are based on robust standard errors.

FIGURE A.21: EPIDEMIOLOGICAL STUDY OF FOREIGN IMMIGRANTS (CONTINUED)

EVENT STUDIES OF FIRST CHILDBIRTH FOR IMMIGRANTS BY COUNTRY OF BIRTH

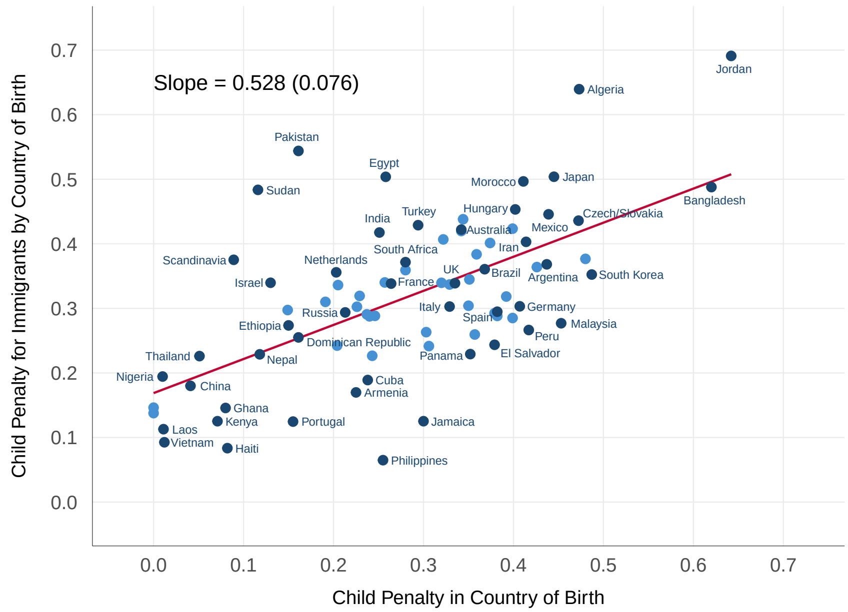


Notes: This figure presents event studies of first childbirth for foreign-born immigrants by country of birth. Each panel displays the child penalty for US immigrants (based on the series shown) and the child penalty in their country of birth (based on Kleven, Landais and Leite-Mariante 2024). The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020. The 95% confidence intervals are based on robust standard errors.

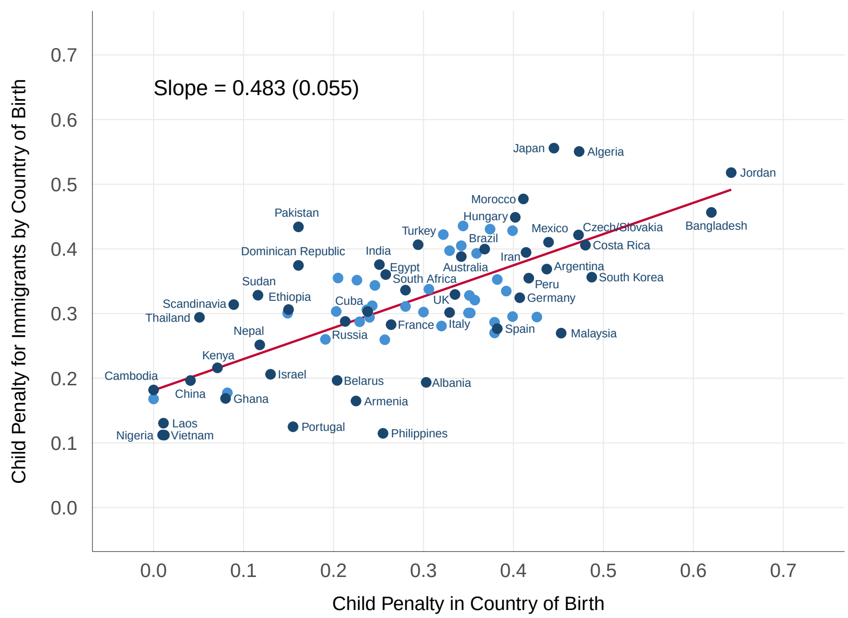
Figure A.22: Epidemiological Study of Foreign Immigrants

Child Penalties for Immigrants vs Child Penalties in Countries of Birth

A. Raw Data



B. With Controls

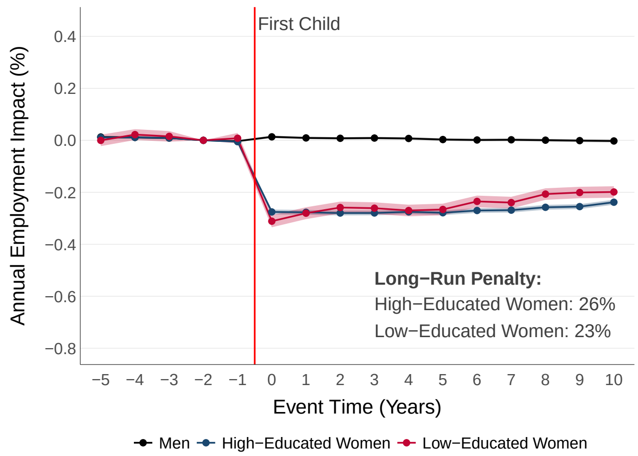


Notes: This figure presents scatter plots of child penalties for foreign-born immigrants against child penalties in country of birth. The underlying event studies for US immigrants are shown in Appendix Figure A.21 and the child penalties in country of birth are taken from Kleven, Landais and Leite-Mariante (2024). Panel A shows raw child penalty estimates, while Panel B controls for differences in education, marriage, race, fertility, age at first birth, and US location across immigrants from different countries. The specification of these control variables corresponds to the variables shown in Table 3. To construct Panel B, immigrant penalties are regressed on birth-country penalties and demographic controls, residualizing the immigrant penalties by the estimated effect of the controls for each country. The average effect of controls across all countries is added back to the residualized outcome to make the levels in Panel A and B comparable. The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020.

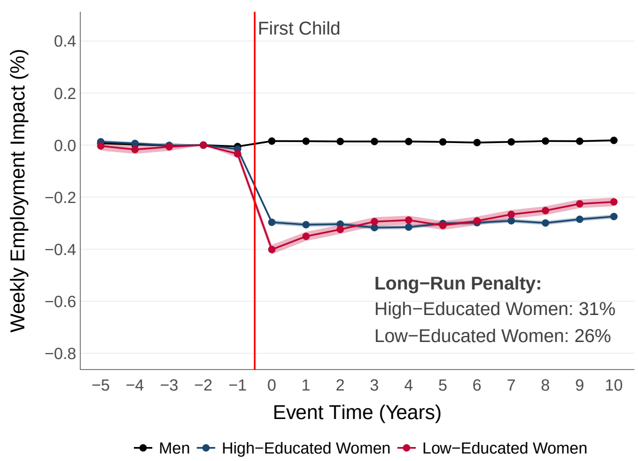
Figure A.23: Event Studies of First Childbirth by Education

Foreign Immigrants

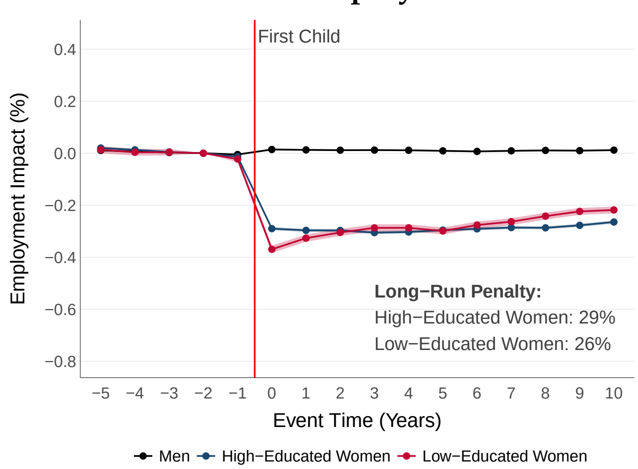
A. Annual Employment



B. Weekly Employment



C. Pooled Employment

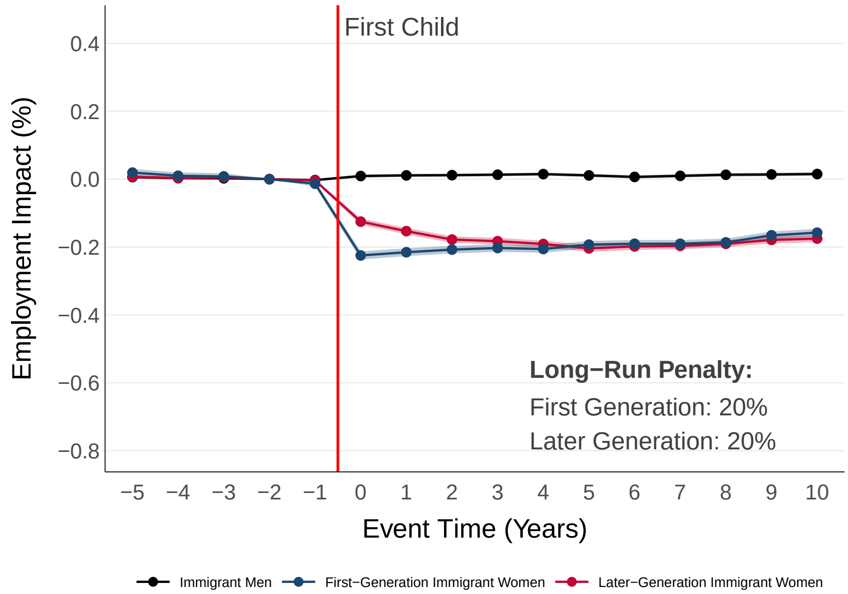


Notes: This figure presents event studies of first childbirth by female education level for foreign-born immigrants. The figure is constructed in the same way as the education part of Figure 8 for the full sample. Results are shown for three labor market outcomes: annual employment, weekly employment, and pooled employment. The analysis is based on ACS data from 2000-2019 and CPS data from 1994-2020.

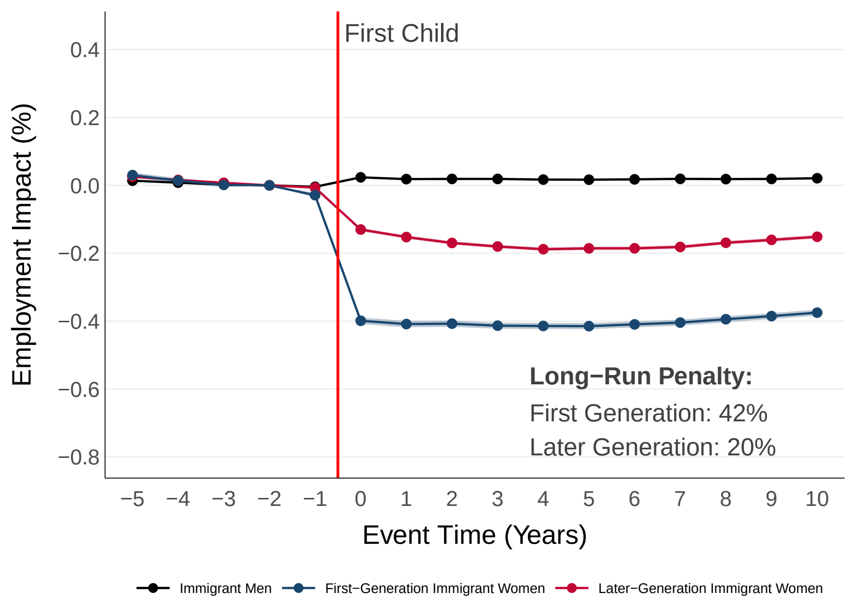
Figure A.24: Cultural Assimilation of Immigrants

First-Generation vs Later-Generation Child Penalties by Origin-Country Penalty

A. Bottom Quartile of Child Penalty in Country of Origin



B. Top Quartile of Child Penalty in Country of Origin



Notes: This figure presents event studies of first childbirth for first-generation and later-generation immigrants by quartile of the child penalty in country of origin. First-generation immigrants are defined as foreign-born US residents, while later-generation immigrants are defined as US-born residents who report foreign ancestry. The analysis is based on the 81 countries shown in Appendix Figure A.21, dividing countries into quartiles of the child penalty using the estimates in Kleven, Landais and Leite-Mariante (2024). The figure is constructed by running the event study specification (6) for first- and later-generation immigrant women separately (within the bottom and top quartiles of origin-country penalties, respectively). Each panel displays long-run child penalties (over event times 5-10) for first- and later-generation immigrants. The outcome is pooled employment (combining information on weekly and annual employment) and the sample is based on ACS data from 2000-2019 and CPS data from 1994-2020. The 95% confidence intervals are based on robust standard errors.