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 October 2024

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 Eva Demsky, 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 October 2024
JEL No. J13, J16, J21, J22, J61

ABSTRACT

This paper develops a new approach to estimating child penalties in labor market outcomes based on cross-sectional data and pseudo-event studies around child birth. The approach is applied to US data and validated against the state-of-the-art panel data approach. Child penalties can be accurately estimated using cross-sectional data, which are widely available and offer more statistical power than typical panel datasets. The approach allows for providing granular evidence on child penalties over time, across geography, and across demographic and cultural groups. Child penalties vary enormously across space: the employment penalty ranges from 12% in the Dakotas to 38% in Utah, while the earnings penalty ranges from 21% in Vermont to 61% in Utah. To investigate if this variation is driven by differences in gender norms, an epidemiological study of movers within the US and immigrants from abroad is presented. The child penalty for US movers is strongly related to the child penalty in their state of birth, adjusting 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), conditional on where they live. 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). Evidence on cultural assimilation is also presented.

Henrik Kleven
Department of Economics
Princeton University
238 Julis Romo Rabinowitz Building
Princeton, NJ 08544
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, Landais, Posch, Steinhauer, and Zweimüller 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 driving child penalties is still in its infancy. We have evidence ruling out explanations such as biology and comparative advantage (Kleven, Landais, and Søgaard 2021) and the incentives created by government policy (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2024), but virtually no evidence conclusively ruling in explanations. A key reason for the paucity of evidence is the data-demanding nature of how child penalties are estimated: event studies around child birth using high-quality panel data. Because of data constraints, child penalty estimates are available for less than a dozen countries and there is hardly any evidence on the variation in child penalties across space and time within countries. To address this gap in knowledge, the present paper develops a new approach to estimating child penalties based on widely available cross-sectional data. The approach is applied to data from the United States.

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 the birth of the first child, 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 $\tau = -n$. Having created a pseudo-panel this way, the event study specification of Kleven, Landais, and Søgaard (2019) is implemented. The results from the pseudo-event study approach are validated against results from an actual event study approach using data from the Panel Study of Income Dynamics (PSID 1968-2019) and the National Longitudinal Survey of Youth (NLSY 1979-2018). The two approaches yield very similar results, but the cross-sectional approach is much more precise due to superior sample size.[1]

The average child penalty in the US is currently 20% in annual employment, 24% in weekly employment, and 31% in earnings. These child penalties are larger than in Scandinavia, but smaller than in central Europe (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019). As in those other countries, US child penalties account for most of the observed gender inequality in labor market outcomes. Similar estimates exist in the literature (Kleven, Landais, Posch, Steinhauer, and Zweimüller 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]

Four main empirical findings are presented. First, child penalties have fallen substantially over the last five decades. The penalties 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, Pan, Shen, and Washington 2018). The literature has discussed a number of explanations, but a conclusive story has not yet emerged. The evidence presented here points to a simple explanation: gender convergence stalled because the decline in child penalties stalled.

Second, child penalties vary enormously across geography. The child penalty in annual employment ranges from 12% in the Dakotas (rural states with Scandinavian heritage) to 38% in Utah (a religiously and culturally conservative state). The child penalty in earnings ranges from 21% in Vermont, another rural state, to 61% in Utah. Interestingly, the range of child penalties across US states aligns closely with the range of child penalties between Scandinavian countries and the culturally conservative countries of central Europe (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019). Looking at the US map of child penalties highlights two potential mechanisms: urbanization and cultural norms. More urban places tend to have larger penalties, perhaps because urban jobs offer less flexibility than rural jobs. Working on a farm in North Dakota

[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.

is a different proposition from working in a bank in Manhattan, irrespective of preferences and norms, and job flexibility matters for gender gaps (Goldin 2014; Goldin and Katz 2016). More culturally conservative places tend to have larger penalties, but many conservative places are also rural and this pulls 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.[3]

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 an almost perfect 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.[4] 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

[3]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).

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

(such as China, Cuba, or Portugal).[5] 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 in her employment penalty of about 5pp. These results are consistent with important effects of childhood culture on child penalties. I show that the effects are unlikely to be driven by differential selection of movers and migrants from different places.

This paper contributes to a large literature on gender inequality, reviewed by Altonji and Blank (1999), Bertrand (2011), and Blau and Kahn (2017). It relates most directly to a burgeoning literature studying the impact of child birth on gender gaps in the labor market, including Angelov, Johansson, and Lindahl (2016), Kleven, Landais, and Søgaard (2019), Kleven, Landais, Posch, Steinhauer, and Zweimüller (2019), Kleven, Landais, and Søgaard (2021), Kleven, Landais, Posch, Steinhauer, and Zweimüller (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, Joensen, Johansen, and Veramendi 2023), and IVF treatment success (Lundborg, Plug, and Rasmussen 2017).[6] While this research agenda has produced many important 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 event times around child birth, thus allowing for the implementation of event study designs. The event studies are compelling in terms of the standard criteria: the pre-event trends are perfectly parallel and the post-event effects are immediate, persistent, and precisely estimated.[7] Given the availability of large cross-

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

[6]The IUD instrument is particularly compelling and it replicates child penalty estimates from event studies almost exactly (Gallen, Joensen, Johansen, and Veramendi 2023). The IVF instrument raises several concerns discussed in recent papers. These include concerns about exogeneity (Groes, Houštecká, Iorio, and Santaeulàlia-Llopis 2024 find observable imbalances by IVF treatment status), concerns about the exclusion restriction (Bögl, Moshfegh, Persson, and Polyakova 2024 show that failed IVF impacts other things than fertility), and concerns about external validity (highly planned births among women who are older, better educated, and have higher incomes than typical first-time mothers). These issues may be responsible for the smaller long-run child penalties found in some recent IVF papers.

[7]I also 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; Borusyak,

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 is 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., Fernández, Fogli, and Olivetti 2004; Fortin 2005; Fernández and Fogli 2009; Blau, Kahn, and Papps 2011; Bertrand 2020).[8] 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 child birth 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 child birth 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 and validates the empirical methodology. Section 4 presents evidence on US child penalties across time, geography, and demographic groups. Section 5 investigates the effect of gender norms on child penalties using difference-in-differences and epidemiological approaches. Section 6 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

Jaravel, and Spiess 2024).

[8]In the context of child penalties, recent studies provide correlational evidence consistent with social norms effects, including cross-country evidence (Kleven, Landais, Posch, Steinhauer, and Zweimüller 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.

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".[9] 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). While annual employment captures extensive margin labor supply, weekly employment captures both extensive and intensive margin labor supply: working some weeks or not at all over the year, and the number of weeks worked over the year. Annual employment 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 specification is validated against a true event study specification 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 NLSY component is taken from the 1979 cohort of the data. 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 under-powered for more granular analyses.

3 Pseudo-Event Study Approach

3.1 Potential Outcomes Framework

I start by developing a potential outcomes framework to elucidate the causal identification of child penalties using event study and pseudo-event study approaches. The potential outcome for

[9] 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.

individual i at time t is denoted by $Y_{it}(D_{it})$ where $D_{it} = 0, 1$ is an indicator for the presence of a child (parenthood). The causal effect of having a child on individual i at time t equals $\Delta_{it} = Y_{it}(1) - Y_{it}(0)$, and the average treatment effect (ATE) is given by

$$\mathbb{E}[\Delta_{it}] = \mathbb{E}[Y_{it}(1) - Y_{it}(0)]. \tag{1}$$

We do not observe Δ_{it} and therefore cannot directly measure the treatment effect. At any point in time, an individual is observed either with or without children, not both. With access to panel data, we do observe the same individual with and without children, but at different times. The idea of the event study approach is to compare outcomes for the same individual at nearby times before and after child birth.

Event Study Approach: To see how the approach works, we need to disinguish between three dimensions of time: (*i*) calendar time t, (*ii*) age a, and (*iii*) event time τ. In this context, event time is measured relative to the birth of the first child, i.e. $\tau = t - T_i = a - A_i$ where T_i is the year of first child birth and A_i is the age of first child birth. Event time corresponds to the age of the oldest child. Crucially, there is independent variation in all three time dimensions: conditional on calendar time t and age a, individuals are observed at different event times τ due to variation in the timing of first birth.

The event study approach is based on sharp changes around the time of first birth at $\tau = 0$. Specifically, with panel data, we can measure the average within-individual change at $\tau = 0$:

$$\hat{\Delta}_0 = \mathbb{E}[Y_{it}(1)|\tau = 0, t, a] - \mathbb{E}[Y_{it}(0)|\tau = -1, t-1, a-1]. \tag{2}$$

Here we compare the same individuals observed one year apart. Since this estimate may be confounded by calendar time and age effects, we control for t and a. This is feasible because, as noted above, there is independent variation in t, a, and τ due to heterogeneity in the timing of first birth. Controlling for t, a amounts to considering a second difference:

$$\hat{\Delta}_{-1} = \mathbb{E}[Y_{it}(0)|\tau = -1, t, a] - \mathbb{E}[Y_{it}(0)|\tau = -1, t-1, a-1]. \tag{3}$$

The estimate of the causal impact is then given by

$$\hat{\Delta}_0 - \hat{\Delta}_{-1} = \mathbb{E}\left[Y_{it}\left(1\right)|\tau = 0, t, a\right] - \mathbb{E}\left[Y_{it}\left(0\right)|\tau = -1, t, a\right]. \tag{4}$$

This is an estimate of the average treatment effect on the treated (ATT). The key difference from $\mathbb{E}\left[\Delta_{it}\right]$ is that we are comparing populations at (marginally) different event times τ. Given we are conditioning on calendar time t and age a, this amounts to comparing populations at (marginally) different ages at first birth. The main identification assumption is that the potential outcome is smooth in age at first birth.

The estimator in equation (4) captures the short-run impact of children, at $\tau = 0$. If there are no dynamics in the effect of children — i.e., if they have the same effect regardless of their age — then this estimate captures all we need to know. In the presence of dynamics, the long-run impact will be different from the short-run impact in equation (4). We could estimate longer-run impacts simply by extending the event time window:

$$\hat{\Delta}_{\tau \gg 0} - \hat{\Delta}_{-1} = \mathbb{E}\left[Y_{it}\left(1\right)|\tau \gg 0, t, a\right] - \mathbb{E}\left[Y_{it}\left(0\right)|\tau = -1, t, a\right]. \tag{5}$$

The problem with this approach is that we are comparing individuals who are far apart in terms of age at first birth. Hence, smoothness of the potential outcome in age at first birth is not sufficient for identification. We therefore take advantage of the fact that we estimate the difference in equation (5) separately for men and women. Letting $\hat{\Delta}^g_\tau - \hat{\Delta}^g_{-1}$ denote the difference between event time $\tau \geq 0$ and event time $\tau = -1$ for gender $g = m, w$, the estimate of the long-run impact is given by

$$\left[\hat{\Delta}^w_{\tau \gg 0} - \hat{\Delta}^w_{-1}\right] - \left[\hat{\Delta}^m_{\tau \gg 0} - \hat{\Delta}^m_{-1}\right]. \tag{6}$$

This difference-in-differences estimator compares men and women between event times $\tau \gg 0$ and $\tau = -1$, conditional on calendar year t and age a. As described, the variation in τ when conditioning on t, a comes from variation in age at first birth. The main identification assumption is parallel trends between men and women with respect to age at first birth.

Pseudo-Event Study Approach: Now consider a setting without panel data, but with repeated cross-sectional data. For individuals with children, we observe the age of their oldest child and therefore know their place in positive event time, $\tau \geq 0$. For individuals without children, we do not observe if and when they will have children and therefore do not know their place in negative event time, $\tau < 0$. The idea of the approach is to use matching to impute negative event times among those without children. The procedure translates the cross-sectional dataset into a pseudo-panel of individuals at different event times τ.

The matching procedure is implemented as follows. Consider parent i observed at event time $\tau = 0$ in calendar year t with age a and demographic characteristics $\boldsymbol{X_i}$. This parent is matched to a childless individual j observed in year $t - n$ with age $a - n$ and the same demographic characteristics $\boldsymbol{X_j} = \boldsymbol{X_i}$. This gives a synthetic observation for $\tau = -n$.[10] By matching each parent at $\tau = 0$ to childless individuals for $n = 1, ..., 5$, we obtain a pseudo-panel with 5 years of pre-child data. Critically, the pseudo-panel data — just like the panel data considered above — feature independent variation in (τ, t, a). Among those observed at $\tau = 0$, there will be variation in t and a due to heterogeneity in the timing of first birth. The variation in t, a conditional on event time among parents ($\tau \geq 0$) creates variation in t, a conditional on event time among matched non-parents ($\tau < 0$). This allows for estimating the impact of child birth using the framework laid out above. That is, the short- and long-run impacts are estimated as shown in equations (4)-(6).

The approach is related to the pseudo-panel (or synthetic-cohort) approach developed by Deaton (1985). As in the synthetic-cohort approach, the matching procedure used here ensures that the estimation sample consists of fixed cohorts over time. The procedure is richer than a synthetic-cohort approach by holding other dimensions fixed too. The key innovation is to use matching to impute event times with respect to child birth for individuals observed without children, thus allowing for event studies of child birth using cross-sectional data.

3.2 Regression Framework

As described above, we conduct event studies around the birth of the first child — indexed to occur at event time $\tau = 0$ — conditional on calendar time t and age a. The approach relies on the existence of independent variation in (τ, t, a) due to heterogeneity in age at first birth. As proposed by Kleven, Landais, and Søgaard (2019), the following regression specification is run

[10] A parent will have multiple possible matches whenever there is more than one childless individual in the specified cell of observables (year, age, and other 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.

separately for men and women:

$$Y_{it}^{g} = \boldsymbol{\alpha^{g}} \cdot \boldsymbol{D}_{i\tau}^{Event} + \boldsymbol{\beta^{g}} \cdot \boldsymbol{D}_{ia}^{Age} + \boldsymbol{\gamma^{g}} \cdot \boldsymbol{D}_{it}^{Year} + \nu_{it}^{g}, \tag{7}$$

where Y_{it}^{g} is the outcome for individual i of gender $g = w, m$ at calendar time t. On the right-hand side, boldface is used to denote vectors. The first term includes dummies for each event time τ, omitting a base year before child birth. The omitted base year is $\tau = -2$, the year before pregnancy, but the choice of base year hardly affects the results as there is virtually no pre-trend in the data. The event time coefficients $\alpha_{\tau}^{g} \in \boldsymbol{\alpha^{g}}$ measure the impact of child birth on gender g in event year τ. The second and third terms include a full set of age and year dummies to control non-parametrically for lifecycle trends and time trends.

The focus is on labor market outcomes such as earnings and employment. Equation (7) is specified in levels rather than in logs to keep observations with zero earnings and employment, thus capturing both intensive and extensive margin responses. The estimated level effects are converted into percentage effects by calculating

$$P_{\tau}^{g} = \frac{\hat{\alpha}_{\tau}^{g}}{\mathbb{E}\left[\tilde{Y}_{it}^{g} \mid \tau\right]}, \tag{8}$$

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.

$$\text{Child Penalty} \equiv \mathbb{E}\left[P_{\tau}^{m} - P_{\tau}^{w} \mid \tau \geq 0\right] - \mathbb{E}\left[P_{\tau}^{m} - P_{\tau}^{w} \mid \tau < 0\right]. \tag{9}$$

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 child birth, 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: The event study framework in equations (7)-(9) relies on a parallel trends assumption. In addition to parallel trends, staggered event studies make implicit assumptions 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; Borusyak, Jaravel, and Spiess 2024). A paper by Melentyeva and Riedel (2023) investigates the issue specifically in the context of child penalties.[11]

Given the staggered event study specification in (7) relies on variation in age at first birth, the potential threat to identification is heterogeneity in this dimension. Using an approach similar to Sun and Abraham (2021) and Melentyeva and Riedel (2023), I address this issue by considering a stacked event study that allows for heterogeneous treatment effects by age at first birth:

$$Y_{it}^{g} = \boldsymbol{\alpha_q^g} \cdot \boldsymbol{D}_{i\tau}^{Event} \cdot \mathbf{1}\left[Q_i = q\right] + \boldsymbol{\beta_q^g} \cdot \boldsymbol{D}_{ia}^{Age} \cdot \mathbf{1}\left[Q_i = q\right] + \boldsymbol{\gamma_q^g} \cdot \boldsymbol{D}_{it}^{Year} \cdot \mathbf{1}\left[Q_i = q\right] + \nu_{it}^{g}. \tag{10}$$

The indicator function $\mathbf{1}\left[Q_i = q\right]$ equals one if the individual's age at first birth belongs to group q. The event time coefficients $\alpha_{q\tau}^{g} \in \boldsymbol{\alpha_q^g}$ measure the impact of child birth in event year τ by gender g and age at first birth q. Defining scaled impacts as $P_{q\tau}^{g} = \hat{\alpha}_{q\tau}^{g} / \mathbb{E}\left[\tilde{Y}_{it}^{g} \mid q, \tau\right]$, the weighted average treatment effect across birth cohorts can be calculated as

$$P_{\tau}^{g} \equiv \sum_{q} \omega_q P_{q\tau}^{g}, \tag{11}$$

where ω_q denotes the sample share of cohort q. The weighted average in (11) does not suffer from the problem, highlighted in the recent econometrics literature, 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). Using this weighted average, the child penalty is defined as in equation (9). As we shall see, the child penalties obtained from the stacked specification are virtually identical to the child penalties obtained from the baseline specification that pools all ages at first birth. This suggests that treatment-effect heterogeneity does not create bias in this context.

[11] Using German data and a heterogeneity-robust approach, they show that child penalties in earnings are even larger than implied by the estimation framework of Kleven, Landais, and Søgaard (2019).

3.3 Descriptive Statistics

As a first step of the empirical analysis, 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, namely 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 event study evidence showing that parenthood has no impact on the labor market outcomes of men (Kleven, Landais, and Søgaard 2019; Kleven, Landais, Posch, Steinhauer, and Zweimüller 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 early cross-sectional literature addressed selection by controlling for observables, but this is not a credible solution due to the possibility of selection on unobservables.[12]

The standard event study approach addresses such selection problems by relying on within-individual variation around first child birth using panel data. Absent panel data, the solution proposed here is to create a pseudo-panel with variation around child birth 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 calendar time (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 child birth on men. Because child birth 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 are chosen to avoid such selection bias. These are gender, education (4 categories), marital status (5 categories), race (4 categories), and state of residence.[13]

[12] See Browning (1992) for a review of the early literature on children and family labor supply. While this literature focused mostly on female labor supply, it also discussed the "positive effect" of children on male labor supply. The arguments provided here suggest that this effect was an artefact of selection.

[13] 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

Table 2 provides descriptive statistics for matched men and women at event times $\tau = 0$ and $\tau = -1$ in the pseudo-panel.[14] By construction, these samples match exactly on education, marital 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 calendar 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. The next section validates the pseudo-panel specification against an actual panel specification.

3.4 Validation of Approach

Cross-Section vs Panel: Figure 1 compares results from the pseudo-event study approach (left panels) to results from an actual event study approach (right panels). The pseudo-event studies are based on CPS and ACS data over the period 1968-2020, while the actual event studies are based on PSID and NLSY data over the same period. Each panel shows an event study for men and women around the birth of their first child at $\tau = 0$, marked by the red vertical line. 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).[15]

The results from the cross-sectional and panel approaches align closely, but the cross-sectional approach is more compelling in terms of pre-trends and statistical precision.[16] It features almost perfectly parallel trends between men and women before child birth and sharp divergence immediately after. Having a child is a non-event for men, but leads to an immediate and persistent drop

categories combines information on race and ethnicity: white (non-Hispanic), black (non-Hispanic), Hispanic, and all others (mostly Asian).

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

[15] There is a slight difference in the matching algorithm for different labor market outcomes. The algorithm described in section 3.1 is used when considering weekly employment (obtained from a question about work activities *last week*), but needs to be adjusted when considering annual employment/earnings (based on a question about earnings *last year*). To account for the retrospective nature of the annual outcomes, the matching of parents at event time 0 (observed in year t with age a) and non-parents (observed in year $t - n$ with age $a - n$) is done for $n = 0, ..., 4$ to obtain synthetic observations for $\tau = -n - 1$. For the same reason, annual outcomes at event times $\tau = 0, ..., \bar{\tau}$ are obtained from parents observed at event times $\tau = 1, ..., \bar{\tau} + 1$.

[16] The difference in statistical precision is mainly due to the vastly different sample sizes. The PSID/NLSY event studies are based on an estimation sample of about 64,000 individual×year observations. The estimation samples used in the CPS/ACS pseudo-event studies are 143 times larger (weekly outcome) or 80 times larger (annual outcomes).

in the labor market outcomes of women. The child penalties 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 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, but the estimates are less precise and the pre-trends are less compelling.

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 introduce selection bias, evidenced by the 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. The problem is strongest for the earnings outcome, but is present for the employment outcomes as well. Adding matching variables reduce the size of these jumps, and the baseline specification in Panel D eliminates them almost entirely.[17]

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 (left panels) to actual event studies (right panels) using PSID/NLSY data for both.

The within-panel validation is exceedingly successful. The pseudo-event studies look very similar to the true event studies and produce virtually identical child penalties. If anything, the pseudo-event studies feature more convincing pre-trends than the true event studies based on the same data: the differences in pre-trends between men and women (mainly for earnings) disappear in the pseudo-event study specification. 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.

[17] Looking closely at the event studies for the baseline specification, there is still a tiny increase in the labor market outcomes of men between $\tau = -1$ and $\tau = 0$. This increase is too small to have any noticeable effect on the results.

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 the following difference-in-differences version of equation (7) estimated on the sample of men and women together:

$$Y_{it} = \alpha \cdot Post_{it} + \alpha^{w} \cdot Post_{it} \cdot \mathbf{1}\left[g = w\right] + \beta_{ag} + \gamma_{tg} + \nu_{it}. \tag{12}$$

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 calendar time, β_{ag} and γ_{tg}, are allowed to vary by gender as in equation (7). The object of interest is α^{w}, which captures the average effect of first child birth on women relative to women. This is the (unscaled) child penalty. Estimating equation (12) in the true panel and in the pseudo-panel, we obtain the estimates $\hat{\alpha}^{w}_{true}$ and $\hat{\alpha}^{w}_{pseudo}$. The Hausman test statistic is given by

$$H = \frac{\left(\hat{\alpha}^{w}_{true} - \hat{\alpha}^{w}_{pseudo}\right)^{2}}{\text{Var}\left(\hat{\alpha}^{w}_{true}\right) - \text{Var}\left(\hat{\alpha}^{w}_{pseudo}\right)}, \tag{13}$$

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 0.88) for annual employment, $H = 0.290$ (p-value of 0.59) for weekly employment, and $H = 0.099$ (p-value of 0.75) for earnings. These results strongly support the consistency of the pseudo-panel estimator.

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 child penalties estimated from pseudo-event studies against child penalties estimated from true event studies in different 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 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 different outcomes. This suggests that the pseudo-event study approach is consistent even in subsamples, lending support to the granular analyses presented below.

Predicted Fertility: Why does the pseudo-event study approach work so well? A potential reason is that the approach accurately predicts fertility (i.e., location in negative event time) among those observed without children. Figure A.4 of the online appendix 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).[18] 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 shown by the event study validations, the discrepancies between predicted and actual fertility do not destroy the accuracy of the approach. The reason is that, conditional on age and year fixed effects, the trajectory of labor market outcomes prior to child birth is virtually flat. In fact, this is a key advantage of the event study specification developed by Kleven, Landais, and Søgaard (2019). Given the flat trajectory prior to parenthood, exactly predicting fertility is not critical for the accuracy of the pseudo-event study approach.

Treatment-Effect Heterogeneity: Finally, we investigate the possibility of bias from treatment-effect heterogeneity by comparing results from the baseline specification to results from the stacked specification 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 10) and calculates a weighted

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

average treatment effect using the sample shares of each cohort (equation 11). 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 the pseudo-event study and event study approaches, but something that applies to both of them, we focus solely on our main design: pseudo-event studies using cross-sectional data.

The results are presented Appendix Figure A.5. The figure shows results from the baseline specification in the left panels and from the stacked specification in the right panels for each of the three labor market outcomes.[19] 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 average child penalty. In particular, the results for annual and weekly employment are virtually identical in the two specifications (the child penalties are exactly the same), 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.

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 almost perfectly parallel *and* post-event effects are almost perfectly persistent following a sharp effect at event time zero. In other words, there is virtually no dynamics in the data, except for the sharp effect precisely at event time $\tau = 0$. As described above, given we control flexibly for age and year effects, the variation in event time comes from heterogeneity in age at first birth. The fact that there is no dynamics outside $\tau = 0$ indicates 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 always 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.

Having provided a range of validation checks, the rest of the paper presents a detailed investigation of the variation in child penalties across time, space, and demographic/cultural groups. The analysis takes advantage of the statistical precision of the pseudo-event study approach to provide very granular evidence. This allows for a better understanding of mechanisms, focusing especially on the role of social norms and culture.

[19] 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.

4 Child Penalties in the United States

4.1 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 (7) is run for different time periods separately.[20] The event studies for each time period and labor market outcome are presented in Figures A.6-A.8 of the appendix. All of these look compelling, featuring parallel trends between men and women before child birth and sharp divergence immediately after child birth.

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). 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.[21] 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.

The preceding findings shed new light on a stylized fact documented elsewhere in the literature: the slowdown of gender convergence in labor market outcomes since the 1990s (e.g., Blau and Kahn 2006, 2017; Kuziemko, Pan, Shen, and Washington 2018). This trend has been viewed as

[20] 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.

[21] 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.

puzzling given the large 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 points to a simple 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.[22] 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.[23] This implies that the evolution of gender inequality in labor market outcomes — especially 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 overlapping 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.

4.2 Child Penalties Across States

To study the variation in child penalties across states, the event time dummies in equation (7) are interacted with state dummies.[24] In this specification, the lifecycle and time trends are estimated

[22] 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 $\equiv \frac{\mathbb{E}[Y_{it}^m|\tau\geq 0]-\mathbb{E}[Y_{it}^w|\tau\geq 0]}{\mathbb{E}\left[Y_{it}^m|\tau\geq 0\right]}$, 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.

[23] 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.

[24] 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."

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 child birth 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, Landais, Posch, Steinhauer, and Zweimüller 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 child birth, diverge immediately and sharply after child birth, 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 very 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 employment penalty ranges from 14.2% to 39.9% across states, and the earnings penalty ranges from 21.1% to 60.8% across states.[25]

As discussed, 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

[25] 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 measure effects precisely on the gender gap and, as we have seen, the 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.

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.[26]

Empirical Bayes Correction: A general concern with highly granular 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 state-level child penalties rather than on the underlying event studies, we focus on the difference-in-differences specification in equation (12), extended to allow for state-specific effects of child birth.[27] 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:

$$CP_s^* = \left(\frac{\sigma^2}{\sigma^2 + SE_s^2}\right) \cdot \widehat{CP}_s + \left(\frac{SE_s^2}{\sigma^2 + SE_s^2}\right) \cdot \mu, \tag{14}$$

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

[26] Notice also that these graphs serve as a basic validation check of the empirical approach, namely that the implied counterfactual employment rates are always below 100%.

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

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.[28]

Mechanisms: What are the underlying mechanisms responsible for the large variation in child penalties across space? The existing literature suggests that labor market structure, and especially the temporal flexibility and family friendliness of jobs, is an important determinant of gender gaps (Goldin 2014; Goldin and Katz 2016).[29] 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 cost 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 much smaller child penalties that urban states like New York, New Jersey, Massachusetts, and Connecticut.

While labor market structure is important, such general equilibrium effects cannot explain child penalties on their own. The flexibility and family friendliness of jobs would affect mothers and fathers equally absent other factors that tilt child care towards women. In other words, the lack of job flexibility may serve as an amplification mechanism, not as a stand-alone explanation. Traditional explanations turn 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, and given job structure is not an independent explanation, then what drives the variation in child penalties? This paper focuses on preference formation, presenting evidence on the role of gender norms and culture in section 5.

[28]In later analyses, I 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. I therefore take the simpler approach of using OLS estimates throughout.

[29]Related, Kleven, Landais, and Søgaard (2019) show that child birth induces women to move into more family-friendly firms and occupations (in exchange for lower pay).

4.3 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 child birth by each of these demographic characteristics 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 (7) is estimated separately for each group. Because child birth 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.[30] This finding contradicts explanations rooted in comparative advantage. If comparative advantage were important for explaining the large and persistent child penalties on women, we would expect low-educated women to have larger child penalties than high-educated women. The absence of such effects is consistent with findings for Denmark in Kleven, Landais, and Søgaard (2021). They use richer data to estimate relative earnings capacity within families, and show that long-run child penalties are unrelated to comparative advantage.

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 penalties for low-educated women relative to high-educated women. This serves to *strengthen* the absence of comparative advantage effects.

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 particu-

[30] 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.

larly 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 opposite Danish evidence presented in Kleven (2021). In Denmark, child penalties on single mothers are much *larger* than on married women. Therefore, while Danish child penalties are smaller than US child penalties on average, the penalties on single women are larger in Denmark than in the US. Kleven (2021) interprets this asymmetry as a side effect of the welfare system, presenting quasi-experimental evidence from US welfare reform in the 1990s. The idea is the following: given someone has to pay for children and 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. This is a luxury single mothers in America cannot afford.

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 larger 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.

5 The Effect of Gender Norms on Child Penalties

5.1 Child Penalties vs Gender Progressivity Over Time and Space

We start by considering the relationship between child penalties and elicited gender norms across time and space. This analysis is correlational and descriptive, not necessarily causal. The main objective is to investigate if existing cross-country evidence on the relationship between child penalties and elicited norms (e.g., Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019) survive when taking a more granular within-country perspective using state×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 the roles of men and women 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 is associated 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:

$$\text{Child Penalty}_{st} = \beta \cdot GPI_{st} + \gamma_s + \boldsymbol{\delta} \cdot \boldsymbol{X}_{st} + \nu_{st}. \tag{15}$$

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 controls $\boldsymbol{X}_{st}$. 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.[31]

Having estimated equation (15), child penalties are residualized using the estimated effect 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.[32] Binscatters for all three labor market

[31]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.

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

outcomes are presented in Figure 11. 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 ($\hat{\beta}$) 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 could have quantitatively important effects on child penalties. The use of granular within-country variation makes a causal interpretation more plausible than for existing cross-country evidence (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019), but the evidence is ultimately correlational. The variation in elicited gender norms is potentially endogenous, and the choice of controls involves a great deal of model uncertainty. Motivated by such concerns, the following sections consider a fundamentally different approach to estimating the impact of social norms: an epidemiological study of US movers and foreign immigrants.

5.2 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.[33] 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 analysis presented here considers the effect on child penalties, relying on granular event studies of child birth that are feasible due to the pseudo-event study methodology. The approach is first

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

applied to US movers and then to foreign immigrants, and together they provide striking visual evidence on the power of gender norms.

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 child birth 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 (7) is run separately for women movers and women stayers, interacting the event time dummies by state-of-birth dummies. Because child birth 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 child birth on a woman's employment is similar to the impact in the state where she was born, even though she lives somewhere else and is not directly exposed to the labor market institutions and public policies of that state. 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:

$$CP_s^{movers} = \beta \cdot CP_s^{stayers} + \gamma \cdot \widetilde{CP}_s^{movers} + \nu_s. \tag{16}$$

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.[34] As shown in Figure 13, the residualized and raw scatter plots look 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 (16), 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 characteristics of mothers who moved from states in the top and bottom quartiles of child penalties. The

[34]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.

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 mitigate concerns about selection on unobservables (Altonji, Elder, and Taber 2005), but such concerns cannot be ruled out entirely.

5.3 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. But the outcome variable is different and more challenging to study. For the analysis to be informative, child penalties have to be convincingly estimated at a granular level. Such estimates have not been available before, but become feasible using the pseudo-event study toolkit.

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.[35] The source-country penalties vary from 0% to 64% in the estimation sample.

[35]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

Figure 14 presents case studies of US immigrants from specific countries. The case studies include 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 child birth 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 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. The evidence is compelling across all countries: the pre-event trends are parallel, the post-event effects are immediate and persistent, and the coefficients are statistically precise.

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%).[36] Figure 16 extends these results to the full distribution of birth-country penalties. It provides binscatters of immigrant penalties against birth-country penalties by decile of birth-country penalties.[37] 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 0.522 implies that, as the employment penalty in a woman's country of birth increases by 10pp, her employment penalty in the US increases by 5.2pp. Because women living in the US are not directly affected by the incentives and institutions of their birth countries, this

the Czech and Slovak Republics allows for the inclusion of immigrants from the former Czechoslovakia, dissolved as an independent country on December 31, 1992.

[36] 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.

[37] 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.

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 that assuage such concerns. To investigate the selection of foreign immigrants, Panel B of Table 3 compares the 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 4.3 showed that child penalties are unrelated to education, alleviating concerns about selection in this dimension.[38] 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.[39] 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.

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

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

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 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 (7) 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 include all descendants with a known country of ancestry regardless of the time at which their ancestors arrived.

6 Conclusion

A recent literature shows that child penalties — the effect of child birth on women relative to men — account for most of the remaining gender inequality in developed countries (Kleven, Landais, and Søgaard 2019; Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019). In other words, eliminating gender inequality is virtually synonymous with eliminating child penalties. Understanding the mechanisms that drive child penalties is therefore one of the most important questions in gender inequality research. This paper contributes methodologically and empirically to this question.

As a methodological contribution, the paper develops a pseudo-event study approach to estimate child penalties that relies only on cross-sectional data. The approach can be validated against a true event study approach using panel data. The two approaches yield similar 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).

As an empirical contribution, the paper provides evidence on the variation in child penalties across time, geography, and demographic/cultural groups. There is large variation in child penalties across 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. These findings are consistent with the correlational evidence on child penalties and elicited gender norms in Kleven, Landais, Posch, Steinhauer, and Zweimüller (2019), but the granular mover/migrant designs presented here make the story much more conclusive.

References

ACS (2000-2019): *American Community Survey 2000-2019*. United States Census Bureau, https://usa.ipums.org/usa/ (accessed February 13, 2022). 1, 6

ALTONJI, JOSEPH G., AND REBECCA M. BLANK (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, chap. 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. 30

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*, ed. by O. Ashenfelter, and D. Card, vol. 4b, chap. 17. Elsevier: Amsterdam. 4

——— (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, 18

——— (2017): "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 55(3), 789–865. 2, 4, 18

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, 27

Bögl, Sarah, Jasmin Moshfegh, Petra Persson, and Maria Polyakova (2024): "The Economics of Infertility: Evidence from Reproductive Medicine," NBER Working Paper No. 32445. 4

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024): "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, Forthcoming. 4, 11

Browning, Martin (1992): "Children and Household Economic Behavior," *Journal of Economic Literature*, 30(3), 1434–1475. 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, 11

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, 27

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

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, 11

Deaton, Angus (1985): "Panel Data from Time Series of Cross-Sections," *Journal of Econometrics*, 30(1), 109–126. 4, 9

Fernández, Raquel (2011): "Does Culture Matter?," in *Handbook of Social Economics*, ed. by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, vol. 1, chap. 11. Elsevier: Amsterdam. 3, 27

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

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. 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. 3, 22

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. 3, 22

GOODMAN-BACON, ANDREW (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225(2), 254–277. 4, 11

GROES, FANE, ANNA HOUŠTECKÁ, DANIELA IORIO, AND RAÜL SANTAEULÀLIA-LLOPIS (2024): "The Unequal Battle Against Infertility: Theory and Evidence from IVF Success," Working Paper. 4

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, 25

KLEVEN, HENRIK (2021): "Lecture 3: Public Policy and Child Penalties," Zeuthen Lecture Series, September 2021. 24

——— (2024): "The EITC and the Extensive Margin: A Reappraisal," *Journal of Public Economics*, 236, 1–28. 6

KLEVEN, HENRIK, CAMILLE LANDAIS, AND GABRIEL LEITE-MARIANTE (2024): "The Child Penalty Atlas," *Review of Economic Studies*, Forthcoming. 2, 4, 5, 20, 30, 31, 34, 40, 53, 54, 77, 78, 79, 81

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, 12, 20, 25, 27, 33, 34

——— (2024): "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation," *American Economic Journal: Economic Policy*, 16, 110–149. 1, 4

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, 9, 11, 12, 16, 22, 33

——— (2021): "Does Biology Drive Child Penalties? Evidence from Biological and Adoptive Families," *American Economic Review: Insights*, 3(2), 183–198. 1, 3, 4, 22, 23

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, 18

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. 11

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

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, 11

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

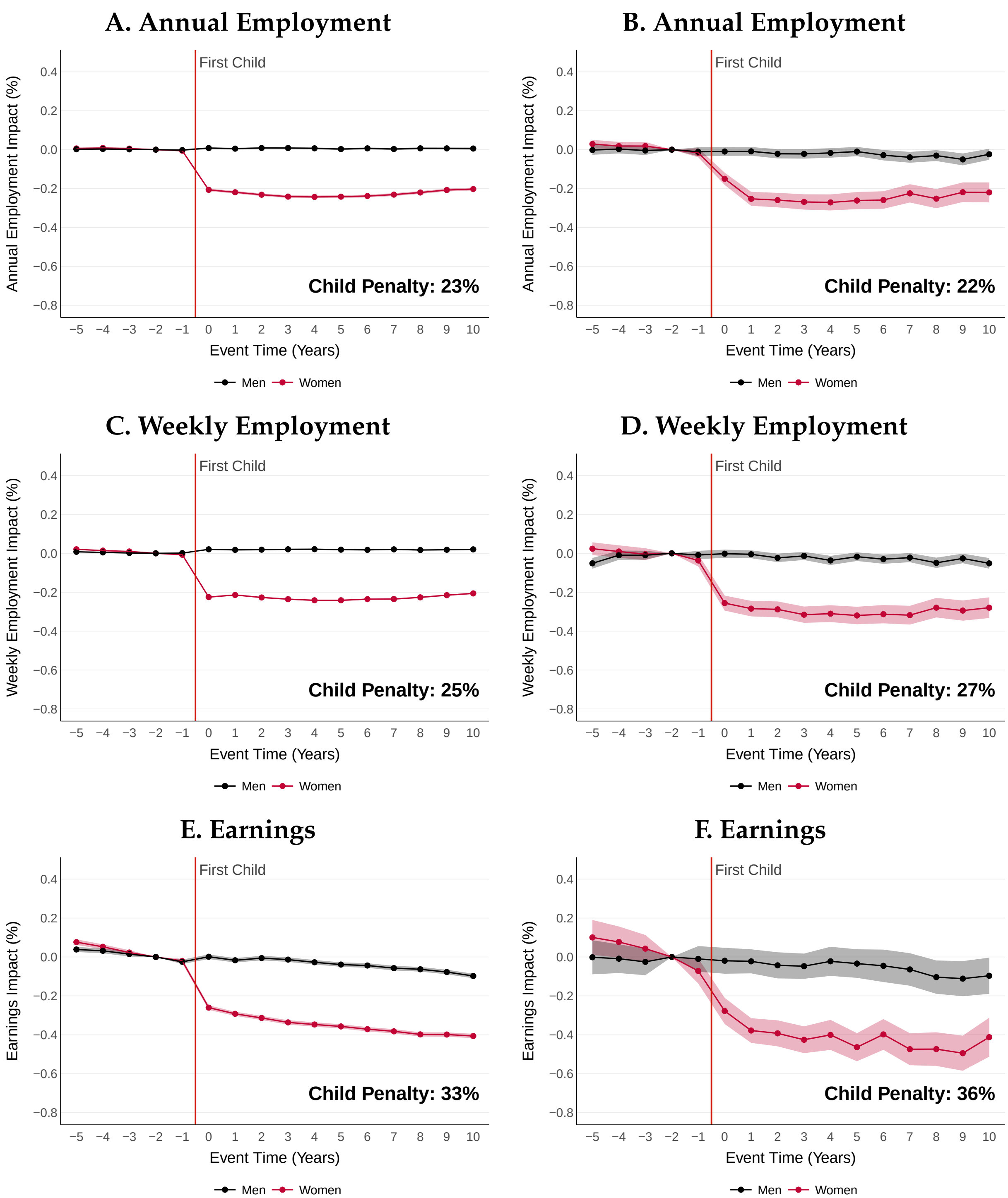
	A. Movers by State of Birth			B. Immigrants by Country of Birth		
	High-Penalty States	Low-Penalty States	**Difference**	High-Penalty Countries	Low-Penalty Countries	**Difference**
Fraction in High-Penalty States	0.25	0.18	**0.07**	0.19	0.20	**-0.01**
Fraction High School or Below	0.12	0.11	**0.01**	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

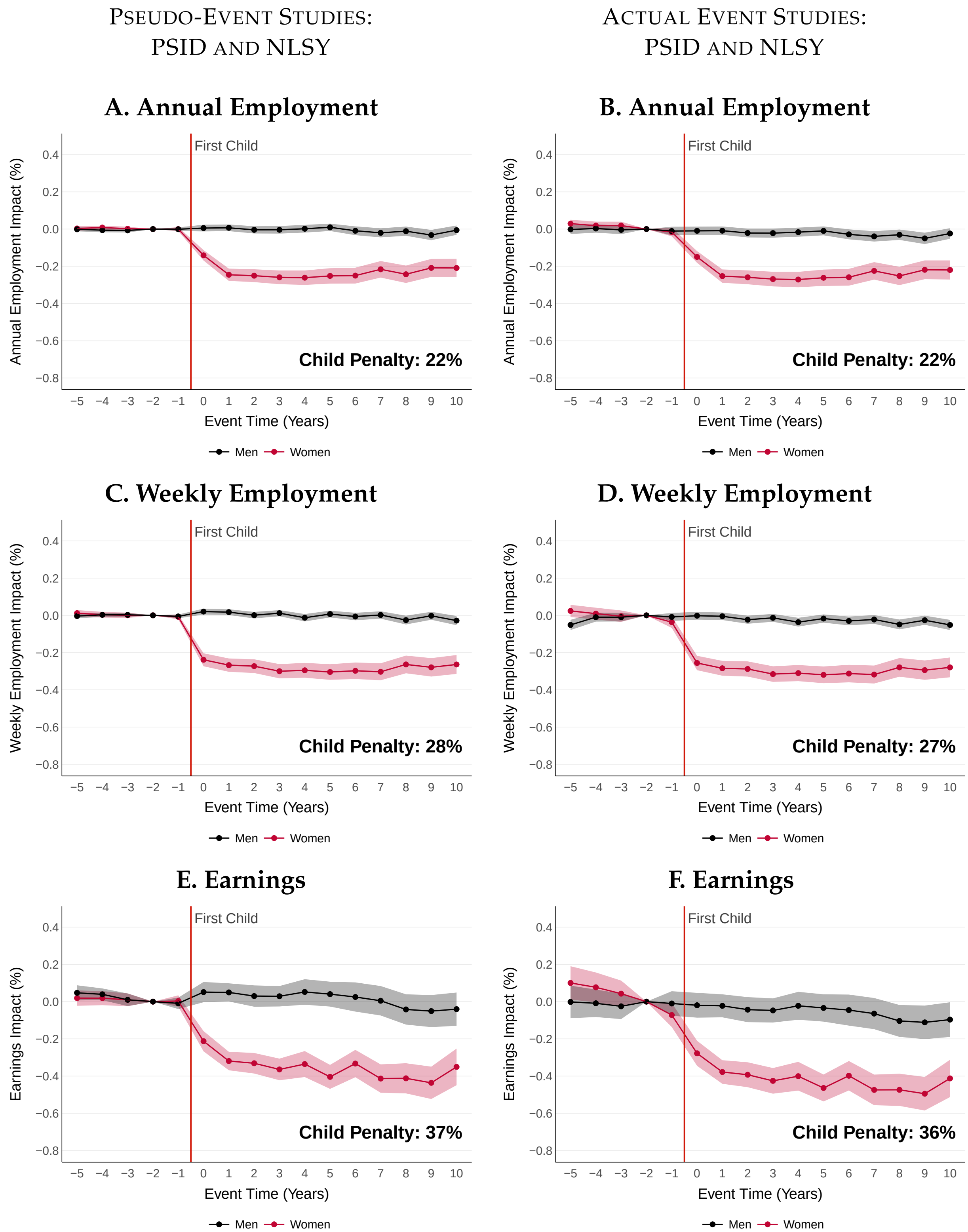
PSEUDO-EVENT STUDIES: CPS AND ACS

ACTUAL EVENT STUDIES: PSID AND NLSY



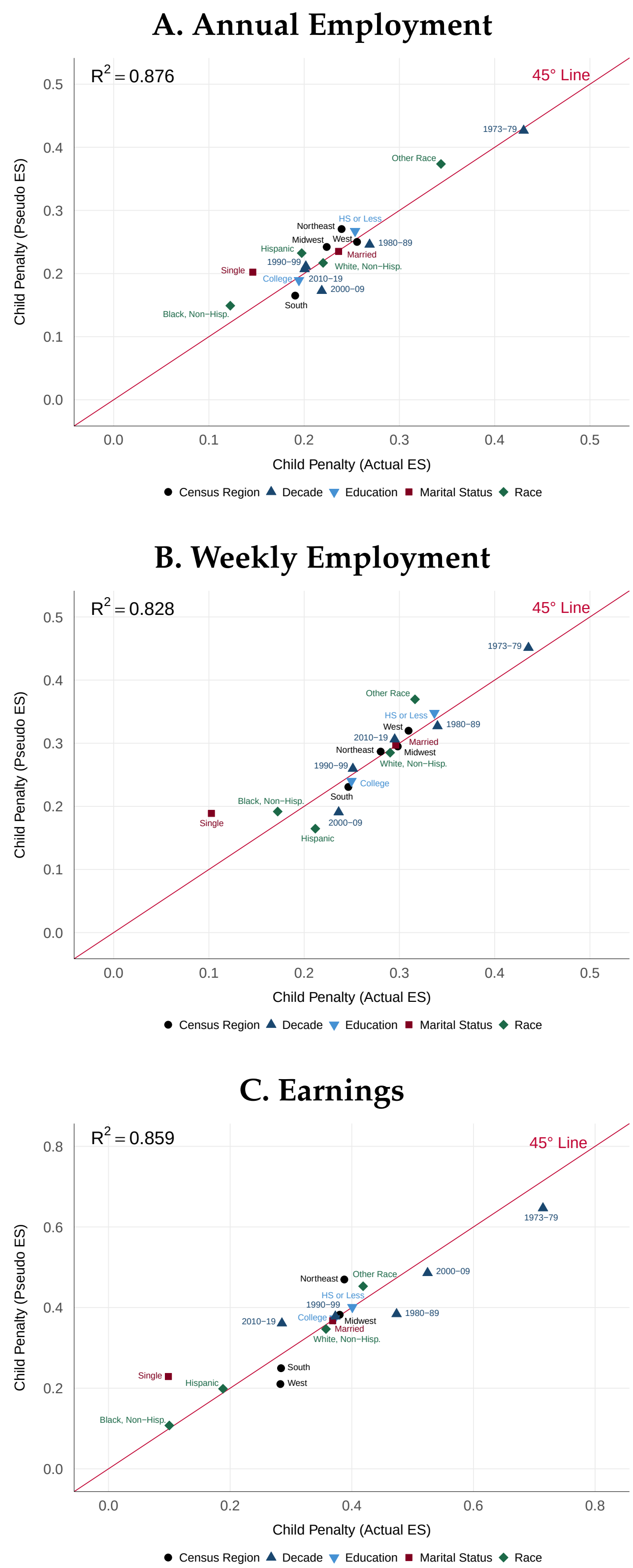
Notes: This figure validates the pseudo-event study approach (left panels) against an actual event study approach (right panels). The pseudo-event studies are based on pooled CPS and ACS data from 1968-2020, while the actual event studies are based on pooled PSID and NLSY data from 1968-2019. Each panel 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 child birth on men and women at each event time τ, i.e. $\hat{P}^m_\tau$ and $\hat{P}^w_\tau$ estimated from equations (7)-(8). Each panel also displays the average child penalty over event times 0-10 defined in equation (9). 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



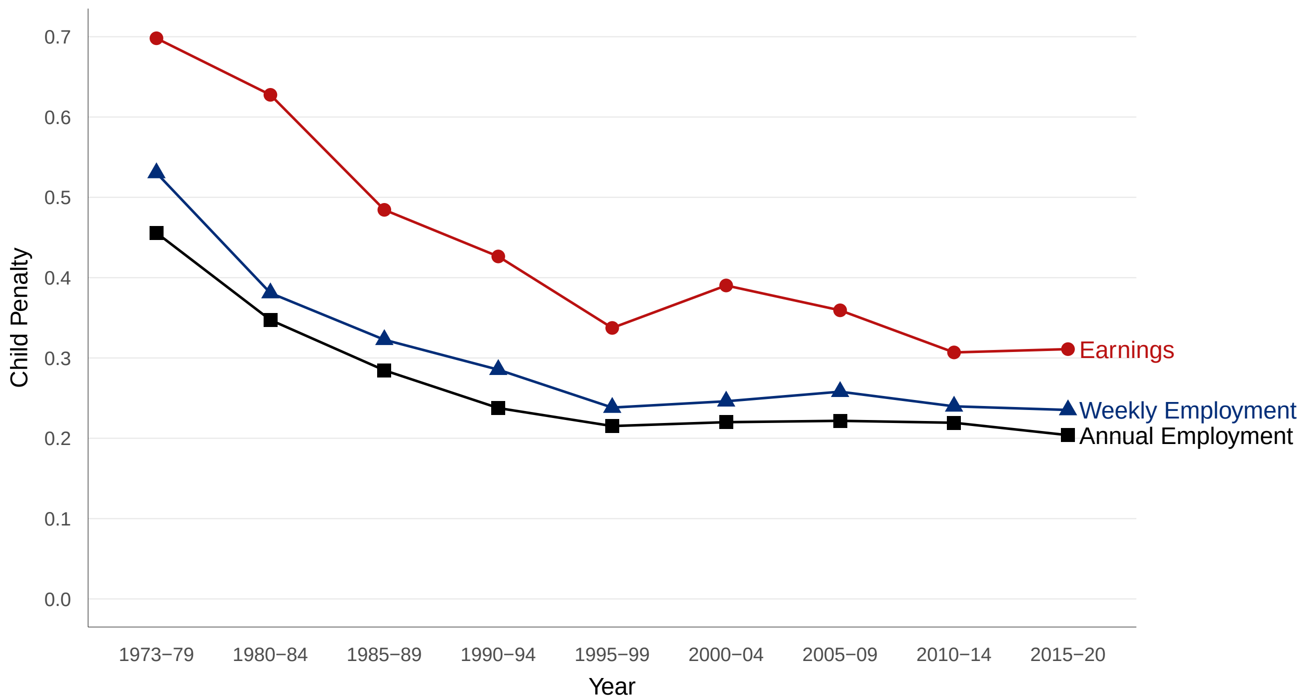
Notes: This figure validates the pseudo-event study approach (left panels) against an actual event study approach (right panels), both using pooled PSID and NLSY data from 1968-2019. Each panel shows an event study for men and women around the birth of their first child at $\tau = 0$. The series show the percentage impact of child birth on men and women at each event time τ, i.e. $\hat{P}^{m}_{\tau}$ and $\hat{P}^{w}_{\tau}$ estimated from equations (7)-(8). Each panel also displays the average child penalty over event times 0-10 defined in equation (9). 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



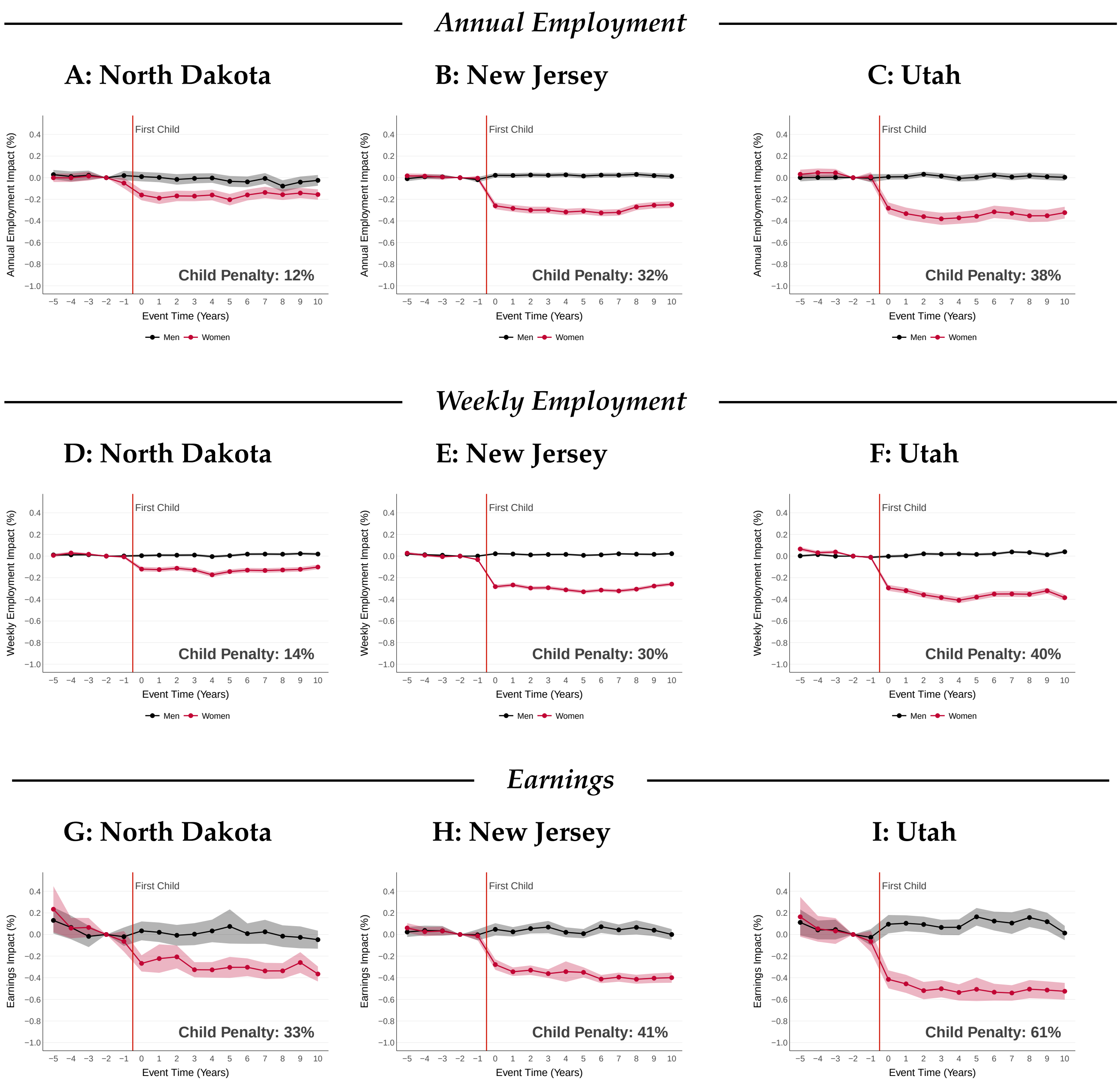
Notes: This figure validates the pseudo-event study approach in subsamples using pooled PSID/NLSY data. For each labor market outcome, the figure plots child penalties estimated from pseudo-event studies against child penalties estimated from true event studies in subsamples. The sample is split by geography (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 9) in different time intervals. These are estimated by splitting the sample of parents by interview year and running the event study specification (7) 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.

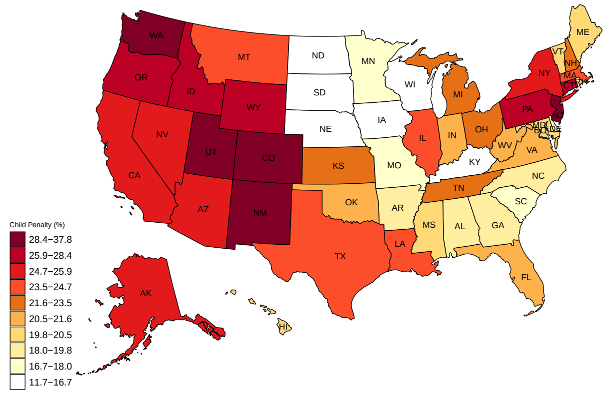
Figure 5: Event Studies of First Child Birth in Selected States



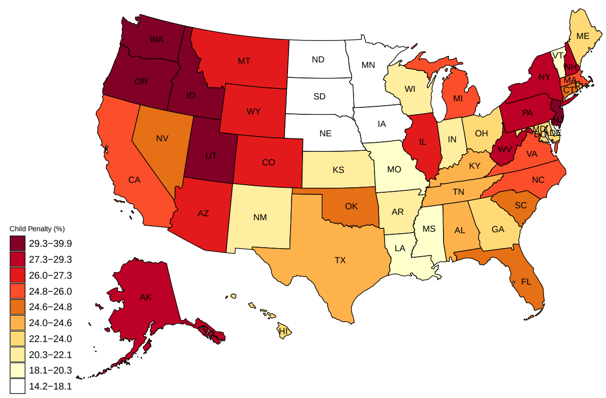
Notes: This figure shows event studies of first child birth for three US states and each of the three labor market outcomes. State-level event studies are constructed by interacting the event time dummies in equation (7) with state dummies, estimating percentage impacts of child birth on men and women at each event time ($\hat{P}_{\tau}^{m}$ and $\hat{P}_{\tau}^{w}$) 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 (7) 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

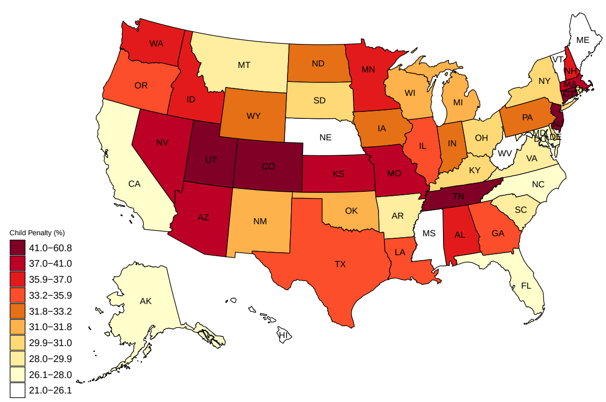
A. Annual Employment



B. Weekly Employment

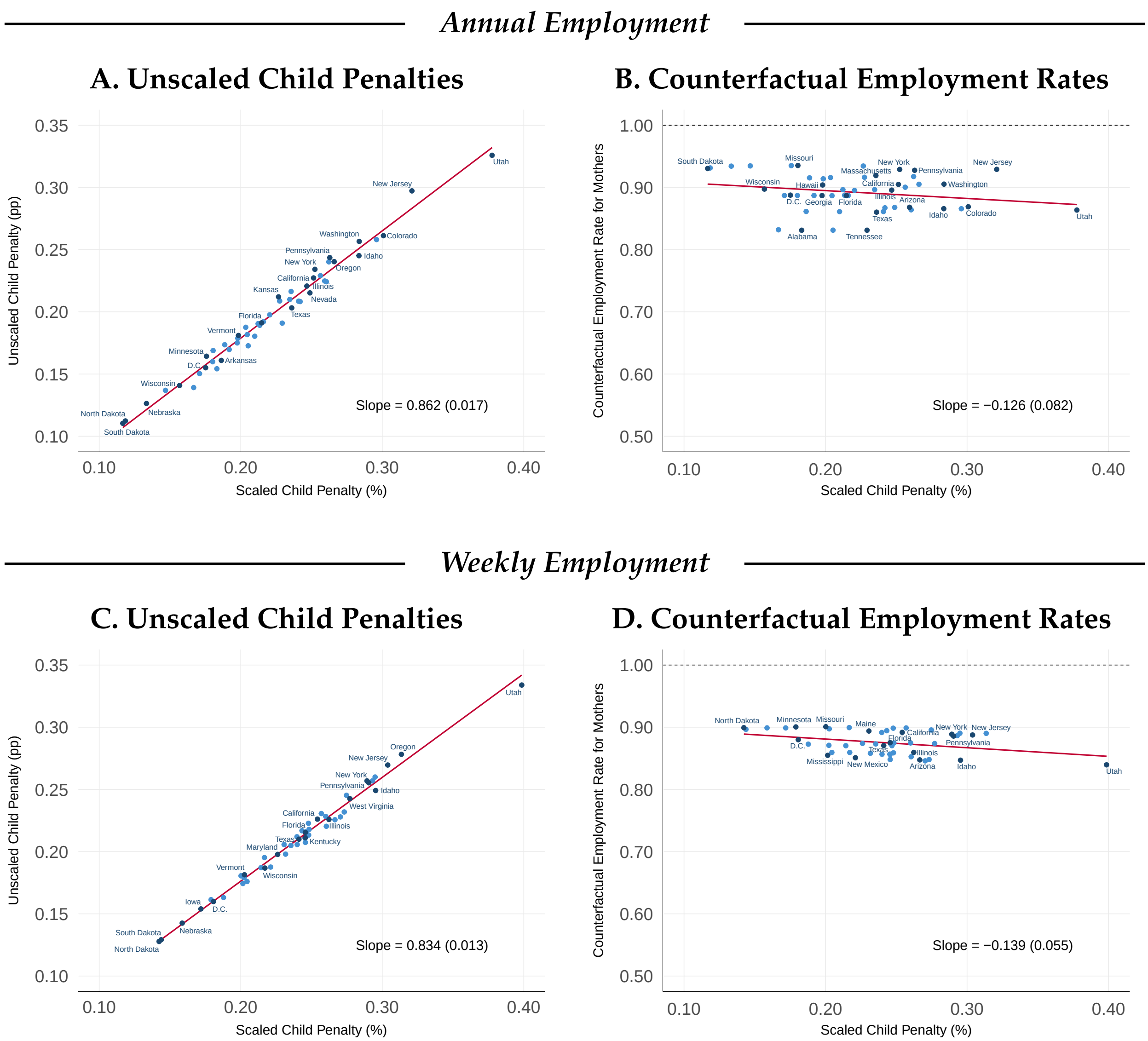


C. Earnings



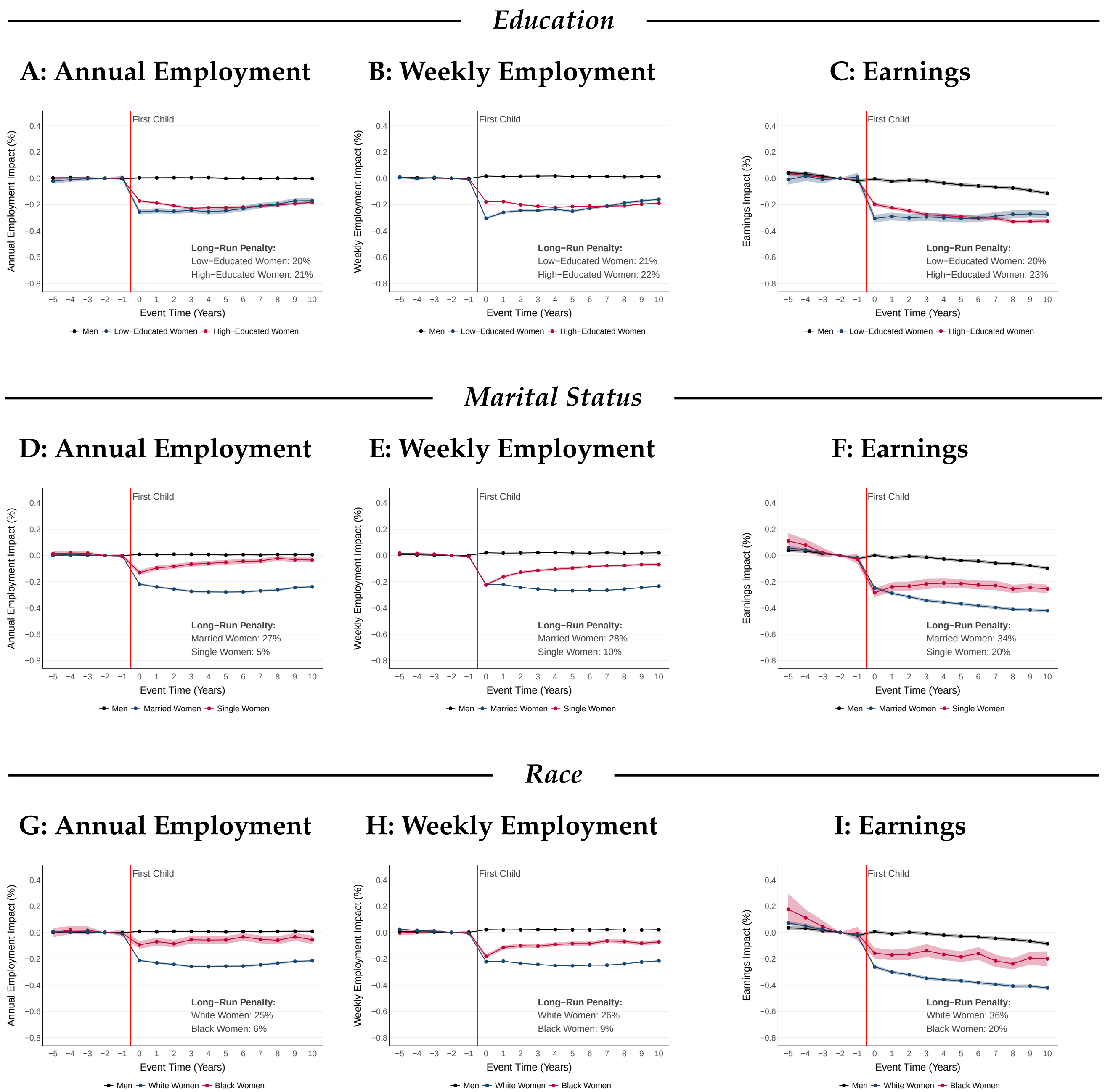
Notes: This figure summarizes the results from the state-level event studies of child birth (shown in Figures A.11-A.13 of the appendix) in heatmaps. In these maps, states are divided into deciles of the child penalty (as defined in equation 9), with darker colors implying larger child penalties.

FIGURE 7: UNSCALED VS SCALED CHILD PENALTIES ACROSS STATES



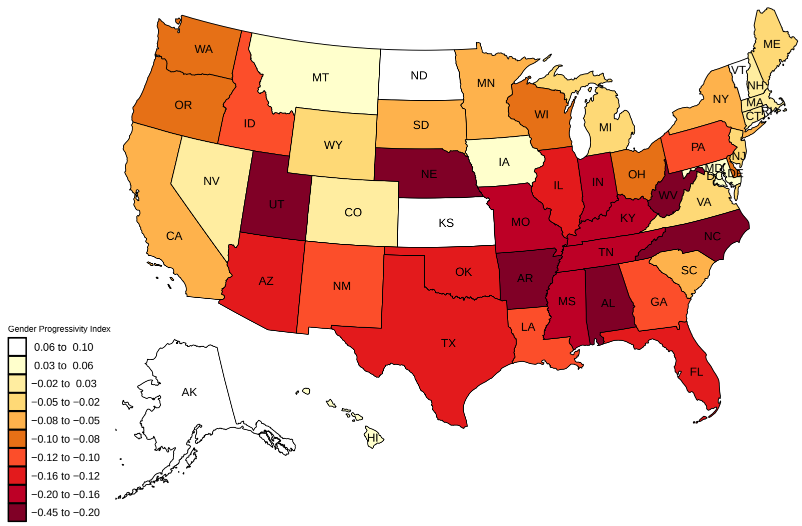
Notes: This figure investigates if the spatial variation in scaled child penalties (effects in percentages) reflects mostly variation in unscaled child penalties (effects in absolute terms) or in the counterfactual levels used for scaling. Results are shown for the two employment outcomes. The left panels 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 (7) 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



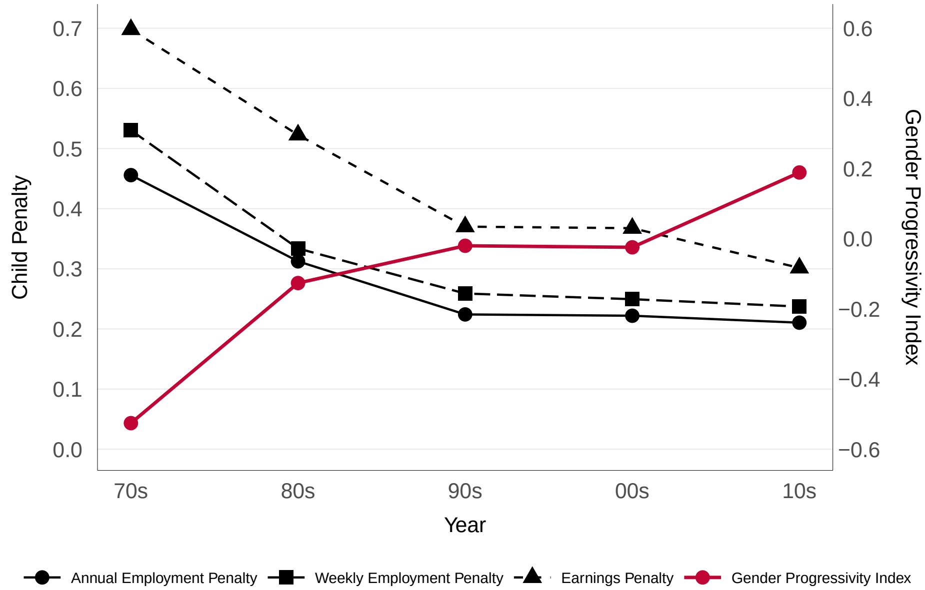
Notes: This figure presents event studies of first child birth by education, marital status, and race. To construct the figure, the sample of women is split into different demographic groups and specification (7) is estimated separately for each group. The sample of men is not split by demographics as child birth 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

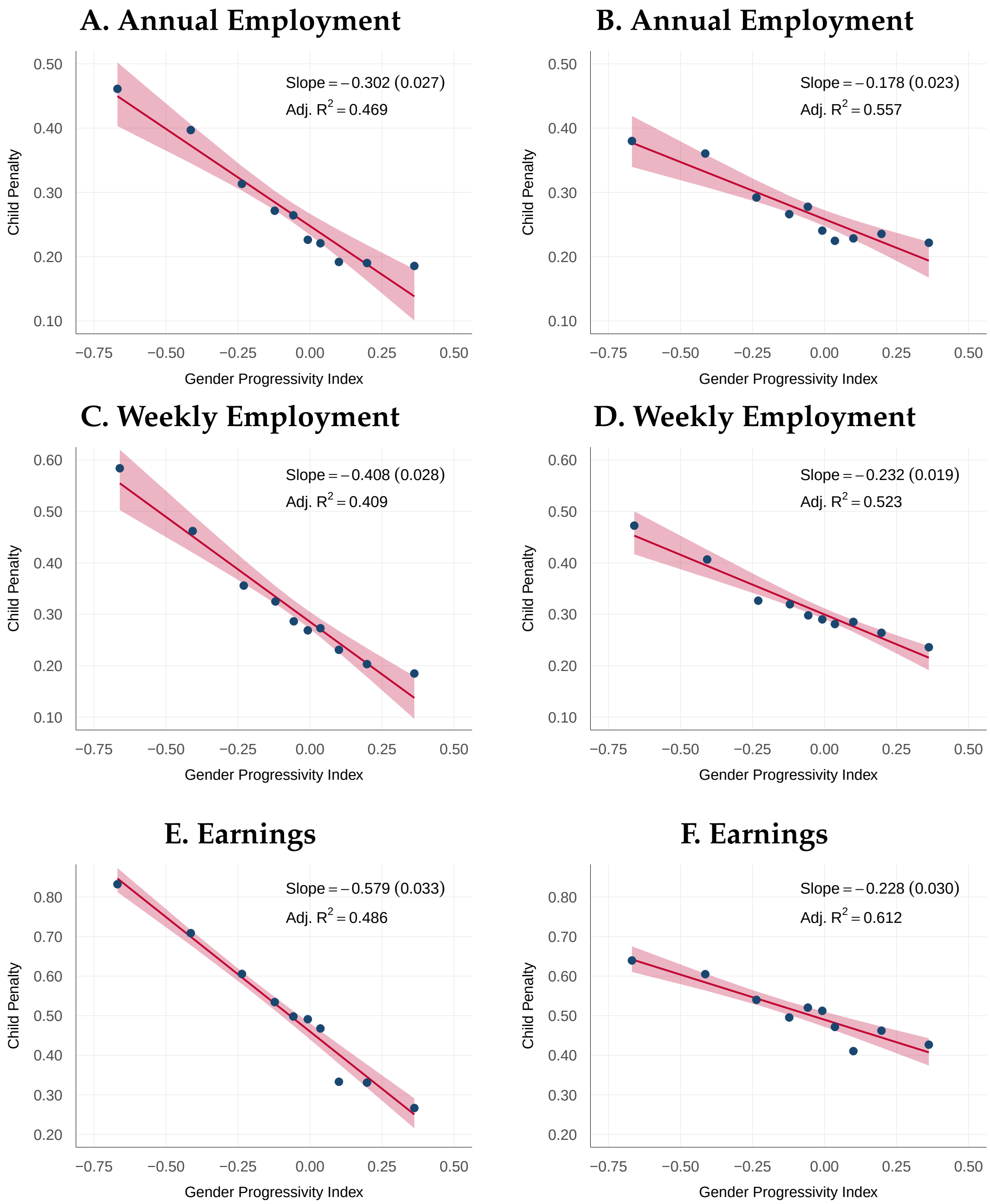


Notes: This figure plots the evolution of child penalties and gender progressivity over the last 50 years. The construction of the Gender Progressivity Index (GPI) is described in the notes to the preceding figure. The GPI time series is obtained by taking an average of state-level GPIs within each decade,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

Figure 11: Child Penalties vs Gender Norms Across States and Time

Binscatter:
State Fixed Effects

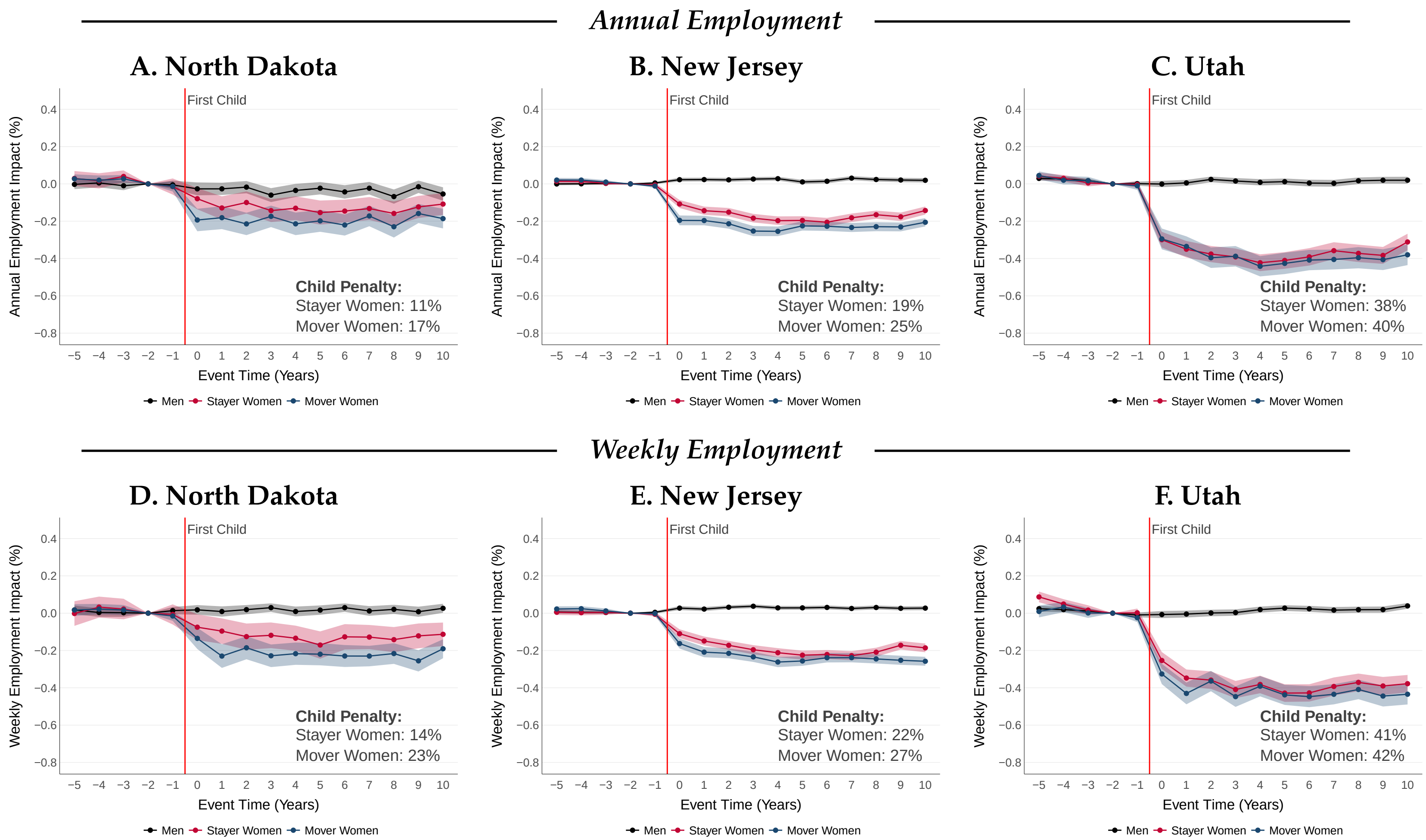
Binscatter:
State Fixed Effects & Demographics



Notes: This figure provides binscatters of child penalties vs gender progressivity across states and time. The analysis is based on equation (15), 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 CHILD BIRTH FOR MOVERS VS STAYERS BY STATE OF BIRTH



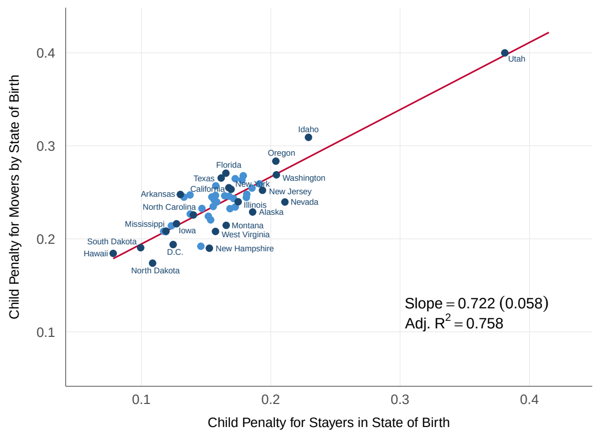
Notes: This figure presents event studies of first child birth for movers and stayers born in different states. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. To construct the figure, specification (7) 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 child birth 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

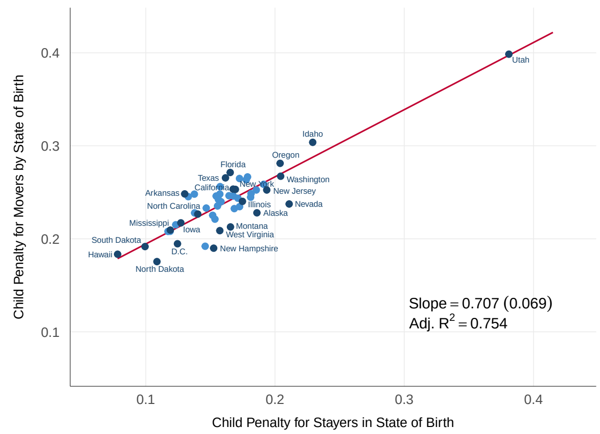
Child Penalties for Movers vs Stayers by State of Birth

Annual Employment

A. Raw Data

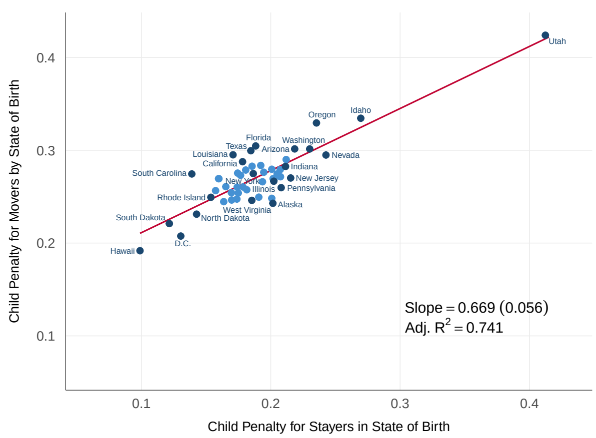


B. Residualize by Effect of State-of-Residence Penalties

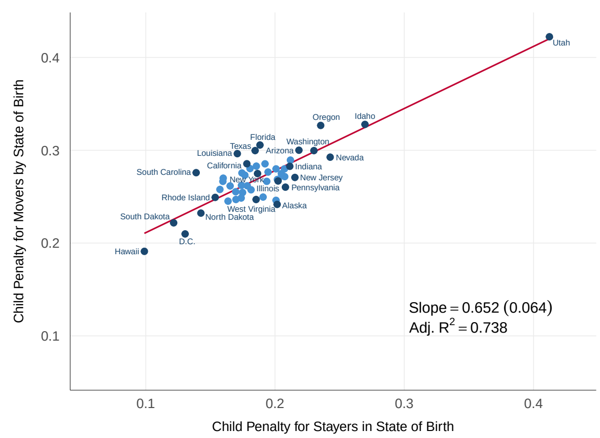


Weekly Employment

C. Raw Data



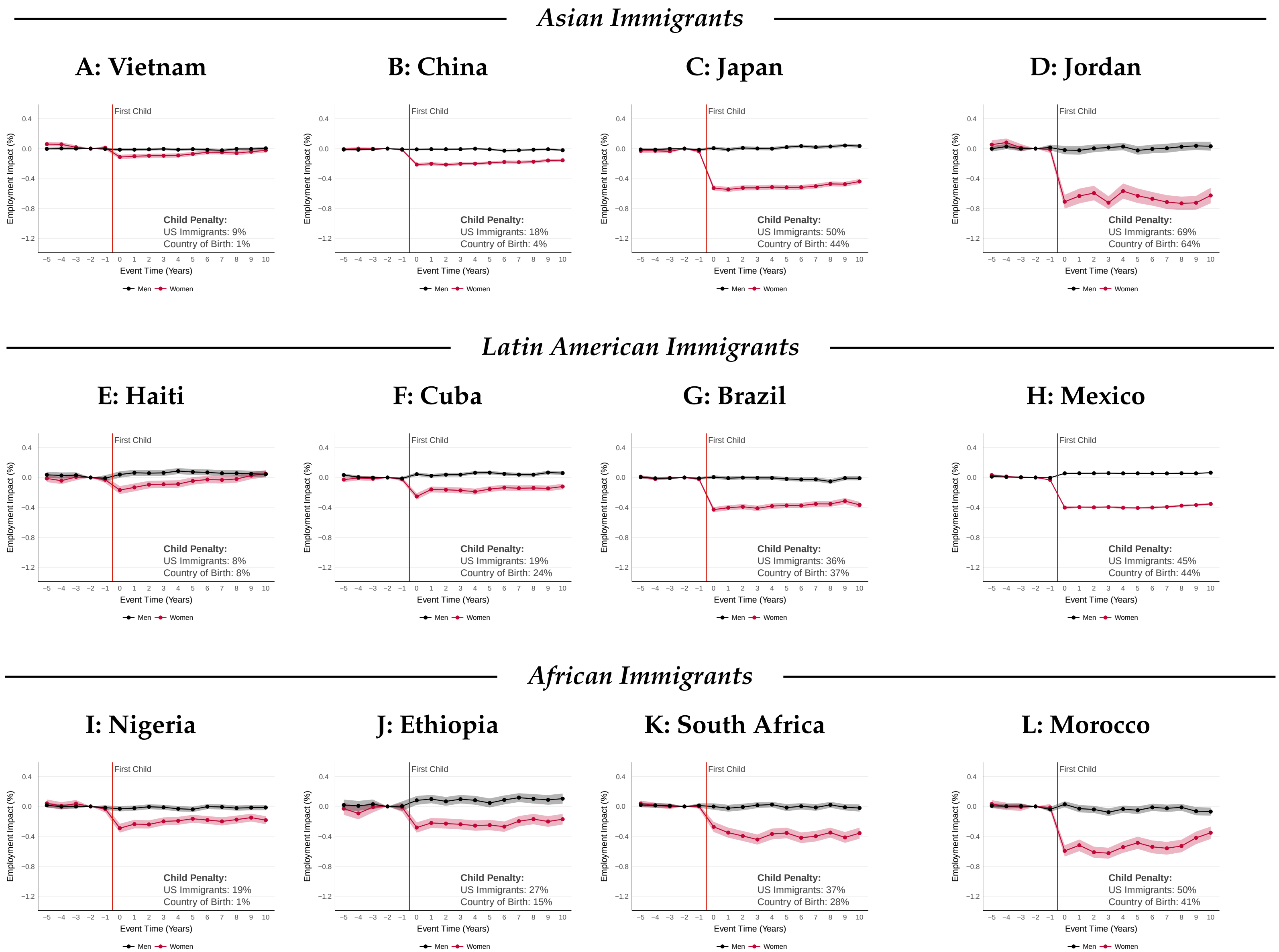
D. Residualize by Effect of State-of-Residence Penalties



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. (16). 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 CHILD BIRTH FOR IMMIGRANTS BY COUNTRY OF BIRTH

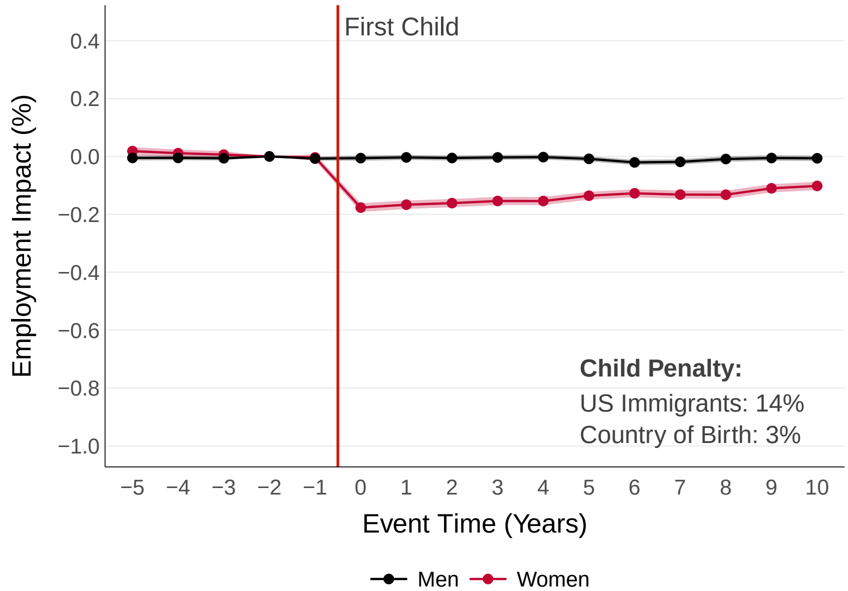


Notes: This figure presents event studies of first child birth for foreign-born immigrants by country of birth. Results are shown for selected countries in Asia (top row), Latin America (middle row), and Africa (bottom row). The results for all 81 countries in the sample are provided in Appendix Figure 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.

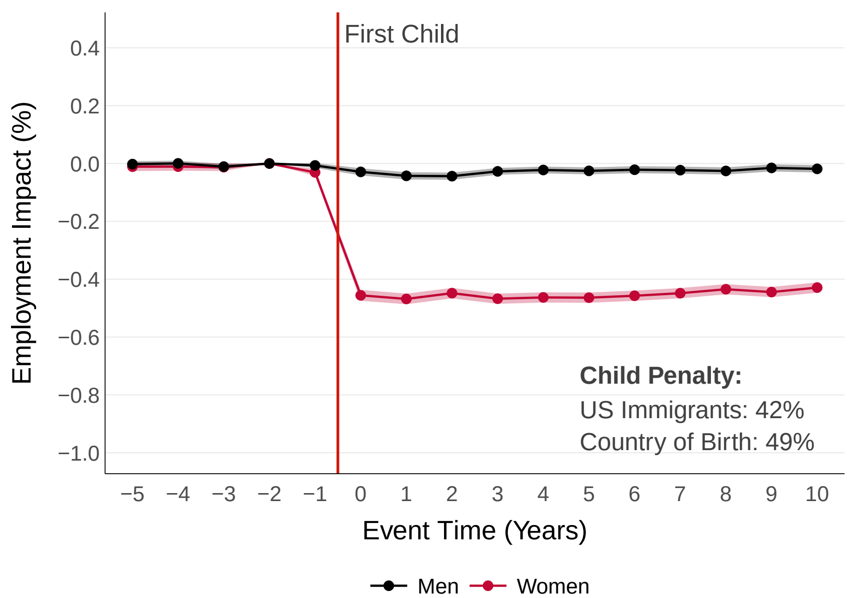
Figure 15: Epidemiological Study of Foreign Immigrants

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

A. Bottom Decile of Child Penalty in Country of Birth



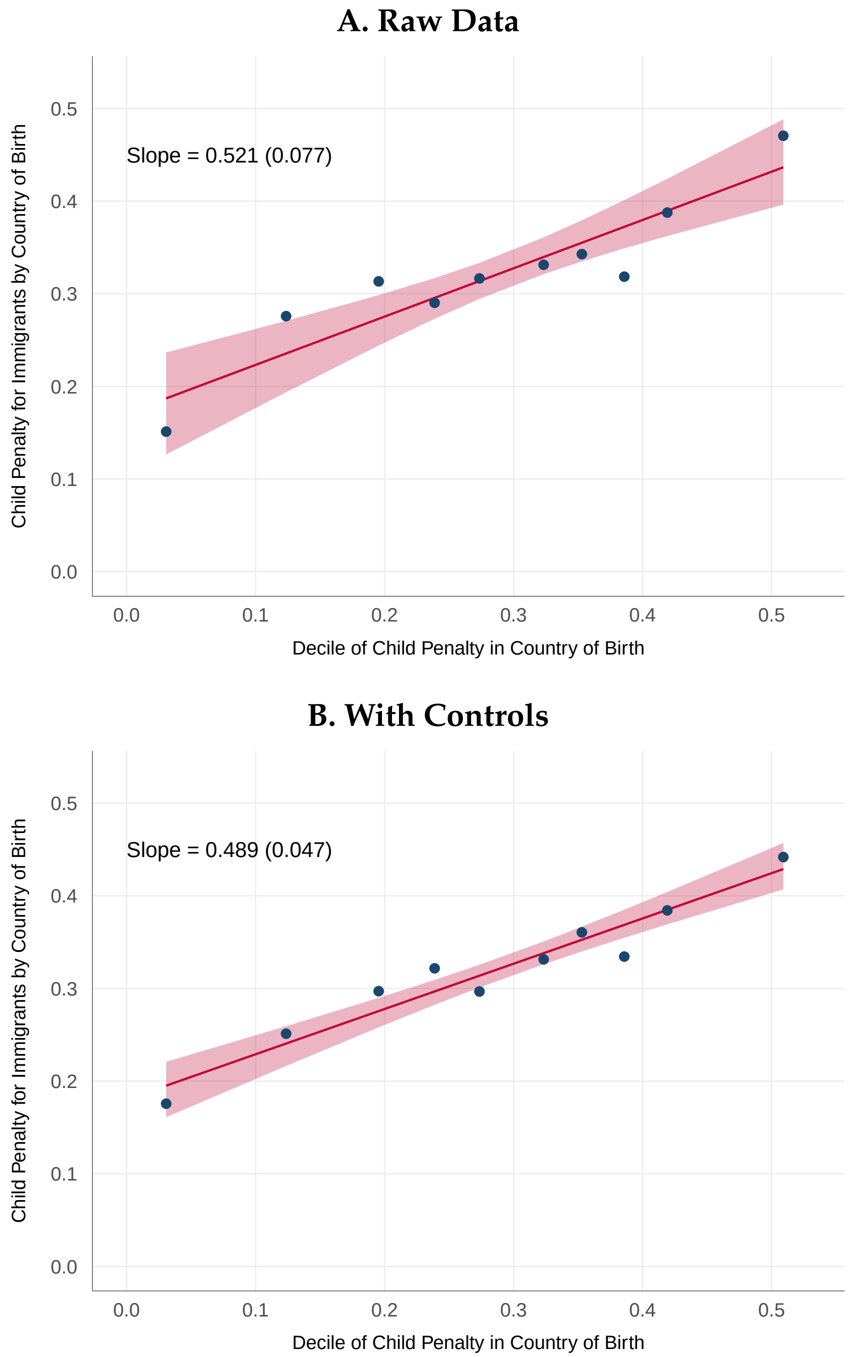
B. Top Decile of Child Penalty in Country of Birth



Notes: This figure presents event studies of first child birth 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 (7) separately for each decile, graphing the percentage impacts on men and women at each event time τ (as defined in equation 8). 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

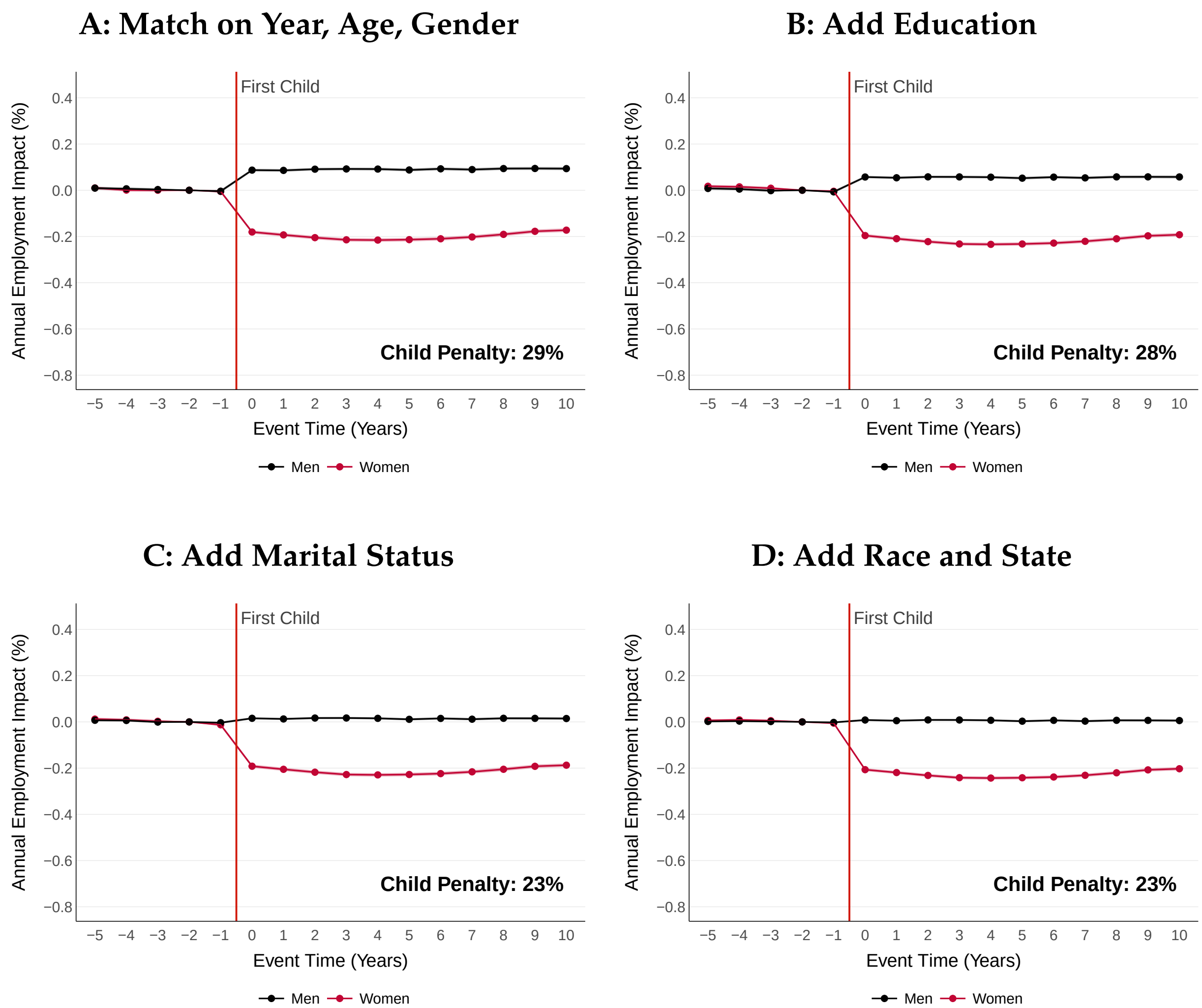


Notes: This figure presents binscatters of child penalties for foreign-born immigrants by decile of the child penalty in country of birth. Panel A shows raw child penalty estimates, while Panel B controls for differences in education, marriage, race, fertility, age at first birth, and US location across 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 (Not for Publication)

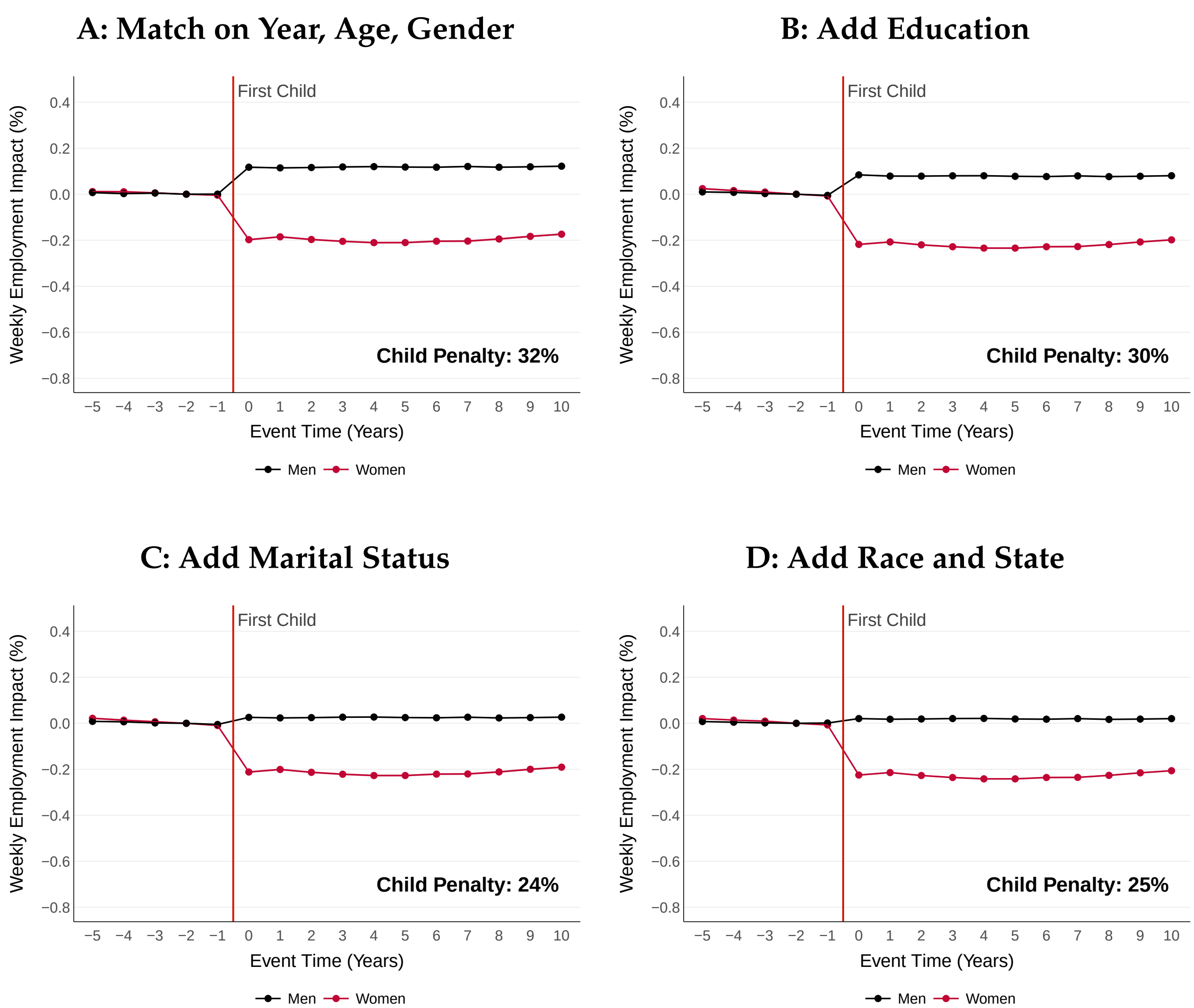
A Supplementary Exhibits

FIGURE A.1: PSEUDO-EVENT STUDIES UNDER DIFFERENT MATCHING SPECIFICATIONS
ANNUAL EMPLOYMENT



Notes: This figure presents pseudo-event studies of first child birth 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.

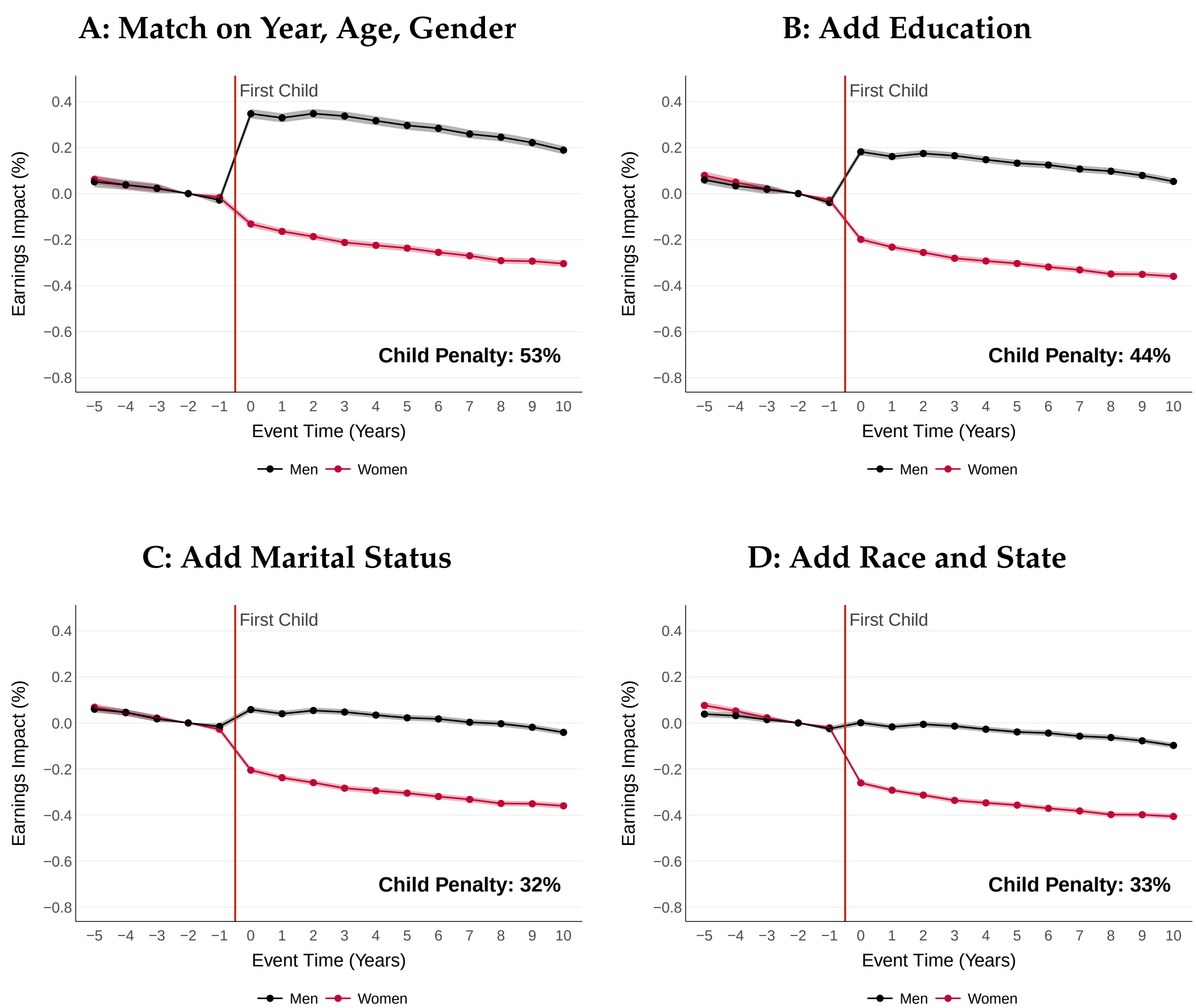
FIGURE A.2: PSEUDO-EVENT STUDIES UNDER DIFFERENT MATCHING SPECIFICATIONS
WEEKLY EMPLOYMENT



Notes: This figure presents pseudo-event studies of first child birth 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 Under Different Matching Specifications

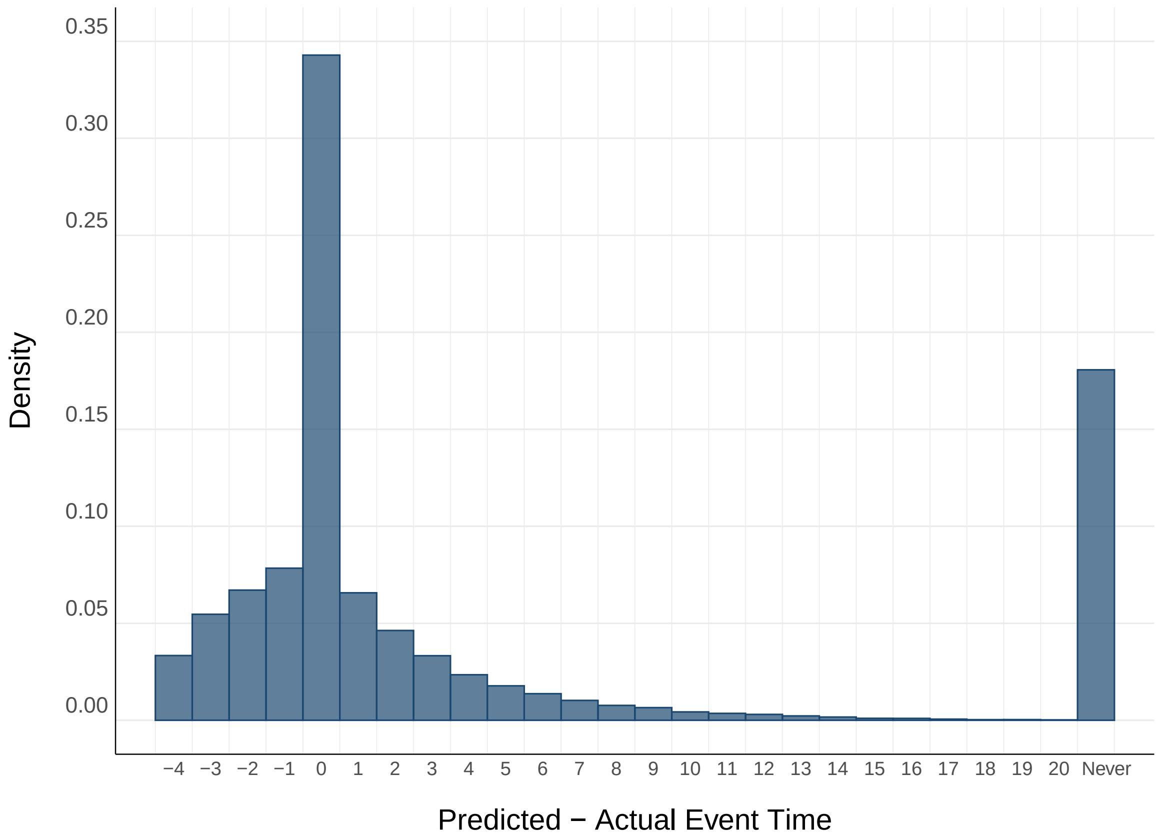
Earnings



Notes: This figure presents pseudo-event studies of first child birth 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

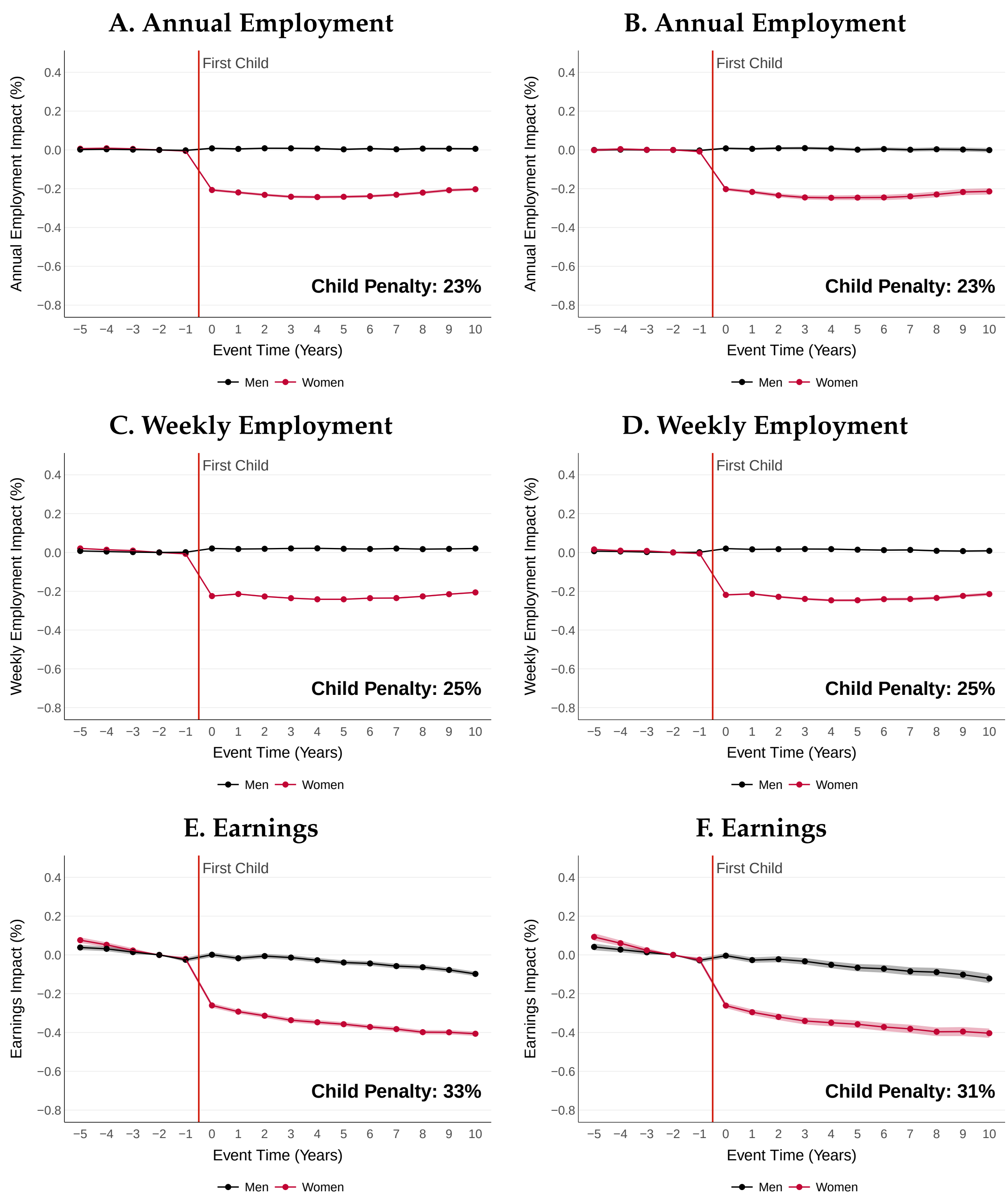
PREDICTED VS ACTUAL EVENT TIMES AMONG CHILDLESS PEOPLE



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

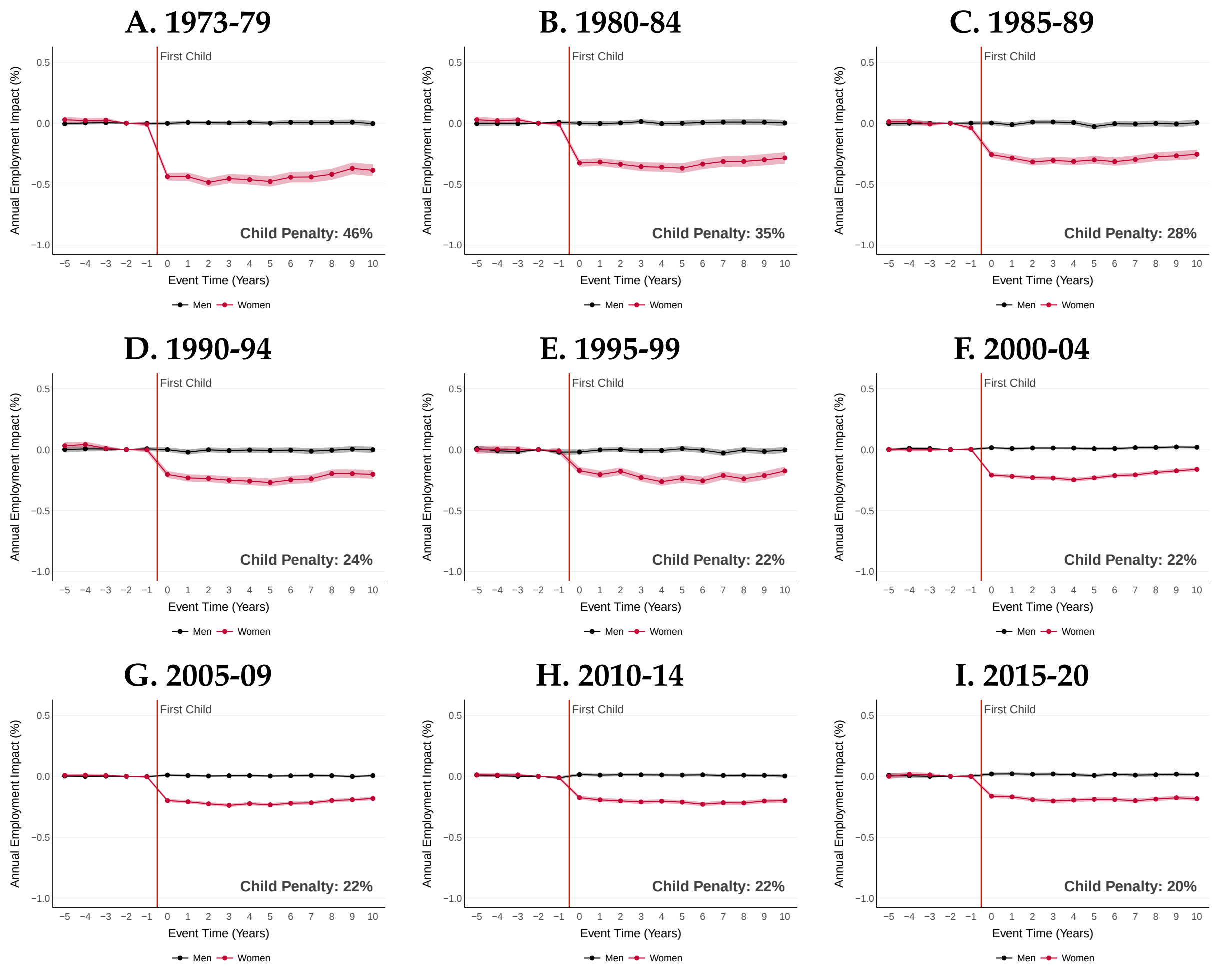
Baseline Specification: Birth Cohorts Pooled | Stacked Specification: Birth Cohorts Stacked



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 10) and calculates a weighted average treatment effect using the sample shares of each cohort (as specified in equation 11). 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 CHILD BIRTH OVER TIME

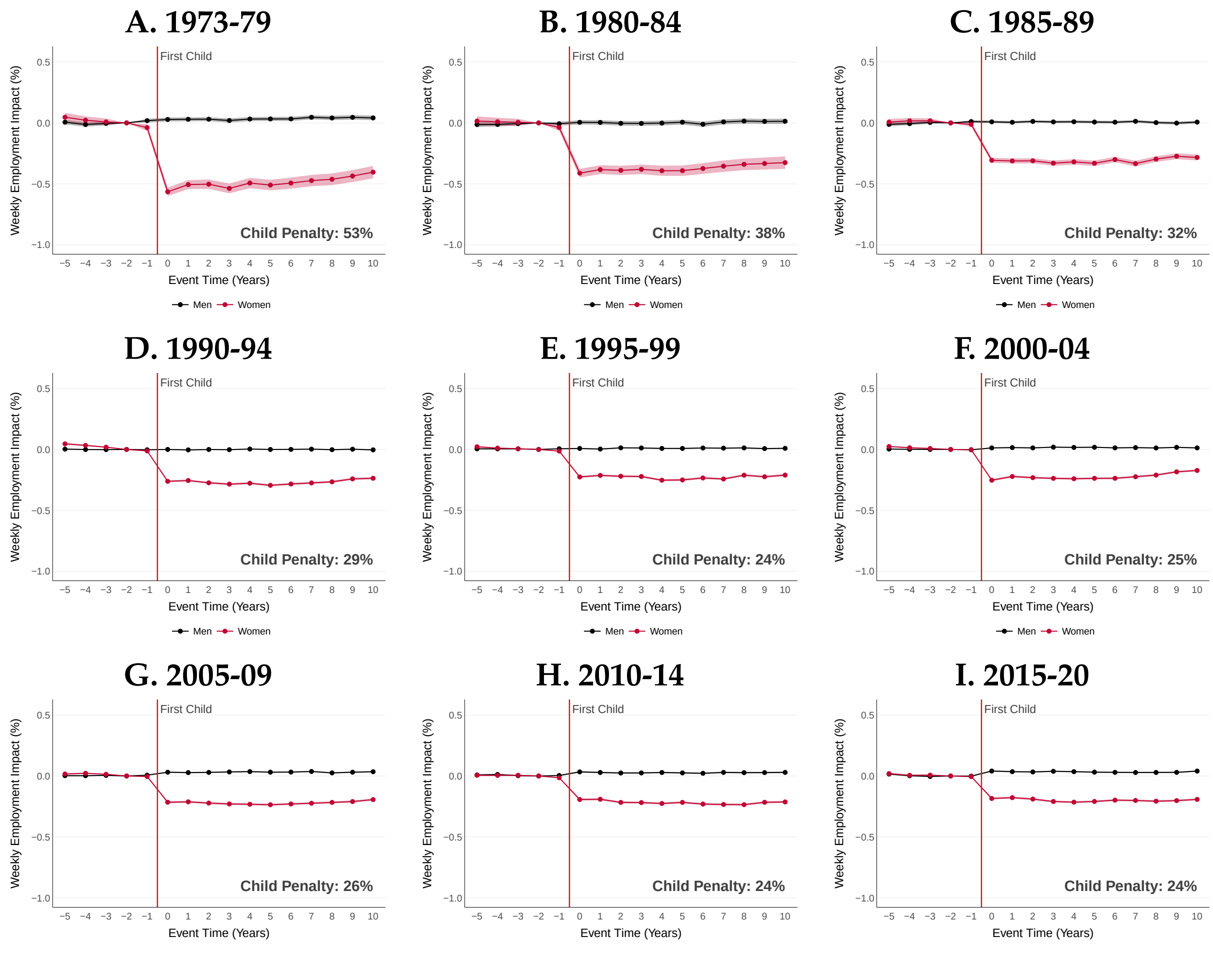
ANNUAL EMPLOYMENT



Notes: This figure shows event studies of first child birth for annual employment in different time periods. The sample of parents is split by interview year and the event study specification (7) 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 9) for the time period in question. The 95% confidence intervals are based on robust standard errors.

FIGURE A.7: EVENT STUDIES OF FIRST CHILD BIRTH OVER TIME

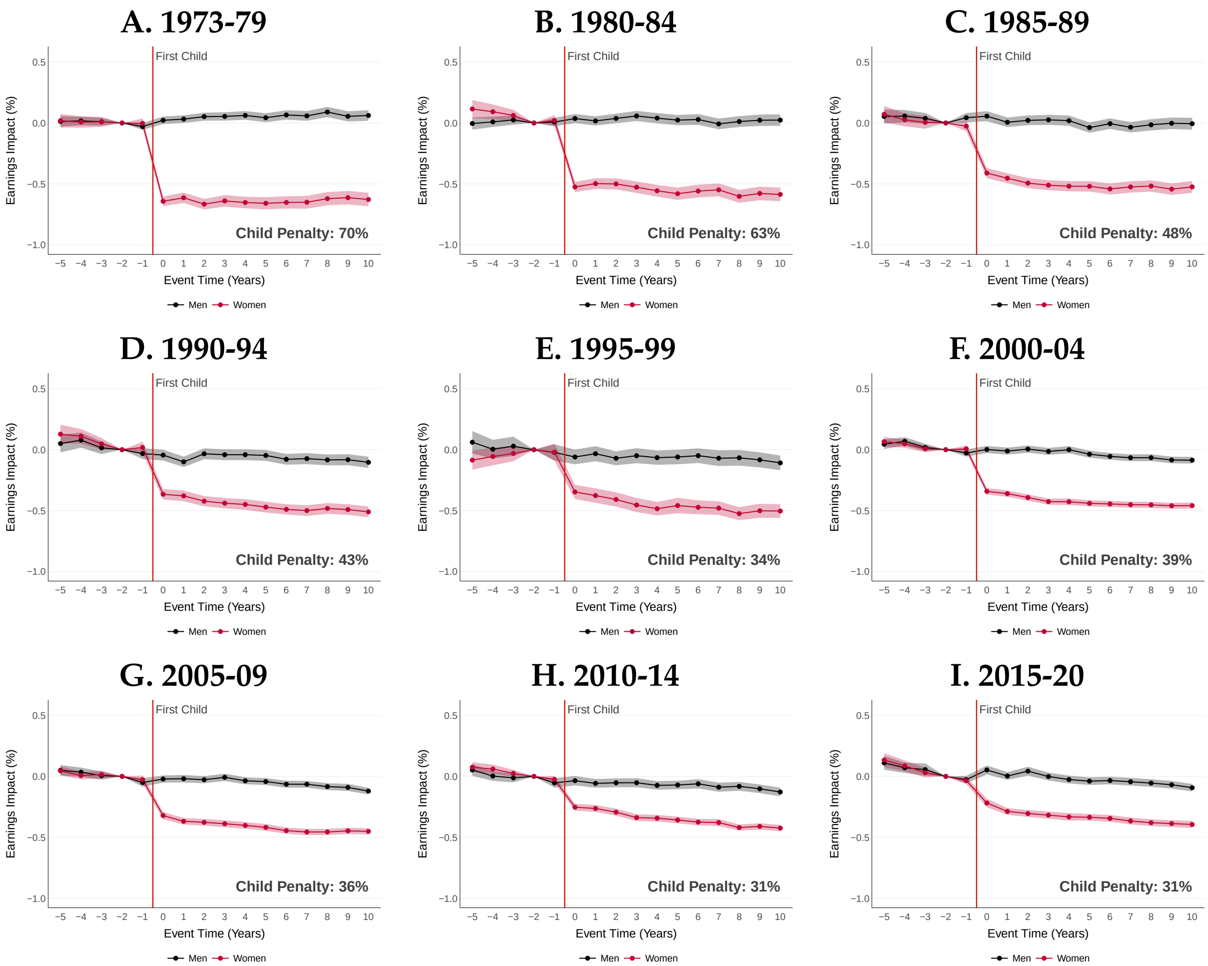
WEEKLY EMPLOYMENT



Notes: This figure shows event studies of first child birth for weekly employment in different time periods. The sample of parents is split by interview year and the event study specification (7) 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 9) for the time period in question. The 95% confidence intervals are based on robust standard errors.

FIGURE A.8: EVENT STUDIES OF FIRST CHILD BIRTH OVER TIME

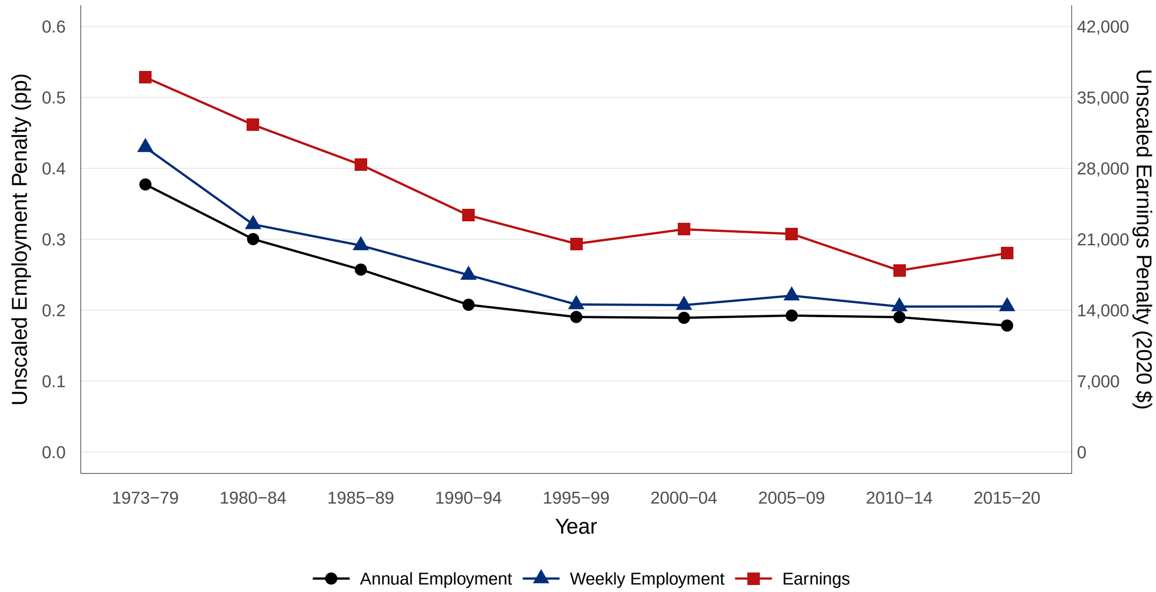
EARNINGS



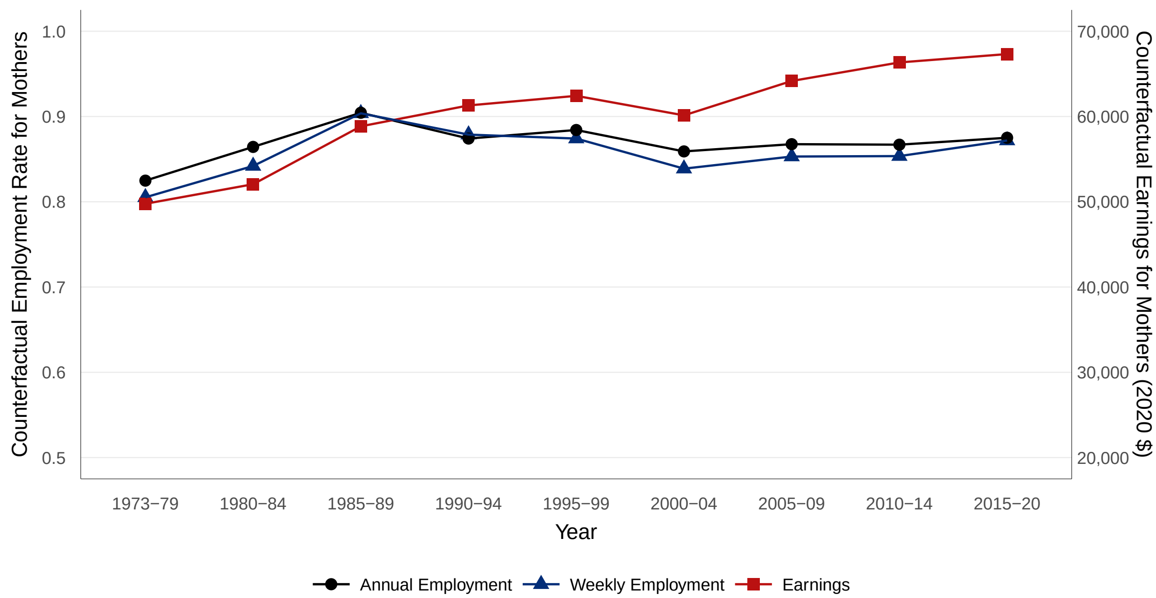
Notes: This figure shows event studies of first child birth for earnings in different time periods. The sample of parents is split by interview year and the event study specification (7) is run separately for each time period. The event studies start in 1973, because the first five years of the data (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 9) 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

A. Unscaled Child Penalties

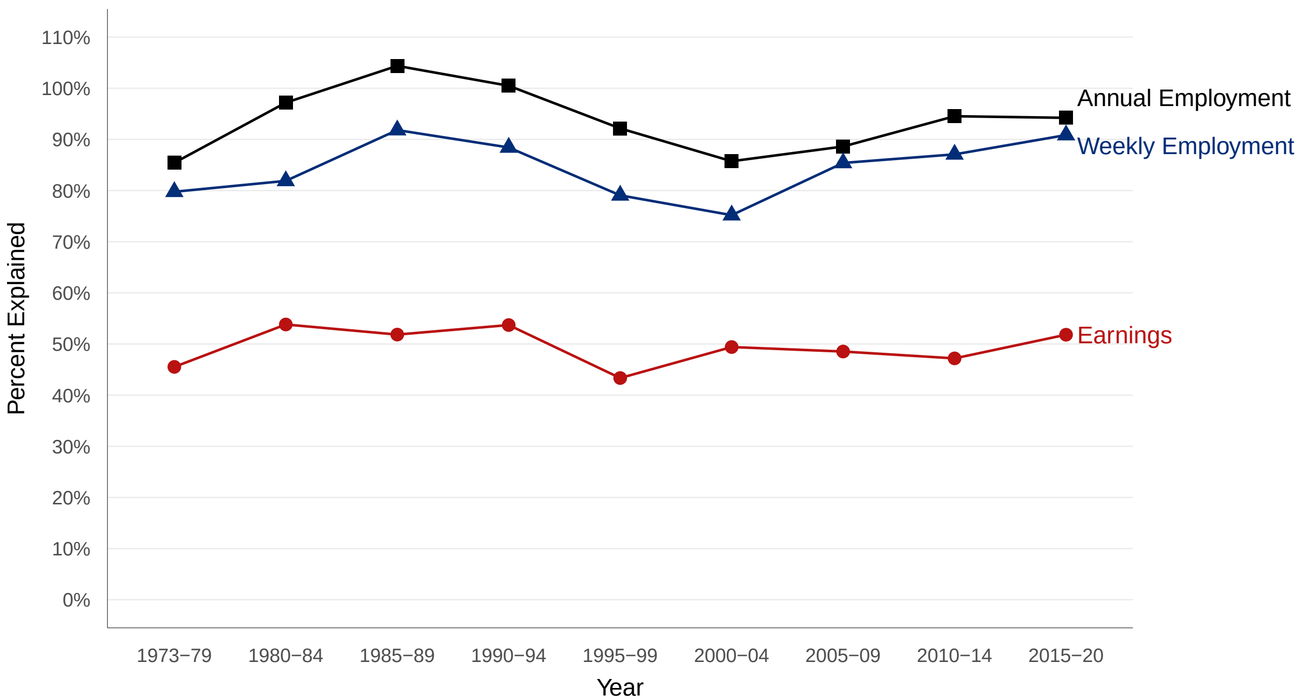


B. Counterfactual Outcomes



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.

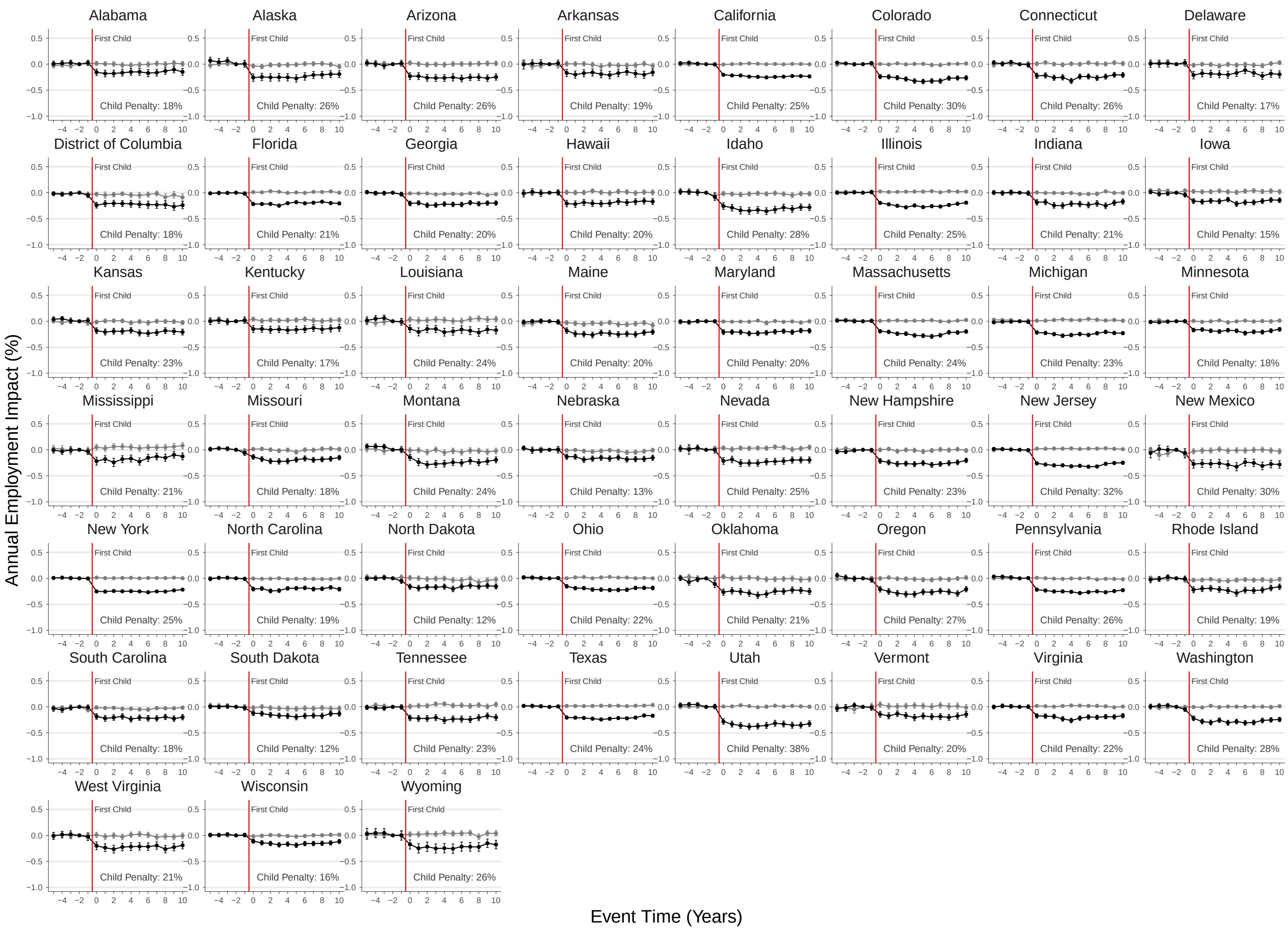
FIGURE A.10: FRACTION OF RAW GENDER GAP EXPLAINED BY CHILD PENALTIES



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 CHILD BIRTH ACROSS STATES

ANNUAL EMPLOYMENT



Notes: This figure shows event studies of first child birth 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 (7) with state dummies, estimating percentage impacts of child birth 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 (7) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

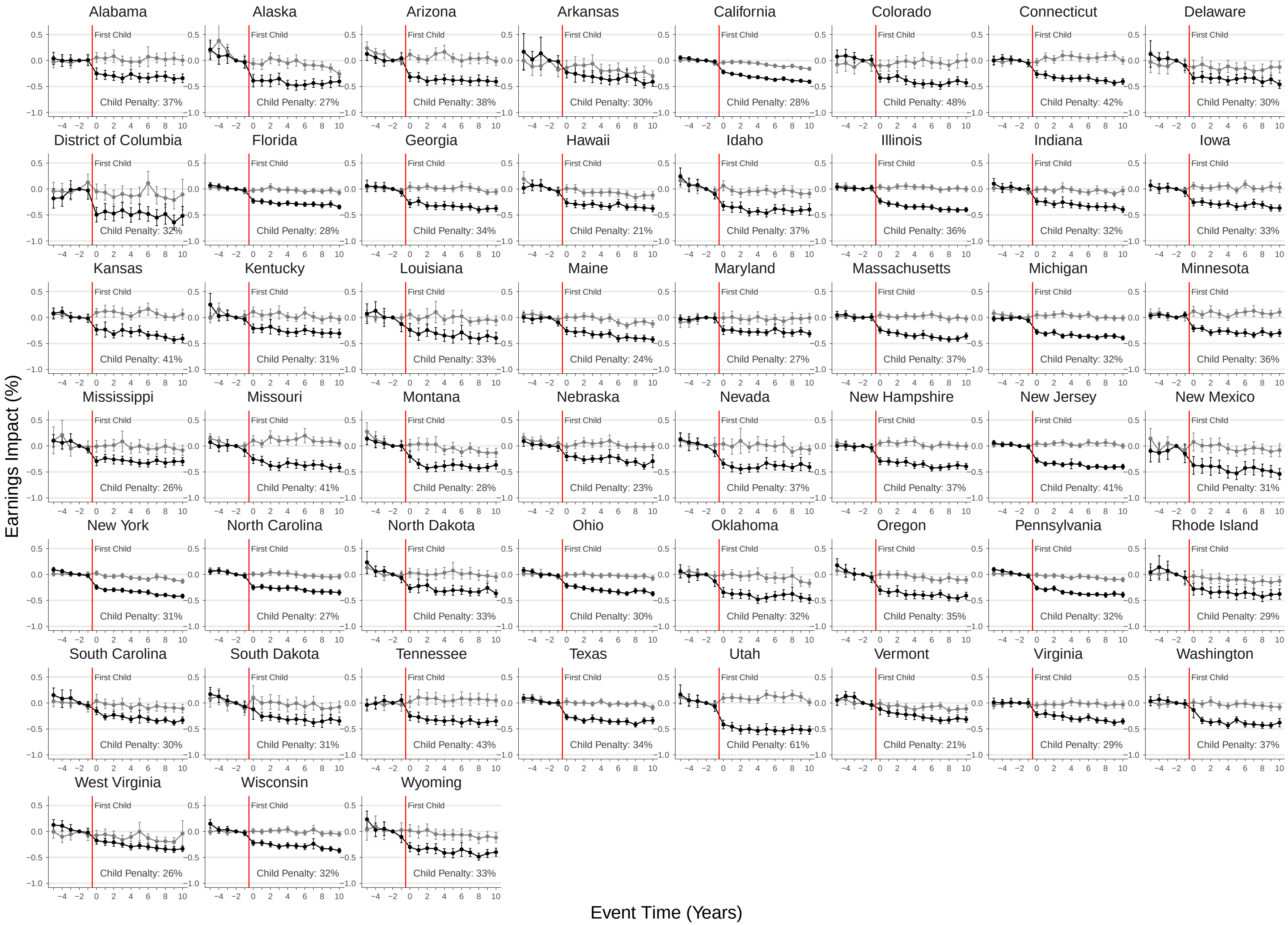
FIGURE A.12: EVENT STUDIES OF FIRST CHILD BIRTH ACROSS STATES

WEEKLY EMPLOYMENT

Notes: This figure shows event studies of first child birth 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 (7) with state dummies, estimating percentage impacts of child birth on men and women at each event time ($\hat{P}_\tau^m$ and $\hat{P}_\tau^w$) 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 (7) 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 CHILD BIRTH ACROSS STATES

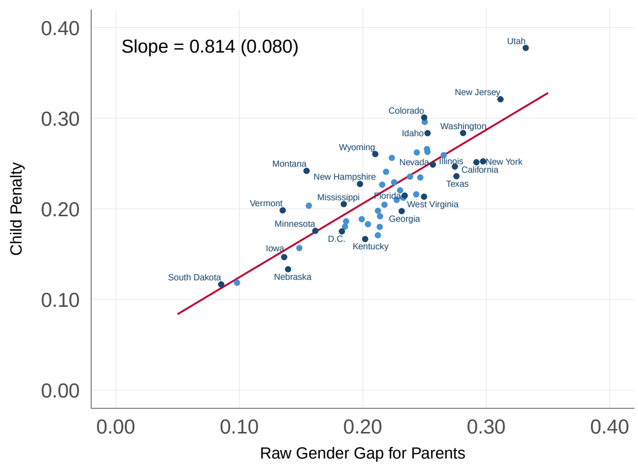
EARNINGS



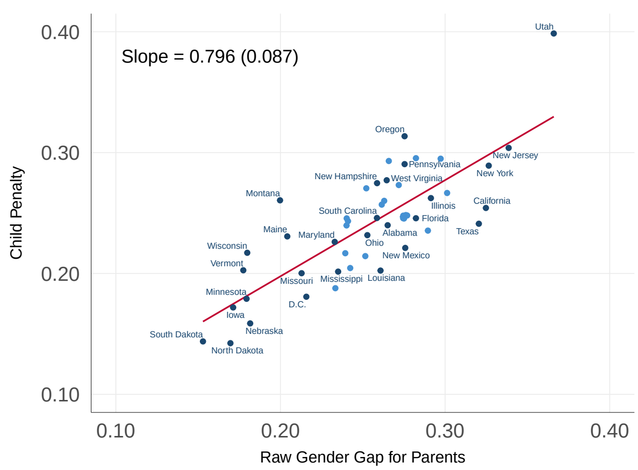
Notes: This figure shows event studies of first child birth in earnings for each of the 51 US states (including the federal district of D.C.). State-level event studies are constructed by interacting the event time dummies in equation (7) with state dummies, estimating percentage impacts of child birth 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 (7) 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

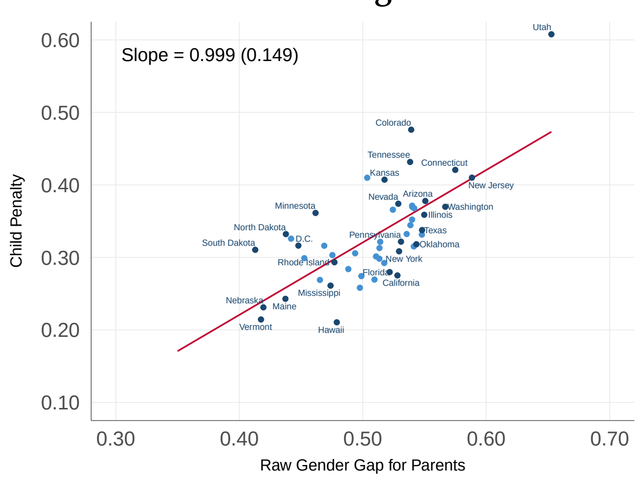
A. Annual Employment



B. Weekly Employment



C. Earnings

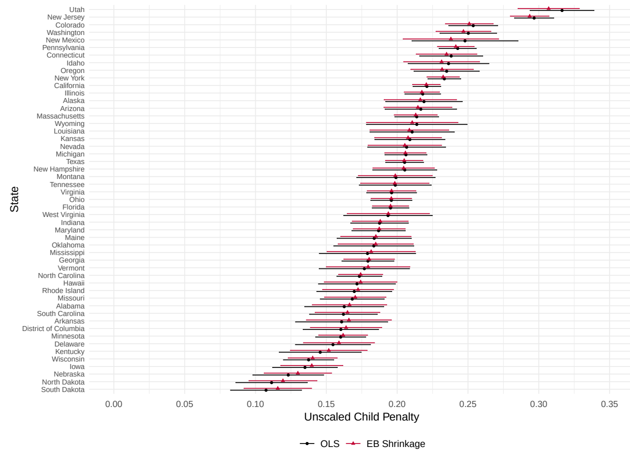


Notes: This figure provides scatter plots of child penalties against raw gender gaps for parents across states. Results are shown for each of the three labor market outcomes: annual employment, weekly employment, and earnings. The raw gender gap is defined as the percentage difference between men and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

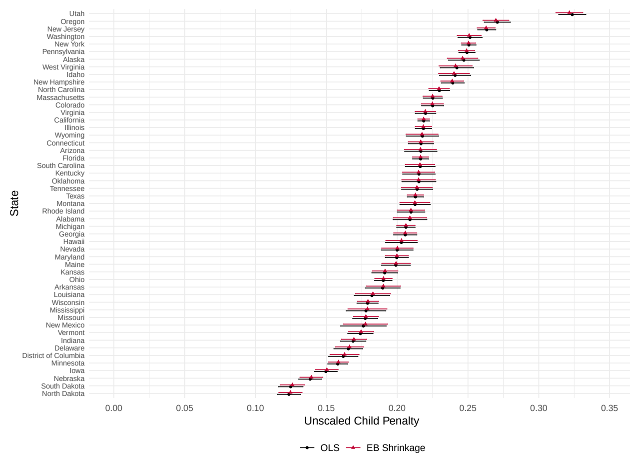
FIGURE A.15: EB VS OLS ESTIMATES OF CHILD PENALTIES

POINT ESTIMATES AND CONFIDENCE INTERVALS BY STATE

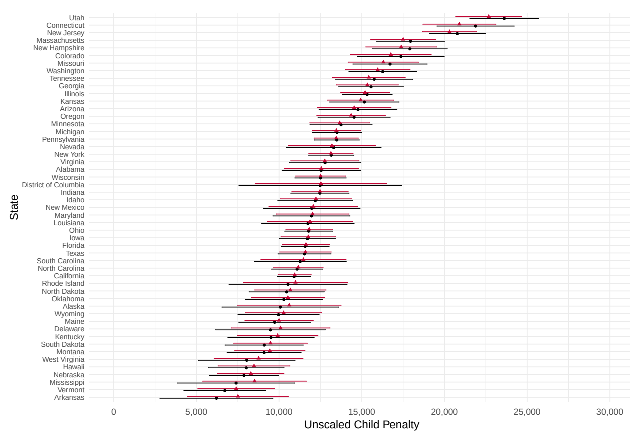
A. Annual Employment



B. Weekly Employment



C. Earnings

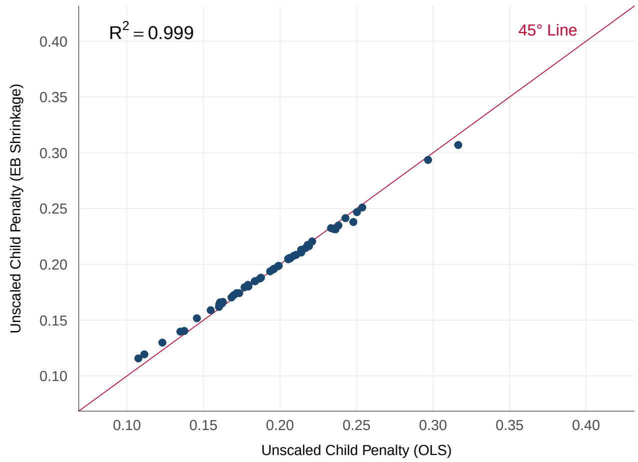


Notes: This figure compares OLS and Empirical Bayes (EB) estimates of unscaled child penalties for each state and each labor market outcome. The EB estimates are based on the linear-shrinkage formula in equation (14). The EB shrinkage adjustment hardly changes the estimates. The reason is the high statistical 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.

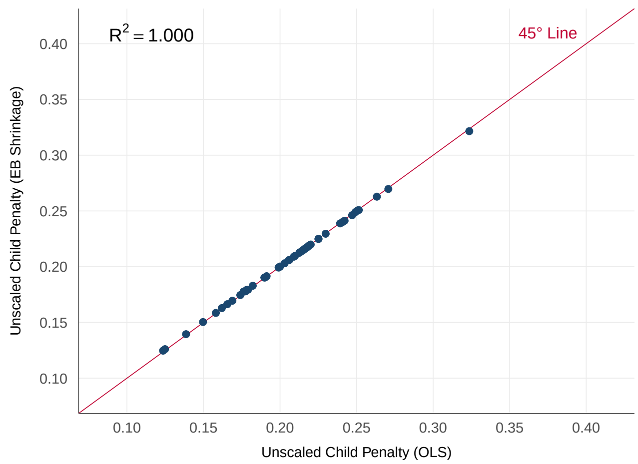
FIGURE A.16: EB VS OLS ESTIMATES OF CHILD PENALTIES

SCATTERPLOTS

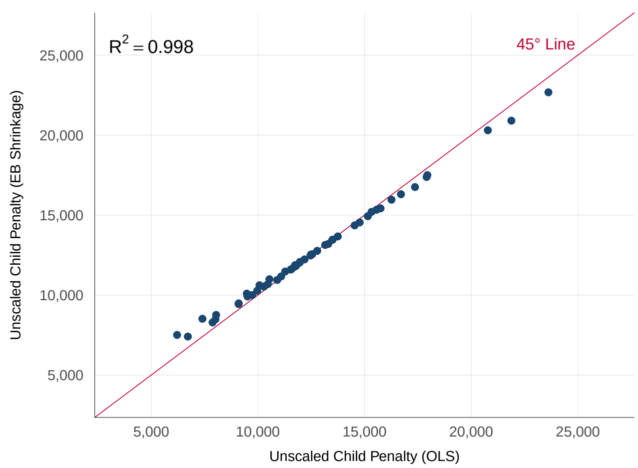
A. Annual Employment



B. Weekly Employment

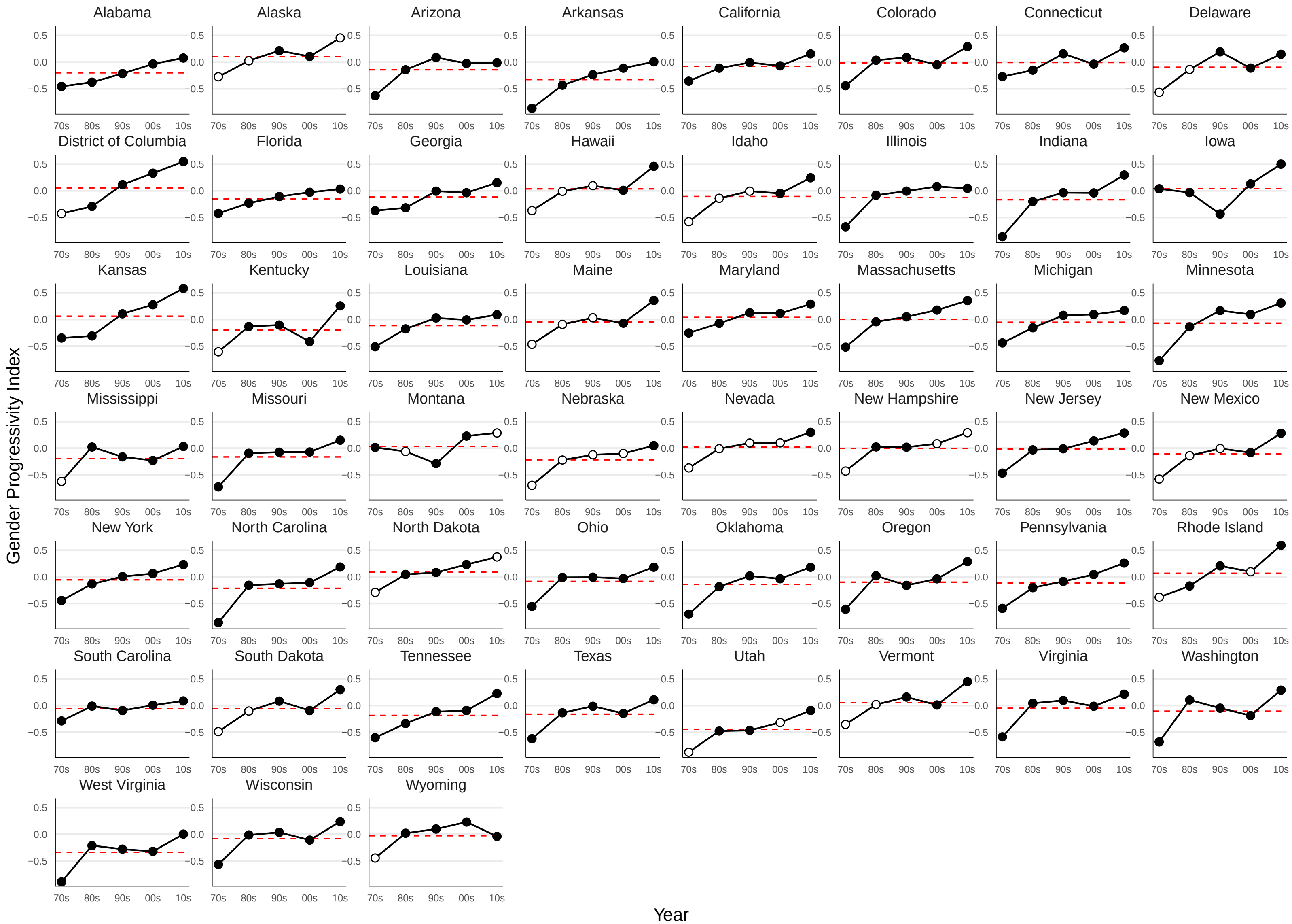


C. Earnings



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 (14). 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.

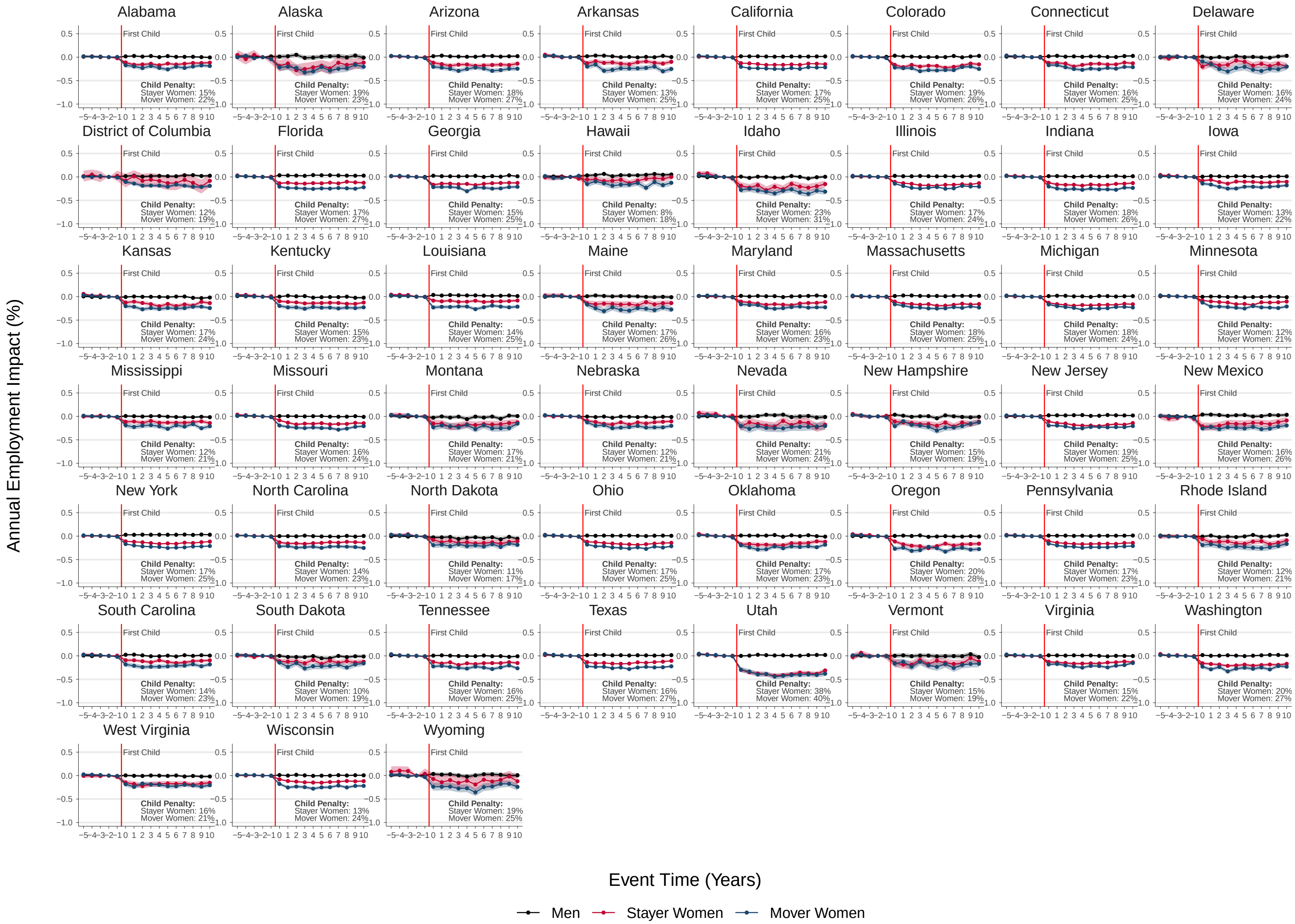
FIGURE A.17: GENDER PROGRESSIVITY INDEX BY STATE AND TIME



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 CHILD BIRTH FOR MOVERS VS STAYERS BY STATE OF BIRTH

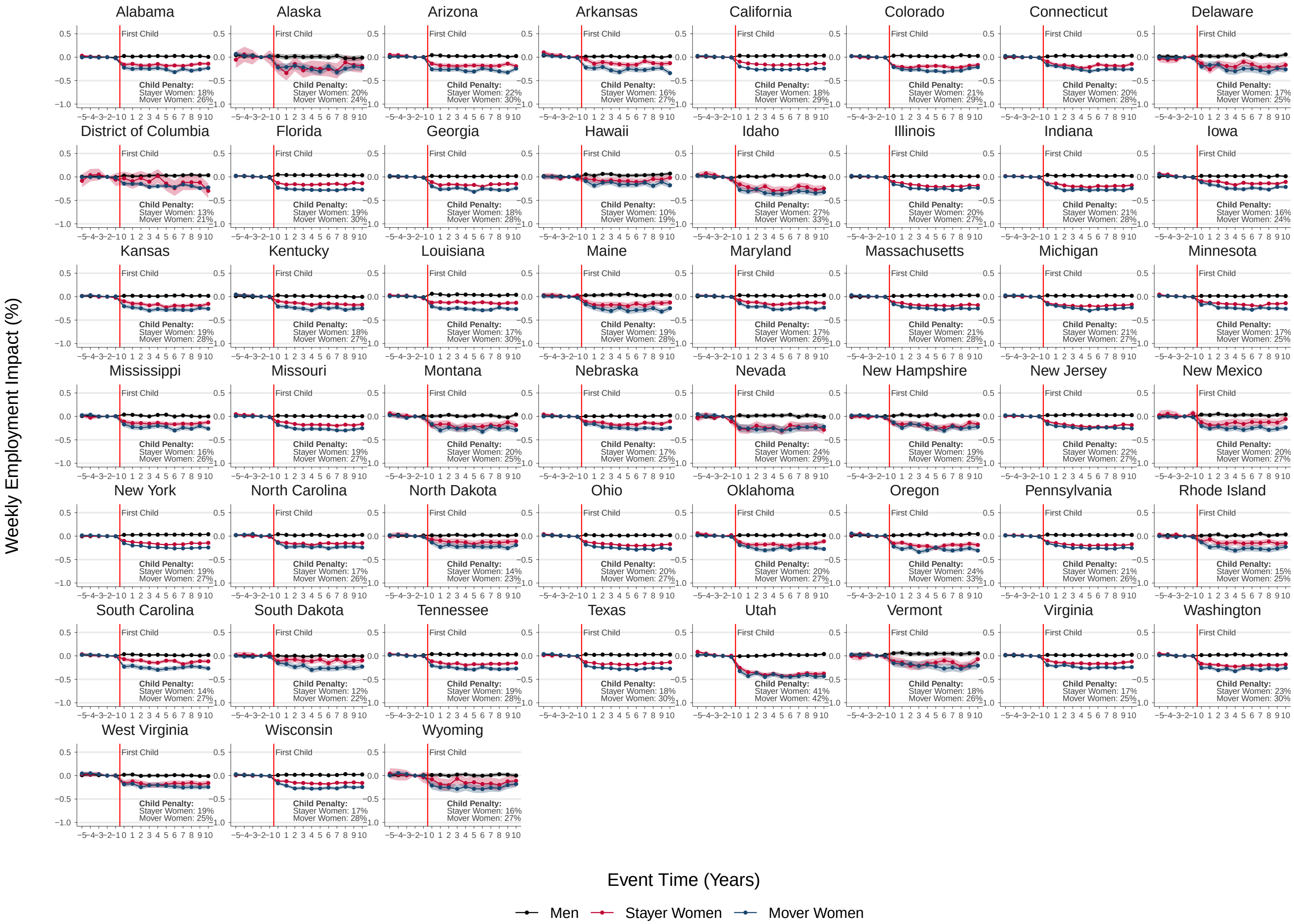
ANNUAL EMPLOYMENT



Notes: This figure presents event studies of first child birth 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 (7) 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 child birth 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 Child Birth for Movers vs Stayers by State of Birth

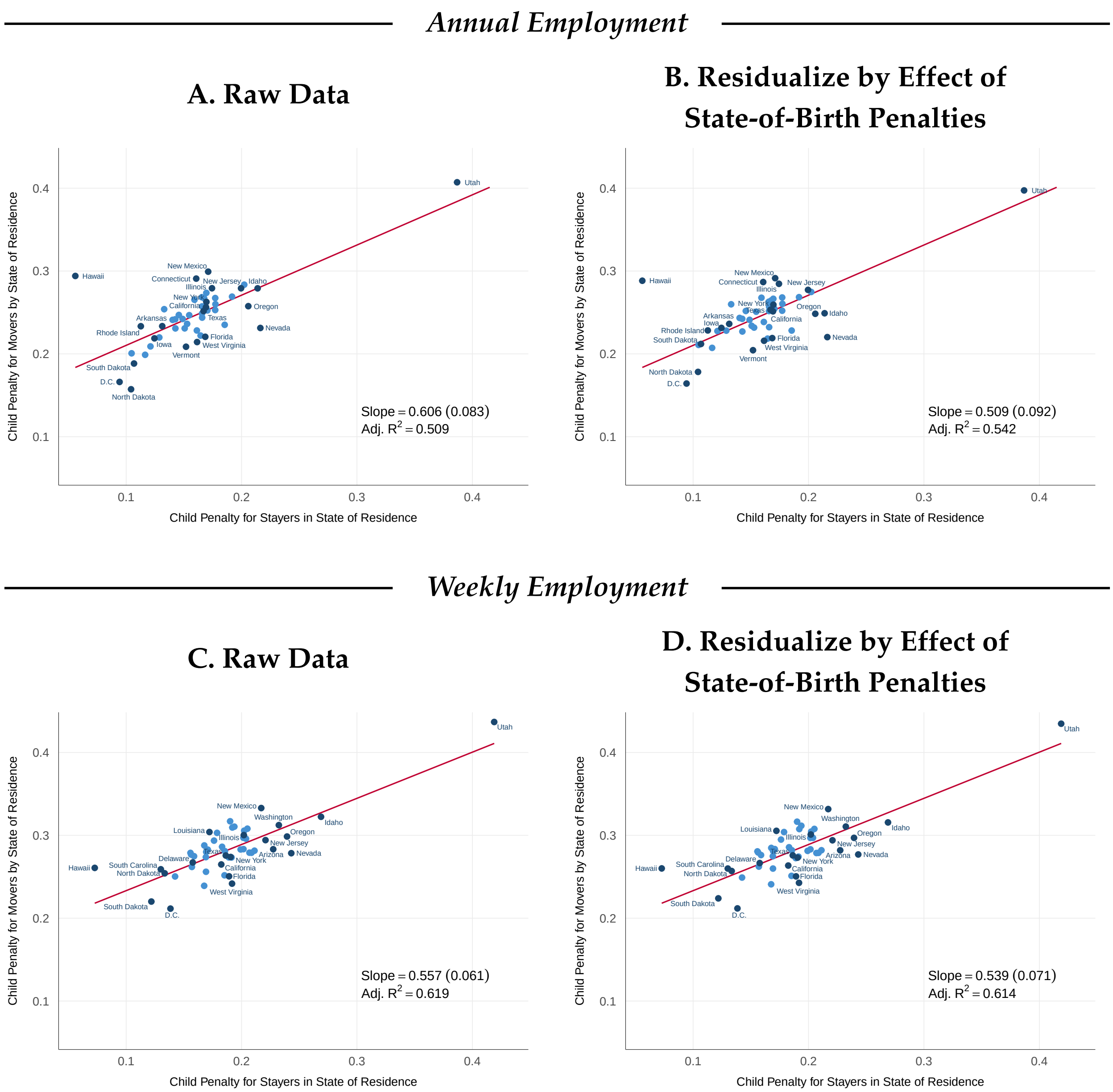
Weekly Employment



Notes: This figure presents event studies of first child birth 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 (7) 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 child birth 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



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. (16). 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.

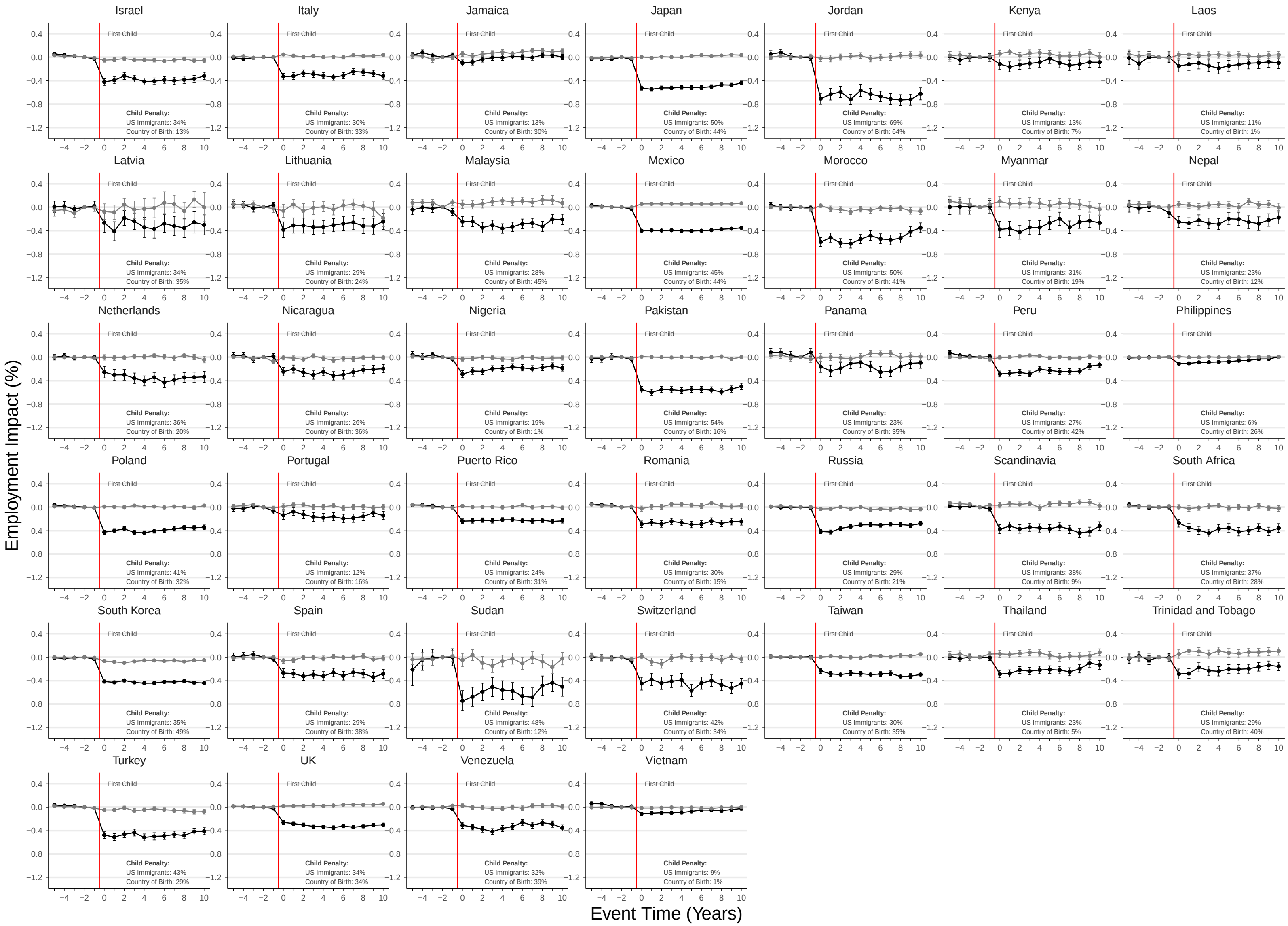
FIGURE A.21: EPIDEMIOLOGICAL STUDY OF FOREIGN IMMIGRANTS

EVENT STUDIES OF FIRST CHILD BIRTH FOR IMMIGRANTS BY COUNTRY OF BIRTH

Notes: This figure presents event studies of first child birth 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 CHILD BIRTH FOR IMMIGRANTS BY COUNTRY OF BIRTH

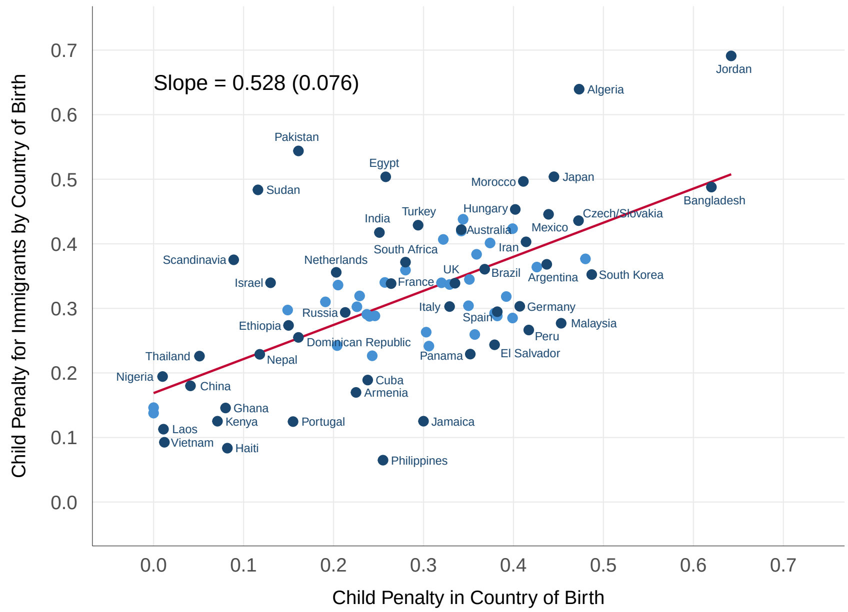


Notes: This figure presents event studies of first child birth for foreign-born immigrants by country of birth. Each panel displays the child penalty for US immigrants (based on the series shown) and the child penalty in their country of birth (based on Kleven, Landais, and Leite-Mariante 2024). The 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.

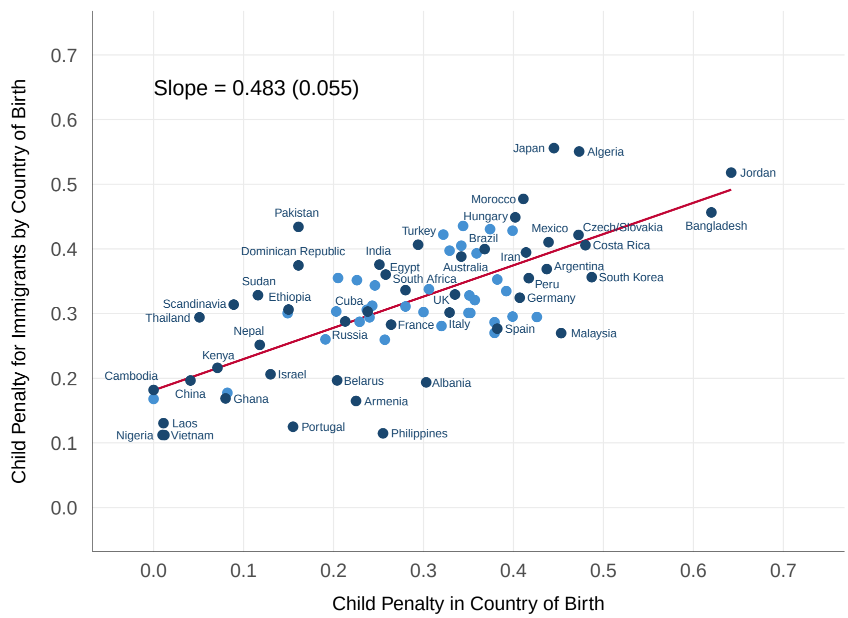
Figure A.22: Epidemiological Study of Foreign Immigrants

Child Penalties for Immigrants vs Child Penalties in Countries of Birth

A. Raw Data



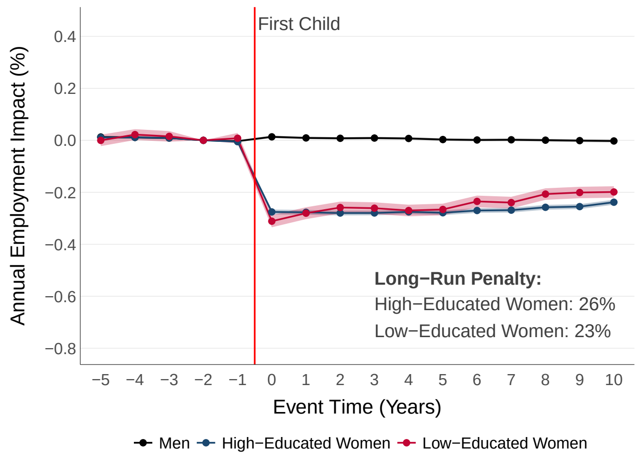
B. With Controls



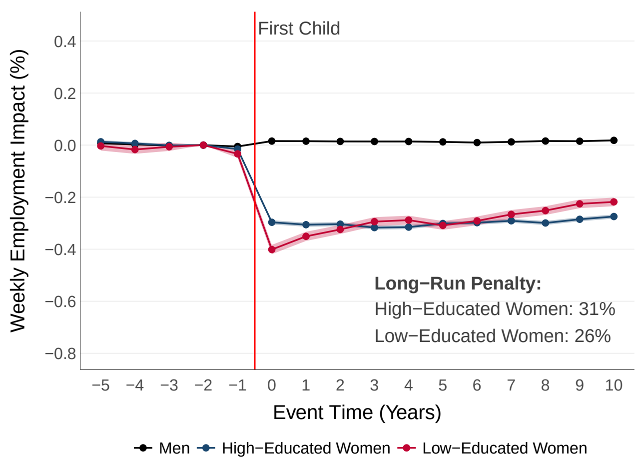
Notes: This figure presents scatter plots of child penalties for foreign-born immigrants against child penalties in country of birth. The underlying event studies for US immigrants are shown in Appendix Figure A.21 and the child penalties in country of birth are taken from Kleven, Landais, and Leite-Mariante (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 CHILD BIRTH BY EDUCATION
FOREIGN IMMIGRANTS

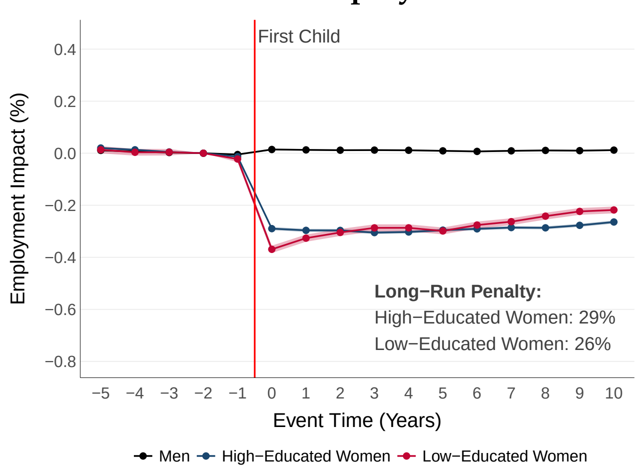
A. Annual Employment



B. Weekly Employment



C. Pooled Employment

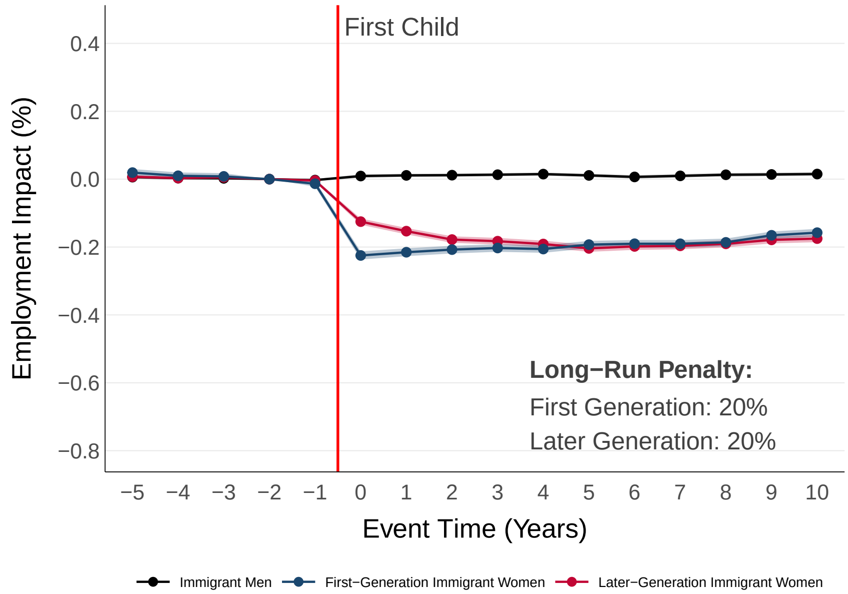


Notes: This figure presents event studies of first child birth by female education level for foreign-born immigrants. The figure is constructed in the same way as the education part of Figure 8 for the full sample. Results are shown for three labor market outcomes: annual employment, weekly employment,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

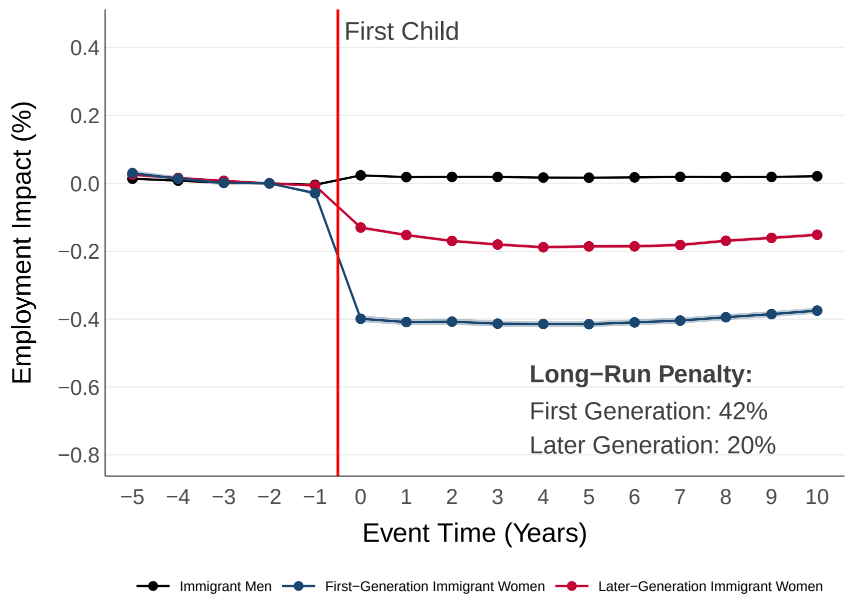
Figure A.24: Cultural Assimilation of Immigrants

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 child birth 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 (7) 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.