The Geography of Child Penalties and Gender Norms: Evidence from the United States*

Henrik Kleven
Princeton University and NBER

July 2022

*I thank Ilyana Kuziemko, Camille Landais, Ale Marchetti-Bowick, Gabriel Leite Mariante, and Owen Zidar for comments and discussions. I also thank Ragini Jain, Madhavi Jha, and Paola Villa Paro for outstanding research assistance. Contact information for the author: Henrik Kleven, Department of Economics and the School of Public and International Affairs, Princeton University, web: www.henrikkleven.com, e-mail: kleven@princeton.edu.

Abstract

This paper develops a new approach to estimating child penalties 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 give more statistical power than typical panel datasets. Five main empirical findings are presented. First, US child penalties have declined significantly over the last five decades, but almost all of this decline occurred during the earlier part of the period. Child penalties have been virtually constant since the 1990s, explaining the slowdown of gender convergence during this period. Second, child penalties vary enormously over 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. Third, child penalties correlate strongly with measures of gender norms. The evolution of child penalties mirrors the evolution of gender progressivity over time, with a greater fall in child penalties in states where gender progressivity has increased more. Fourth, an epidemiological study of gender norms using US-born movers and foreign-born immigrants 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 Rhode Island), 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 Mexico or Iran) have much larger child penalties than immigrants born in low-penalty countries (such as China or Sweden). Evidence is presented to show that these effects are not driven by selection. Finally, immigrants assimilate to US culture over time: A comparison of child penalties among first-generation and later-generation immigrants shows that differences by country of origin eventually disappear.

1 Introduction

The recent literature on gender inequality highlights the importance of child penalties: causal 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 2020). 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 2021), 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 Current Population Survey (CPS) data from 1968-2020 and American Community Survey (ACS) data from 2000-2019. Building on Kleven, Landais, and Søgaard (2019), the objective is to provide event studies around the birth of the first child, indexed as event time $t = 0$. 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 matching to create a pseudo-panel: Each person observed at event time $t = 0$ is matched to a childless person n years younger n years before and with the same demographic characteristics to obtain surrogate observations for $t = -n$. Having created a pseudo-panel, 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) and the National Longitudinal Survey of Youth (NLSY). The two approaches yield very similar results, but the cross-sectional approach is much more precise due to superior sample size.

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 2020), 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. Down the road, it allows for the possibility of mapping child penalties across most regions and countries in the world.[1]

Five 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% 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 finding sheds 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). The literature has discussed a variety of explanations for this puzzle, but conclusive evidence has been elusive. The evidence presented here provides a simple explanation: Gender convergence stalled because the decline in child penalties stalled.

Second, child penalties vary enormously over space. 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 align closely with the range of child penalties between Scandinavian countries and the culturally conservative countries of central Europe (Kleven 2021b). 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

[1] In ongoing work, Kleven, Landais, and Mariante (2022) use the approach to build a global database of child penalties.

a farm in North Dakota 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, being careful to address the confounding effects of urbanization and other factors.[2]

Third, the relationship between child penalties and gender norms is analyzed using General Social Survey (GSS) data from 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 of gender progressivity mirrors the time series of 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. To deal with confounders such as urbanization, state fixed effects and demographic composition are controlled for. Even with such controls, the effects are strong: 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.

Fourth, the paper develops an epidemiological study of gender norms using US-born movers and foreign-born immigrants.[3] 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 larger child penalties than those born in low-penalty states (such as the Dakotas or Rhode Island), conditional on where they reside. 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 6pp. Similarly, the child penalty for foreign immigrants is strongly related to

[2]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).

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

the child penalty in their country of birth. Immigrants born in high-penalty countries (such as Mexico or Iran) have larger child penalties than immigrants born in low-penalty countries (such as China or Sweden).[4] This effect is as large as for US movers: a 10pp increase in the employment penalty in a woman's country of birth translates into an increase in her employment penalty in the US of about 6pp. These results are consistent with important effects of cultural norms on child penalties. It is shown that these effects are unlikely to be driven by differential selection of movers and migrants from different places.[5]

Finally, evidence on cultural assimilation is provided by comparing child penalties among first-generation immigrants and later-generation immigrants. Immigrants assimilate to US culture over time: The stark differences in first-generation child penalties by country of origin are almost non-existent in later-generation penalties. The assimilation could take many generations to materialize, however, as later-generation immigrants include all descendants with a known country of ancestry regardless of the time at which their ancestors arrived.

This paper contributes to the literature on gender inequality, reviewed by Altonji and Blank (1999) and Bertrand (2011). It relates most closely to recent work on the impact of child birth and parenthood (Bertrand, Goldin, and Katz 2010; Kleven, Landais, and Søgaard 2019; Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019; Cortés and Pan 2020), and to work on the influence of social norms and culture (Fernández, Fogli, and Olivetti 2004; Fernández and Fogli 2009; Bertrand 2020; Boelmann, Raute, and Schönberg 2021). Two key findings motivate the research question posed here: The fact that child penalties account for most of the remaining gender inequality in developed countries, and the fact that these penalties cannot be explained by traditional mechanisms rooted in biology, comparative advantage, or government policies (Kleven, Landais, and Søgaard 2021; Kleven, Landais, Posch, Steinhauer, and Zweimüller 2021). In light of these findings, this paper asks if child penalties are better understood through the lens of gender norms and culture. To make progress on this difficult question, the paper develops a simple cross-sectional approach to estimating child penalties that allows for granular analyses of child penalties across time, space, and cultural groups. There is large variation in child penalties across

[4] Child penalties across countries are estimated by Kleven, Landais, and Mariante (2022).

[5] Related to the movers design presented here, Boelmann, Raute, and Schönberg (2021) explore the impact of culture on child penalties using movers between East and West Germany following the reunification of the two countries. The two parts of the country diverge significantly in terms gender norms (due to their different political institutions during separation), facilitating a study of movers who grew up in one child-penalty culture and moved to a different child-penalty culture. Compared to the study of movers in Germany, the epidemiological study presented here is more granular and exploits richer variation arising from the large heterogeneity in child penalties across different US states and different countries.

these dimensions, facilitating an epidemiological study of US movers and foreign immigrants. The analysis provides some of the first direct evidence that gender norms and culture have important effects on child penalties. The applicability of the approach is wide-ranging due to the minimal data requirements.

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 documents the variation in gender norms across time and geography, and investigates their effect 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 (CPS) between 1968-2020 and the American Community Survey (ACS) between 2000-2019. The CPS component includes data from both the basic monthly files and the Annual Social and Economic Supplement (ASEC), or "March files".[6] The pooled dataset includes about 44 million households over the entire period, which gives sufficient statistical power for granular event studies and investigating mechanisms.

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.

[6] 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 (2021a) for details.

The pseudo-event study specifications are validated against actual event study specifications using pooled data from the Panel Study of Income Dynamics (PSID) between 1968-2019 and the National Longitudinal Survey of Youth (NLSY) between 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 event studies at the level of the entire country — used to validate the pseudo-event study approach — but is under-powered for more granular analyses. The panel data contains similar labor labor market outcomes as the cross-sectional data, allowing for a validation of the approach for each outcome.

The analysis of social norms uses data from the General Social Survey (GSS) between 1972-2018. It includes about 65,000 households over the entire period. The GSS elicits information on a wide range of attitudes and beliefs among American residents, including beliefs about gender roles in families with children. The analysis presented below creates a measure of gender progressivity using a subset of questions that are particularly relevant for child penalties and consistently available over time. Based on this measure, the association between child penalties and gender progressivity over time and space is studied.

3 Methods

3.1 Event Study Approach

The event study approach to estimating child penalties uses panel data on men and women who become parents. The estimation is based on sharp changes in the outcomes of women relative to men around the birth of the first child, indexed to occur at event time $t = 0$. As proposed by Kleven, Landais, and Søgaard (2019), the following specification is run separately for men and women:

$$Y_{it}^{g} = \boldsymbol{\alpha^{g}} \cdot \boldsymbol{D}_{it}^{Event} + \boldsymbol{\beta^{g}} \cdot \boldsymbol{D}_{it}^{Age} + \boldsymbol{\gamma^{g}} \cdot \boldsymbol{D}_{it}^{Year} + \nu_{it}^{g}, \tag{1}$$

where Y_{it}^{g} is the outcome for individual i of gender $g = w, m$ at event time t. On the right-hand side, boldface is used to denote vectors. The first term includes dummies for each event time t, omitting a base year before child birth. The event time coefficients $\alpha_{t}^{g} \in \boldsymbol{\alpha^{g}}$ measures the impact of child birth on gender g in event year t, relative to the base year.[7] The second and third terms include a full set of age and year dummies to control non-parametrically for lifecycle trends and

[7]Throughout the paper, the omitted event time dummy is chosen as $t = -2$, the year before pregnancy. The choice of base year hardly impacts the results as there is virtually no pre-trend in the data.

time trends. The conditions for causal identification in this framework were laid out and validated against IV-approaches in Kleven, Landais, and Søgaard (2019).

The focus is on labor market outcomes such as earnings and employment. Equation (1) 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

$$P_t^g \equiv \frac{\hat{\alpha}_t^g}{\mathsf{E}\left[\tilde{Y}_{it}^g \mid t\right]}, \tag{2}$$

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. 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.

$$\text{Child Penalty} \equiv \mathsf{E}\left[P_t^m - P_t^w \mid t \geq 0\right] - \mathsf{E}\left[P_t^m - P_t^w \mid t < 0\right]. \tag{3}$$

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 contributes causally to increasing the gender gap in a given outcome.

3.2 Pseudo-Event Study Approach

The event study approach described above is straightforward to implement given access to high-quality panel data. Such data are not always available, however, which explains why compelling estimates of child penalties exist for just a handful of (developed) countries and typically not at a very granular level. This limits our understanding of how child penalties vary across geography and makes it difficult to study mechanisms. Motivated by these limitations, this section develops a pseudo-event study approach based on cross-sectional data. The idea is to use matching techniques to convert cross-sectional data into a pseudo-panel of men and women at different event times, thus allowing for the implementation of the event study specification in (1). The approach is implemented using cross-sectional data from CPS and ACS, and validated against panel data from PSID and NLSY.

Before describing the approach, it is useful to discuss the main identification challenge when estimating the impact of children, namely selection into parenthood. Table 1 provides descriptive statistics for men and women observed with and without children in cross-sectional data. To understand the nature of the selection problem, it is particularly informative to consider the outcomes of men. As can be seen from the table, men with children have better labor market and demographic outcomes than men without children. For example, their employment rate 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 virtually identical in the cross-section, despite the fact that child penalties pull mothers down, all else equal. The early literature addressed selection by controlling for observables, but this is not a credible solution due to the possibility of selection on unobservables.[8]

The first step of the approach is to create a pseudo-panel of men and women before and after the birth of their first child. For individuals with children, we observe the age of their oldest child and therefore know their place in positive event time, $t \geq 0$. 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, $t < 0$. We create surrogate observations of negative event times through matching. Specifically, consider parent i observed at event time 0 in calendar year y with age a and demographic characteristics $\boldsymbol{X_i}$. This parent is matched to a childless individual j observed in year $y - n$ with age $a - n$ and the same demographic characteristics $\boldsymbol{X_j} = \boldsymbol{X_i}$. This gives a surrogate observation for $t = -n$.[9] By matching each parent at event time 0 to childless individuals for $n = 1, ..., 5$, a pseudo-panel with 5 years of pre-child data is created.[10] This procedure implies

[8] 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 driven by selection. The modern literature has addressed selection by using instruments for fertility such as sibling sex mix (Angrist and Evans 1998) and IVF treatment success (Lundborg, Plug, and Rasmussen 2017). These approaches are compelling as far as they go, but they are limited in terms of external validity (as they study particular samples and treatment effects) and statistical power. The questions posed in this paper cannot be feasibly analyzed using available IV-approaches. Kleven, Landais, and Søgaard (2019) discuss these approaches in detail and use them to validate the event study study approach laid out in the preceding section.

[9] 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.

[10] To clarify, there is a slight difference in the matching protocol for different labor market outcomes. The protocol described above 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 y with age a) and non-parents (observed in year $y - n$ with age $a - n$) is done for $n = 0, ..., 4$ to obtain

that calendar time and age are changing along the event time dimension, exactly as in the panel data approach of Kleven, Landais, and Søgaard (2019). The effect of time and lifecycle trends are absorbed through the inclusion of age and year dummies in equation (1).

To implement the approach, the set of demographic variables used for matching needs to be specified. Importantly, the choice of matching 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 of the data is the effect of the first child on men. As shown in previous papers, child birth is a non-event for men. Therefore, if the pseudo-event study is associated with a positive jump in the labor market outcomes of men around $t = 0$, this reflects bias from positive selection. The set of matching variables used below are chosen to avoid such selection bias. The variables used are gender, education (4 categories), marital status (5 categories), race (4 categories), and state of residence (51 states, including the federal district of D.C.).[11]

Table 2 provides descriptive statistics for matched men and women at event times $t = 0$ and $t = -1$ in the pseudo-panel.[12] By construction, these samples match exactly on education, marital status, race, age at first birth, and cohort. Also by construction, individuals at event time $t = 0$ are exactly one year older than those at event time $t = -1$. The samples do not match on labor market outcomes, nor are they supposed to: Those observed at $t = 0$ 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.

3.3 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

surrogate observations for $t = -n - 1$. For the same reason, annual outcomes at event times $t = 0, ..., T$ are obtained from parents observed at event times $t = 1, ..., T + 1$.

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

[12]The sample restricts attention to parents whose age at first birth lies between 25 and 45. Throughout the analysis, births before age 25 are excluded to have information on pre-birth labor market trends during adulthood. The event study approach is less suited to studying the impact of early births (such as teenage births) due to the limited pre-birth labor market history in such cases.

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 $t = 0$, marked by the red vertical line. The event time horizon shown in these and subsequent graphs goes from $t = -5$ to $t = 10$. The average child penalty across event times 0-10, defined as in equation (3), is displayed in each panel. Three outcomes are shown: annual employment (top panels), weekly employment (middle panels), and earnings (bottom panels).

The results from the cross-sectional and panel approaches align closely. If anything, the cross-sectional approach is more compelling in terms of statistical precision and pre-trends. It features 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 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 coefficient estimates are less precise and the pre-trends are less compelling.

To assess 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 key insight is that the more parsimonious specifications introduce selection bias, evidenced by the positive jumps in the labor market outcomes of men between $t = -1$ and $t = 0$. 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.[13]

Panel vs Panel: The preceding validation exercise compares results from different datasets. It is possible that the pseudo-event studies align with the true event studies due to offsetting effects of the approach and 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

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

a validation, comparing pseudo-event studies (left panels) to actual event studies (right panels) using PSID and NLSY data for both. This validation also works exceedingly well. The pseudo-event studies look similar to the true event studies and produce similar-sized child penalties. In fact, 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. Overall, the validation exercises presented in Figures 1-2 lend strong support to the empirical approach taken in this paper.

Why does the pseudo-event study approach work so well? A sufficient but not necessary condition is that the approach accurately predicts fertility (specifically, location in negative event time) among those observed without children. Appendix Figure A.4 investigates this point, comparing predicted and actual event times among childless people in PSID and NLSY data. The figure shows the distribution of within-person differences between predicted event times (obtained from matching) and actual event times (directly observed).[14] 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 are sufficiently modest to not destroy the accuracy of the pseudo-event studies given the chosen matching specification.

Having validated the approach, the rest of the paper provides 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.

4 Child Penalties in the United States

4.1 Child Penalties Over Time

Figure 3 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 (1) is run for different time periods separately.[15] The event studies for each time period and labor market outcome are

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

[15]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

presented in Figures A.5-A.7 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. Importantly, almost all of the decline in child penalties occurred prior to the mid-1990s, followed by a long period of stagnation. This finding sheds 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 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 provides a simple explanation: Gender convergence stalled because the decline in child penalties stalled.

How much of gender convergence can be attributed to changes in child penalties? For each labor market outcome, Appendix Figure A.8 shows the fraction of the gender gap for parents explained by child penalties over time. 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.[16] This implies that the evolution of gender inequality in the labor market is explained largely by child penalties.

This explanation is admittedly very reduced-form. It shifts the research question one level up: if child penalties determine the rate of gender convergence, we would like to know what determines child penalties. This paper is motivated by precisely this question. A number of empirical analyses will be presented to highlight the critical role of social norms and culture.

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.

[16]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 that men.

4.2 Child Penalties Across Space

To study the variation in child penalties across states, the event time dummies in equation (1) are interacted with state dummies.[17] In this specification, the lifecycle and time trends are estimated at the level of census divisions by interacting the age and year dummies with census division dummies. Estimating lifecycle and time trends at the state level produces similar results, but the event studies for some of the smaller states (specifically for the earnings outcome) become noisier under such a granular specification.

As a first glimpse of the spatial variation, Figure 4 presents event studies for five selected states across the country: California (Pacific), New Jersey (Northeast), North Dakota (Northern Midwest), Texas (South), and Utah (Mountain West). Results are shown for annual employment, weekly employment, and earnings. The impact of child birth on labor market outcomes is sharp and persistent in all five states, but varies greatly in magnitude. Child penalties are relatively small in a state like North Dakota (close to Scandinavian levels) and extremely large in a state like Utah (close to central European levels).[18] For example, the annual employment penalty equals 12% in North Dakota and 38% in Utah. The penalties are also very large in New Jersey, one of the most urban states in the country, and at intermediate levels in California and Texas.

Figures A.9-A.11 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 5. 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. How do these child penalties relate to raw gender gaps? This is shown in Appendix Figure A.12, which provides scatterplots 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 outcomes. In other words, variation in gender inequality across

[17] 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."

[18] Child penalties across European countries have been documented in Kleven, Landais, Posch, Steinhauer, and Zweimüller (2019) and Kleven (2021b).

space is a reflection of variation in child penalties.

Figure 6 provides another way of illustrating the spatial variation in the data. The left panels plot estimated employment penalties against observed employment rates for mothers across states. There is a strong negative relationship between the two, with a slope of -1.00 in annual employment and a slope of -0.75 in weekly employment. In other words, a higher maternal employment rate maps almost one-for-one into a lower child penalty. The right panels plot counterfactual employment rates (absent the impact of children) against observed employment rates for mothers across states. Because of the negative relationship between child penalties and observed employment rates, there is relatively little variation in counterfactual employment rates across states. This is particularly true of weekly employment: Absent children, the maternal employment rate would be between 85% and 90% in all states. These graphs also serve as a basic validation check of the empirical approach, namely that the implied counterfactual employment rates are always below 100%. The graphs confirm that this is satisfied.

What drives 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).[19] 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,

[19] 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).

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.

4.3 Child Penalties Across Demographic Groups

This section presents evidence on heterogeneity in child penalties by demographic characteristics. As we shall see, there is virtually no heterogeneity by female education, but lots of heterogeneity by marital status and race. These findings are important for thinking about mechanisms and for addressing possible confounders in the next section on the role of gender norms.

Education: Figure 7 presents event studies by female education level. The figure is constructed by running specification (1) separately for low-educated women (high school degree or less) and high-educated women (college degree or more). Because child birth is a non-event for men regardless of their education, the sample of men is not split by education. The figure shows results for all three labor market outcomes and reports the long-run child penalty (over event times 5-10) for each outcome.

The short-run impact of parenthood is 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, although the difference is not statistically significant.[20] These findings contradict theories based on specialization by comparative advantage in market work vs child care. If comparative advantage were important, we would expect low-educated women to have larger child penalties than high-educated women.

A few remarks on interpretation are worth making. First, because female education has increased over time, the low-educated sample tends to be selected from earlier years than the high-educated sample. This serves to *strengthen* the absence of comparative advantage effects. 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-

[20]The penalties for both education groups in Figure 7 are smaller than the penalties for the full sample in Figure 1. The main reason is that here we consider *long-run* penalties (t = 5-10) rather than penalties over the full event time window (t = 0-10).

educated women. Second, splitting women by education without conditioning on their spouses' education implies that the results could be influenced by sorting in the marriage market. Positive sorting on education reduces the degree to which within-family comparative advantage varies by female education group. However, because child birth is a non-event for men regardless of their education, the results are not materially affected by sorting. Finally, the findings presented here are consistent with findings from Denmark using richer data to measure comparative advantage within families. Kleven, Landais, and Søgaard (2021) estimate male and female earnings capacity using data on education level, education field, and labor market experience. They show that child penalties are unrelated to relative earnings capacity within families: women with higher earnings capacity (relative to their spouses) face similar child penalties as women with lower earnings capacity.

Marital Status: Figure 8 presents event studies by marital status. The figure is constructed in the same way as the preceding one: Specification (1) is estimated separately for single and married women, but jointly for men regardless of their marital status. The definition of being single 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 the three outcomes, but the magnitudes are particularly stark for the employment outcomes. For example, the child penalty in annual employment equals 27% for married women and only 5% for single women. These findings highlight that child penalties are closely linked to the possibility of specialization between spouses, even if this specialization is not governed by comparative advantage as shown above.

Why are child penalties on single women much smaller than on married women in the US? Interestingly, the US evidence is exactly opposite Danish evidence presented in Kleven (2021d). 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 (2021d) 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: Figure 9 presents event studies by race, comparing child penalties among 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.

Fertility: The event study approach is centered on the birth of the first child, but it does not condition on total fertility. As a result, the child penalty estimates include the impact of subsequent children over the event study horizon. Kleven, Landais, and Søgaard (2019) discuss identification in this setup and provides several validation checks. While estimates of short-run penalties rely only on the sharp discontinuity at event time zero, estimates of longer-run penalties rely on parallel trends between men and women conditional on the controls for lifecycle and time trends. With a full set of age and year dummies, this amounts to assuming parallel trends with respect to the timing of first birth. As we move to the right on the event time scale, we are considering men and women who had their first child at an earlier age.

With cross-sectional data, it is not possible to condition on completed fertility. Still, it is possible to investigate how child penalties vary by fertility by conditioning on the maximum number of children at any given event time. Such results are presented in Appendix Figure A.13: For each labor market outcome, event studies are presented for women with at most one child, women with at most two children, and women with any number of children (baseline specification). As one would expect, child penalties are strictly increasing in the number of children and converge to the baseline estimates that allow for any number of children. The heterogeneity by number of children is more pronounced for employment than for earnings.

5 The Effect of Gender Norms on Child Penalties

5.1 Child Penalties vs Gender Progressivity Over Time and Space

To investigate the effect of gender norms, we first consider the relationship between child penalties and a measure of gender progressivity obtained from GSS data between 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. The resulting time series of the GPI for each state are provided in Figure A.14 of the appendix.[21]

Figure 10 presents a heatmap of gender norms in the United States. States are divided into deciles of the GPI, with darker colors representing more 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

[21] In this figure, actual state-decade observations of the GPI are marked by filled dots and imputed observations are marked by empty dots. The figure shows that gender progressivity has increased over time in all states, but that there are substantial differences in the rate and timing of these increases. There are also substantial differences in the average level of gender progressivity, marked by the dashed red line.

the two is not strong. Figure A.15 in the appendix shows scatterplots of child penalties vs gender progressivity across states. The slope is negative in all outcomes — more gender progressivity is associated with smaller child penalties — but the relationship is relatively weak. This does not invalidate the effect of social norms, however, because the cross-sectional relationship is likely affected by confounders. As discussed above, one such confounder is labor market structure, including the temporal flexibility and family friendliness of jobs. Such aspects are likely related to the degree of urbanization. Rural states tend to be gender conservative (increasing the child penalty, all else equal), but offer jobs with greater flexibility for families (reducing the child penalty, all else equal). Other confounding factors may play a role too. To address these issues, the spatial variation in gender norms will be combined with time variation.

Figure 11 compares the time series of gender progressivity to the time series of child penalties in employment (Panel A) and earnings (Panel B). The evolution of gender progressivity is a mirror image of the evolution of 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, combines 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.

Figure 12 brings together the state and time variation. The 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 specification:

$$\text{Child Penalty}_{st} = \beta \cdot GPI_{st} + \gamma_s + \boldsymbol{\delta} \cdot \boldsymbol{X}_{st} + \nu_{st}. \tag{4}$$

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.[22] The previous section also presented evidence on hetero-

[22] 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.

geneity by number of children, but fertility is not included among the controls as it may be a transmission mechanism for gender norms.

Having estimated equation (4), 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.[23] Binscatters for all three labor market outcomes are presented in Figure 12. The left panels include only state fixed effects, while the right panels include both state fixed effects and 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 may have important effects on child penalties. The US evidence is consistent with the cross-country evidence on child penalties and elicited gender norms in Kleven, Landais, Posch, Steinhauer, and Zweimüller (2019), but the within-country analysis presented here is geographically more granular and exploits time variation as well. This makes causal identification more plausible, but not conclusive. The variation in elicited gender norms is potentially endogenous, posing a threat to interpretation. The choice of demographic controls to address possible confounders 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.[24] 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

[23] When plotting residualized child penalties by bin of the GPI, the average effect of the controls, i.e. $\mathsf{E}\left[\hat{\gamma}_s + \hat{\boldsymbol{\delta}} \cdot \boldsymbol{X}_{st}\right]$, 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.

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

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. The next section considers immigrants.

As a first visualization of the results, Figure 13 presents case studies for 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 (1) 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. The child penalties for movers and stayers reported in the figure are thus based on comparing coefficients for women movers and stayers, respectively, to coefficients for all men. 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 regarding gender roles in families with children.

Event studies for movers and stayers are presented for all states in Figures A.16-A.17 of the appendix. The results from these event studies are summarized in Figure 14, which provides scatterplots of the child penalty for movers against the child penalty for stayers by state of birth. Annual employment is shown in Panel A and weekly employment is shown in Panel B. The relationship between mover penalties 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 Rhode Island). 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 7.1pp, although she lives and has children somewhere else. The effect of the birth-state penalty in weekly employment is slightly weaker, but still very strong.[25]

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 may be more likely to move to another high-penalty state, whereas people born in low-penalty states may be more likely to move to another low-penalty state. If moves are selected in this way, the estimated effect of childhood environment may biased by effects of adulthood environment. Such place effects could reflect another dimension of cultural transmission — an effect of the culture experienced as adult — but may also reflect entirely different mechanisms. This includes the effect of job flexibility in one's current labor market as discussed previously.

The issue of selection on state of residence is addressed in Figure 15. To construct this figure, the child penalties for movers by state of birth are regressed on child penalties for stayers along with controls for the fraction of movers residing in different deciles of stayer penalties. The controls absorb variation in residence choices across movers born in different states. Having run the regression, mover penalties are residualized using the estimated residence controls and plotted against stayer penalties by state of birth.[26] The results are presented in binscatters, dividing stayer penalties into ten deciles. Controlling for differences in where movers reside does not qualitatively alter the findings. For annual employment, the relationship between mover and stayer penalties has a slope coefficient of 0.55. This implies that, as the employment penalty in a woman's state of birth increases by 10pp, her employment penalty increases by 5.5%. For weekly employment, the effect is slightly stronger.

Of course, there could be selection in other dimensions than state of residence. 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 a causal effect of culture. To investigate the importance 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. Strikingly, the table shows that the only relevant dimension of selection is where people move: Those born in high-penalty states are more likely to move to other high-penalty states. This is precisely the dimension of selection addressed above. For all other variables, there is very little selection. Movers born in low-penalty and high-penalty states are strikingly similar in terms of education levels, marriage rates, racial composition, fertility rates,

[25] The results are qualitatively similar for earnings penalties (not shown), although this outcome is more noisy.

[26] When plotting the residualized mover penalties, the average effect of the estimated controls is added to the residuals to make the levels in Figure 15 comparable to those in the preceding figure.

age at first birth, age, and cohort.

The absence of selection on observables provides strong support to the epidemiological approach presented in this section. The findings suggest that culture and gender norms have strong effects on child penalties.

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 identified 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. However, the outcome variable considered here is different and more challenging to study. The reason is that child penalties, unlike the observables typically studied, represent causal effects that have to be estimated quasi-experimentally.

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 Mariante (2022) estimate child penalties in employment for roughly 140 countries.[27] Child penalties exist in almost every country, but their magnitudes vary enormously. For example, employment penalties are small in countries such as China (0%) or Sweden (10%), but very large in countries such as Mexico (43%) or Jordan (47%). 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 of child penalties for as many source countries as possible, information on weekly employment (working last week) and annual employment (working 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 66 source countries where event studies of US immigrants are feasible *and* where Kleven, Landais, and Mariante (2022) provide estimates of source-country child penalties.

Figure 16 presents event studies for US immigrants from a selected set of countries.[28] These

[27] Preliminary findings from this paper are presented in Kleven (2021c).

[28] Figure A.18 in the appendix provides event studies for all 66 countries included in the analysis.

countries represent some of the largest sources of US immigration, and they span a wide range of geography and institutions. The top row shows Asian immigrants, the middle row shows South American immigrants, and the bottom row shows European immigrants. In each row, the event studies have been ordered by the size of the child penalty in country of birth. Each panel displays child penalties for US immigrants (obtained from the event study shown) and for people in their country of birth. The estimates imply a strong and monotonic relationship between immigrant penalties and birth-country penalties within each continent. For Asian immigrants, as we move from the lowest to the highest birth-country penalty (from China to Jordan), the child penalty increases from 18% to a staggering 84%. For South American immigrants, as we move from the lowest to the highest birth-country penalty (from Cuba to Mexico), the child penalty increases from 22% to 44%. The results for European immigrants are similar, with relatively modest child penalties among Swedish immigrants (a low-penalty country) and very large child penalties among Polish immigrants (a large-penalty country).

Figure 17 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 1%) and 55% in the top decile (where the average birth-country penalty is 48%).[29] Building on these results, Figure 18 shows results for the full distribution of birth-country penalties. It provides binscatters of immigrant penalties vs birth-country penalties by decile of the distribution of birth-country penalties.[30] 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 0.696 implies that, as the employment penalty in a woman's country of birth increases by 10pp, her employment penalty in the US increases by 7pp. Because women living in the US are not directly affected by the incentives and institutions of their birth countries, this evidence is most naturally interpreted as an effect of childhood culture on preferences.

As discussed above, a concern with epidemiological studies is the possibility of differential selection of movers/migrants from different places. For US movers, this was not an issue in practice: As shown in Panel A of Table 3, movers born in low-penalty and high-penalty states are virtually

[29] 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.

[30] Appendix Figure A.19 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 18.

identical on average, except in terms of their state of residence which was directly addressed in the empirical analysis. Panel B of Table 3 provides evidence on the selection of immigrants from different countries. This panel compares the demographic characteristics of mothers immigrating from countries in the bottom and top quartiles of child penalties. In this case, there *is* selection on some observables. Mothers from high-penalty countries have less education and a different racial composition than mothers from low-penalty countries. The rest of the observables are 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 alleviates selection concerns by showing that child penalties are unrelated to female education. Appendix Figure A.20 reproduces this finding for the sample of immigrants. On the other hand, section 4.3 also showed that there is heterogeneity in child penalties by race, although this may itself reflect cultural differences across racial groups.

To address potential selection effects, Panel B of Figure 18 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.[31] 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 has only a small effect. The slope coefficient is slightly smaller, but the relationship between immigrant and birth-country penalties is more stable and linear in this specification. 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 preceding analysis of elicited gender attitudes — suggest that gender norms and culture are crucial 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.

[31] These controls are specified as shown in Table 3.

5.4 Cultural Assimilation

How persistent are cultural norms? Do immigrants retain their ancestral culture over time or do they assimilate to their surrounding culture? Figure 19 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 (1) separately for first-generation and later-generation immigrants within each quartile. As in the preceding section, the outcome variable is pooled employment. The figure compares child penalties for first- and later-generation immigrants in the bottom quartile (Panel A) and in the top quartile (Panel B).

The results show that immigrants do assimilate, but asymmetrically so. Immigrants from high-penalty countries start out with large child penalties in the first generation, but have much smaller child penalties in later generations. Conversely, immigrants from low-penalty countries do not significantly increase their child penalty across generations: They start out with small child penalties in the first generation and continue to have small child penalties in later generations. Prima facie, this suggests that immigrants assimilate down, but not up. However, the asymmetry needs to be interpreted cautiously for the following reason: In general, foreign-born individuals have larger child penalties than US-born individuals, perhaps due to a fixed effect of migration on child penalties. This implies that, without any cultural assimilation, later-generation immigrants would have smaller penalties than first-generation immigrants regardless of their country of origin. Accounting for such effects, the evidence in Figure 19 is consistent with cultural assimilation among both low- and high-penalty immigrants. Rather than focusing on the aforementioned asymmetry, it is the narrowing of the gap between quartiles as we go from first-generation to later-generation penalties that provides evidence on cultural assimilation effects. Importantly, these effects could take many generations to materialize given later-generation immigrants include all descendants with a known country of ancestry regardless of the time that their ancestors arrived.

6 Conclusion

Recent work on gender inequality shows that child penalties — the causal impact of parenthood on women relative to men — account for most of the remaining 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.

Why are child penalties so large and persistent? There is evidence ruling out explanations rooted in comparative advantage (Kleven, Landais, and Søgaard 2021) and government policy (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2021), but there is little evidence ruling in explanations. Goldin (2014) argues that the last chapter of gender convergence must involve changes in labor market structure and job flexibility. Such aspects of labor market equilibrium may amplify or alleviate pre-existing differences between men and women, but they can only have an effect if *something else* tilts child care towards women. The unanswered question is why job flexibility matter for women and not for men? More fundamentally, why is the gendered homemaker-breadwinner institution so strong? This paper studies the role of social norms and culture from different angles. This includes studying elicited gender attitudes over time and space, and it includes conducting epidemiological studies of movers within the US and immigrants from abroad. The evidence consistently points to strong effects of gender norms and culture. As a result, the last chapter of gender convergence seems to require changes in cultural beliefs.

To produce these findings, the paper also makes a methodological contribution: A pseudo-event study approach to estimating child penalties that relies on widely available cross-sectional data. The applicability of the approach is wide-ranging due to the minimal data requirements. It allows for the possibility of mapping child penalties within and across countries around the world (Kleven, Landais, and Mariante 2022). This will facilitate a deeper understanding of the mechanisms that drive gender inequality at different levels of development and under different political institutions.

References

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

ANGRIST, JOSHUA D., AND WILLIAM N. EVANS (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 88, 450–477. 8

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. 4

BERTRAND, MARIANNE, CLAUDIA GOLDIN, AND LAWRENCE KATZ (2010): "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, 2, 228–255. 4

BLAU, FRANCINE, AND LAWRENCE KAHN (2006): "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence," *Industrial and Labor Relations Review*, 60(1), 45–66. 2, 12

——— (2017): "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 55(3), 789–865. 2, 12

BOELMANN, BARBARA, ANNA RAUTE, AND UTA SCHÖNBERG (2021): "Wind of Change? Cultural Determinants of Maternal Labor Supply," Working Paper. 4

BROWNING, MARTIN (1992): "Children and Household Economic Behavior," *Journal of Economic Literature*, 30, 1434–1475. 8

CORTÉS, PATRICIA, AND JESSICA PAN (2020): "Children and the Remaining Gender Gaps in the Labor Market," *Journal of Economic Literature*, p. forthcoming. 1, 2, 4

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, 21

FERNÁNDEZ, RAQUEL, AND ALESSANDRA FOGLI (2009): "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics*, 1(1), 146–177. 4

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

Goldin, Claudia (2014): "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104, 1091–1119. 3, 14, 27

Goldin, Claudia, and Lawrence Katz (2016): "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation," *Journal of Labor Economics*, 34, 705–746. 3, 14

Kleven, Henrik (2021a): "The EITC and the Extensive Margin: A Reappraisal," NBER Working Paper No. 26405. 5

——— (2021b): "Lecture 1: The Child Penalty," Zeuthen Lecture Series, September 2021. 2, 13

——— (2021c): "Lecture 2: The Geography of Child Penalties," Zeuthen Lecture Series, September 2021. 23

——— (2021d): "Lecture 3: Public Policy and Child Penalties," Zeuthen Lecture Series, September 2021. 16

Kleven, Henrik, Camille Landais, and Gabriel Leite Mariante (2022): "The Child Penalty Atlas," Working Paper, forthcoming. 2, 4, 23, 27, 33, 49, 50, 52, 71, 72

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, 8, 13, 20, 27

——— (2021): "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation," NBER Working Paper No. 28082. 1, 4, 27

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019): "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11(4), 181–209. 1, 4, 6, 7, 8, 9, 14, 17, 27

——— (2021): "Does Biology Drive Child Penalties? Evidence from Biological and Adoptive Families," *American Economic Review: Insights*, 3, 183–198. 1, 3, 4, 15, 16, 27

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, 12

LUNDBORG, PETTER, ERIK PLUG, AND ASTRID WÜRTZ RASMUSSEN (2017): "Can Women Have Children and a Career? IV Evidence from IVF Treatments," *American Economic Review*, 107(6), 1611–37. 8

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

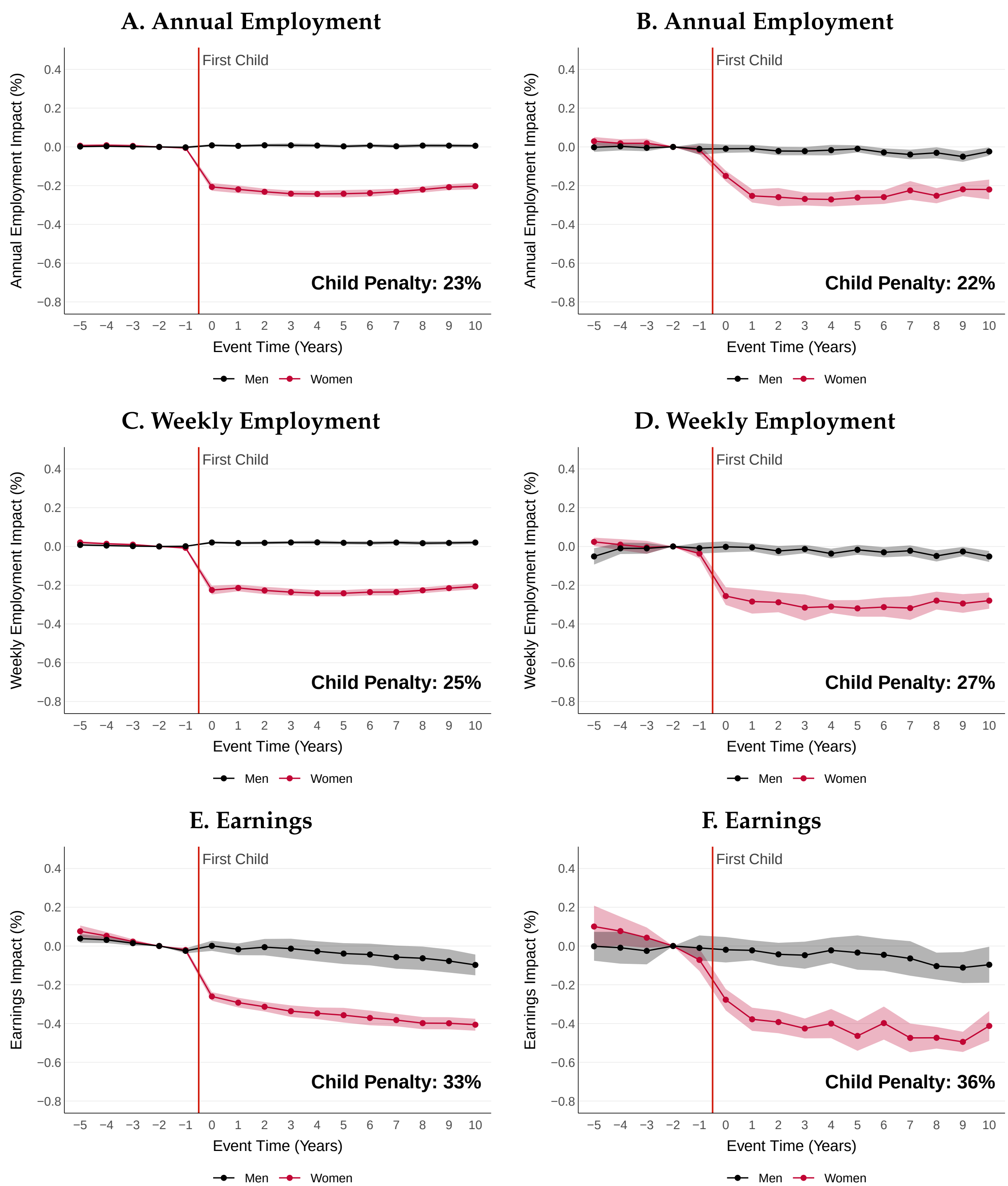
TABLE 3: SELECTION OF MOVERS AND IMMIGRANTS BY PLACE OF BIRTH

	A. US Movers by State of Birth			B. US Immigrants by Country of Birth		
	High-Penalty States	Low-Penalty States	**Difference**	High-Penalty Countries	Low-Penalty Countries	**Difference**
Demographic Characteristics of Mothers:						
Fraction Living in High-Penalty States	0.22	0.16	**0.06**	0.22	0.27	**-0.05**
Fraction High School or Below	0.13	0.12	**0.01**	0.47	0.31	**0.17**
Fraction College	0.60	0.61	**-0.01**	0.33	0.48	**-0.15**
Fraction Married	0.84	0.84	**0.00**	0.82	0.84	**-0.03**
Fraction Black	0.04	0.07	**-0.04**	0.01	0.15	**-0.14**
Fraction White	0.91	0.87	**0.04**	0.67	0.24	**0.44**
Fertility	1.78	1.74	**0.04**	1.72	1.63	**0.08**
Age at First Birth	31.16	31.19	**-0.03**	30.66	31.32	**-0.66**
Age	37.36	37.44	**-0.08**	36.54	37.10	**-0.56**
Cohort	1973.46	1973.18	**0.28**	1972.74	1972.75	**-0.01**
Number of Observations	81,339	81,367		189,744	112,723	

Notes: This table provides evidence on 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.16), while the child penalties used to split immigrants by country of birth are taken from Kleven, Landais, and Mariante (2022). The table compares the demographic characteristics of mothers by place of birth. 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.18.

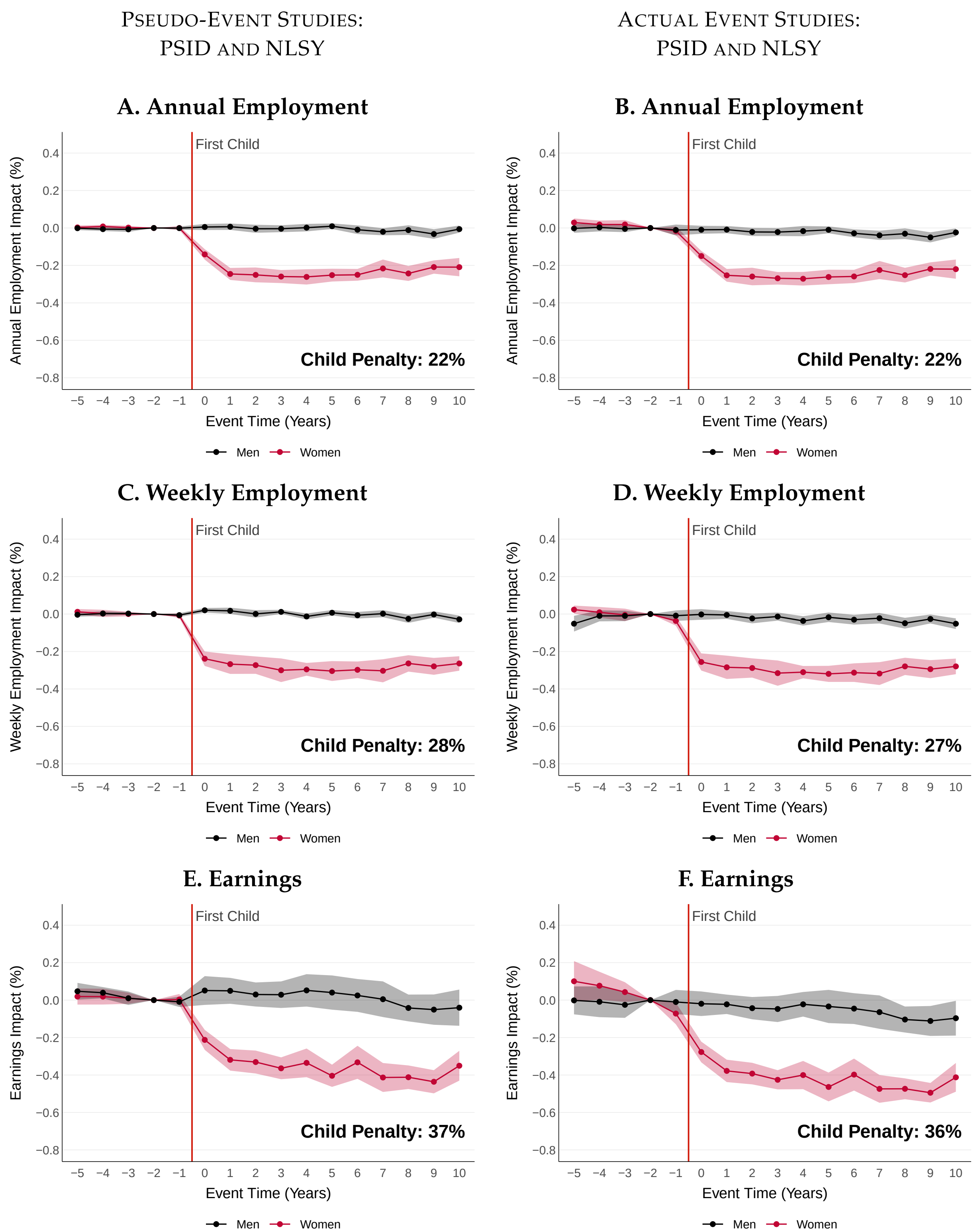
FIGURE 1: VALIDATION OF PSEUDO-EVENT STUDY APPROACH

PSEUDO-EVENT STUDIES: CPS AND ACS | ACTUAL EVENT STUDIES: PSID AND NLSY



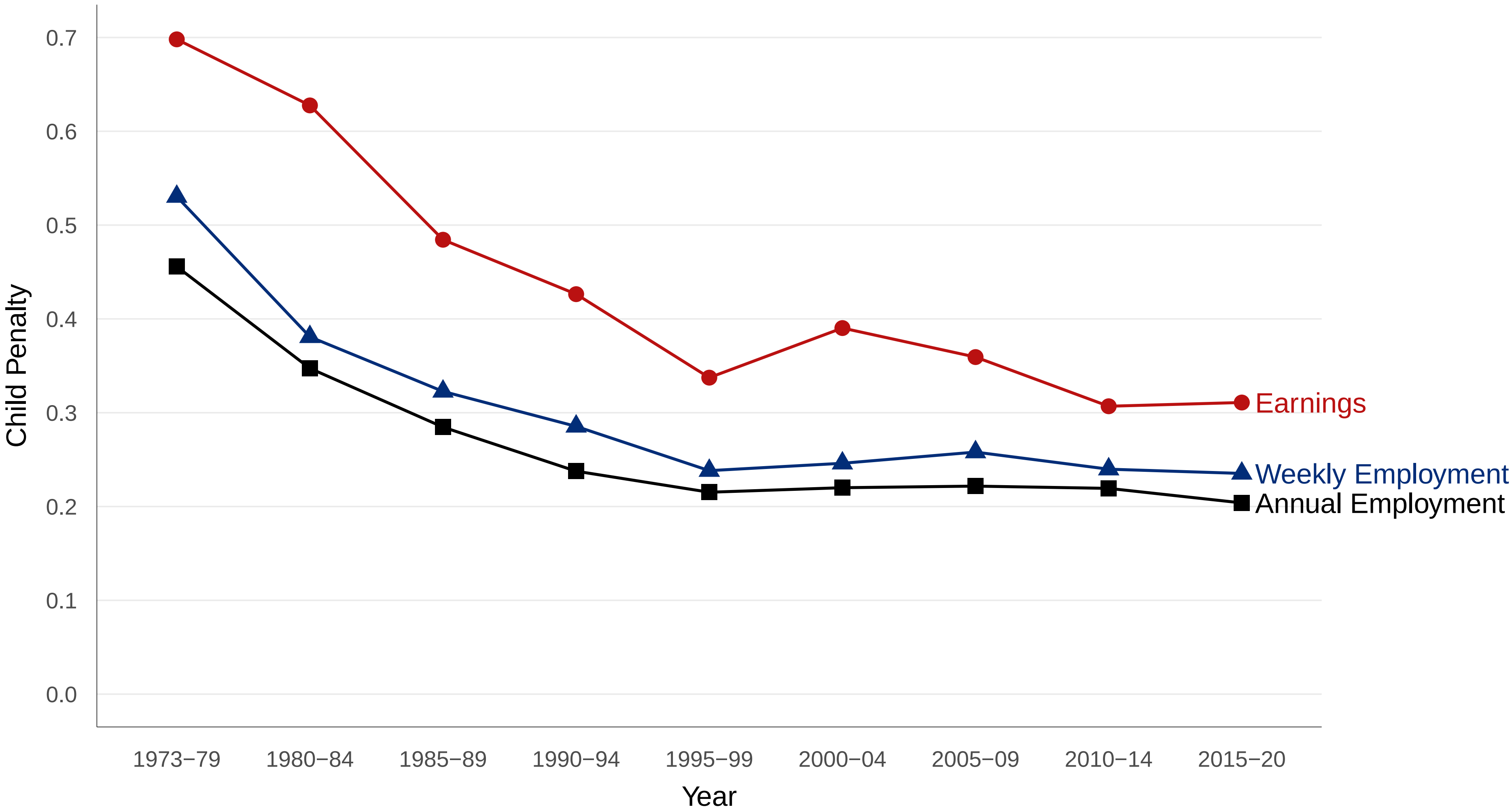
Notes: This figure validates the pseudo-event study approach (left panels) against an actual event study approach (right panels). The pseudo-event studies are based on pooled CPS and ACS data from 1968-2020, while the actual event studies are based on pooled PSID and NLSY data from 1968-2019. Each panel shows an event study for men and women around the birth of their first child at $t = 0$. The series show the percentage impact of child birth on men and women at each event time t, i.e. $\hat{P}_t^m$ and $\hat{P}_t^w$ estimated from equations (1)-(2). Each panel also displays the average child penalty over event times 0-10 defined as in equation (3). 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



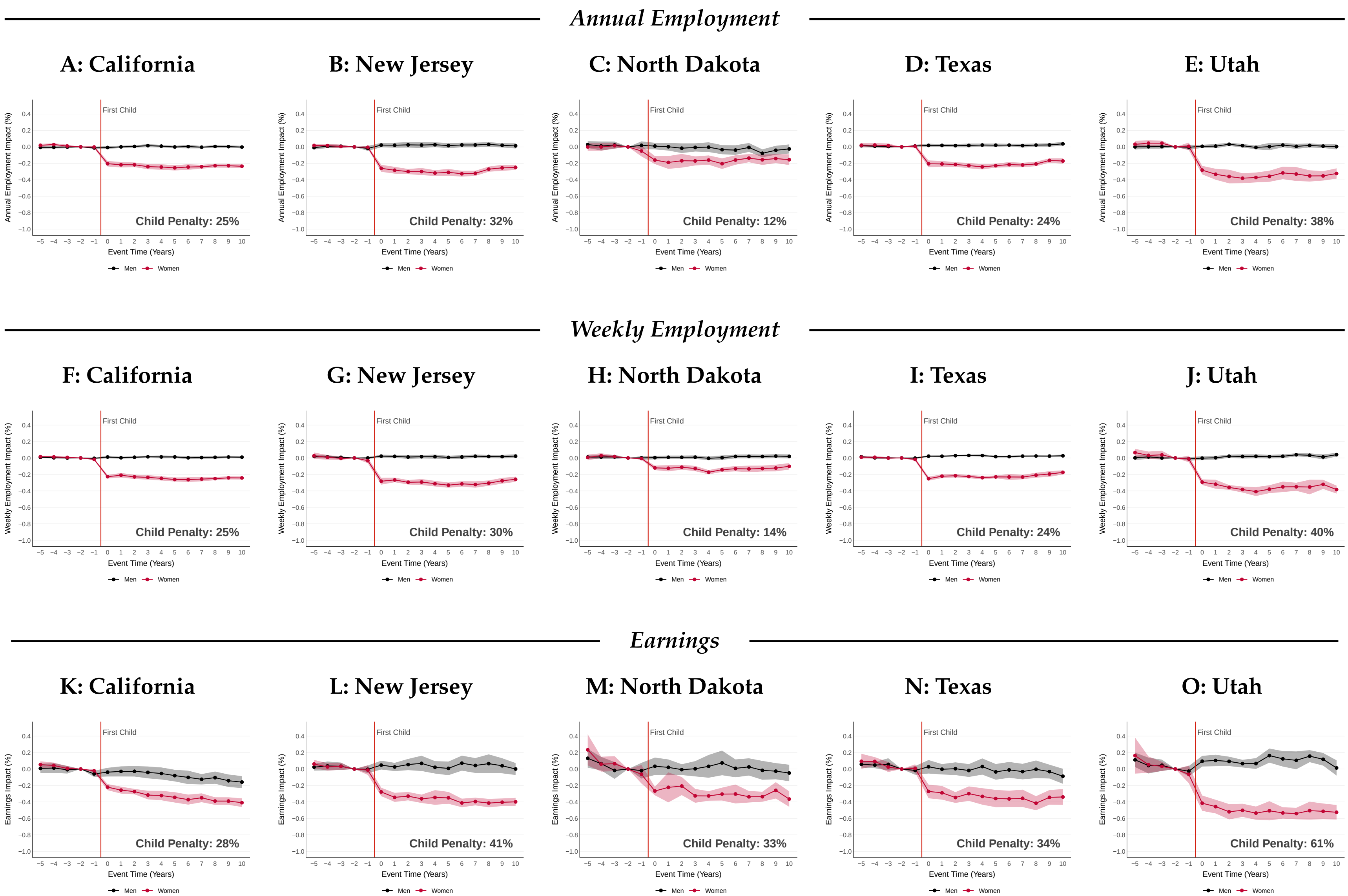
Notes: This figure validates the pseudo-event study approach (left panels) against an actual event study approach (right panels), both using pooled PSID and NLSY data from 1968-2019. Each panel shows an event study for men and women around the birth of their first child at $t = 0$. The series show the percentage impact of child birth on men and women at each event time t, i.e. $\hat{P}_t^m$ and $\hat{P}_t^w$ estimated from equations (1)-(2). Each panel also displays the average child penalty over event times 0-10 defined as in equation (3). 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: 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 show the average child penalty over event times 0-10 (defined in equation 3) in different time intervals. These are estimated by splitting the sample of parents by interview year and running the event study specification (1) 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 surrogate 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.5-A.7.

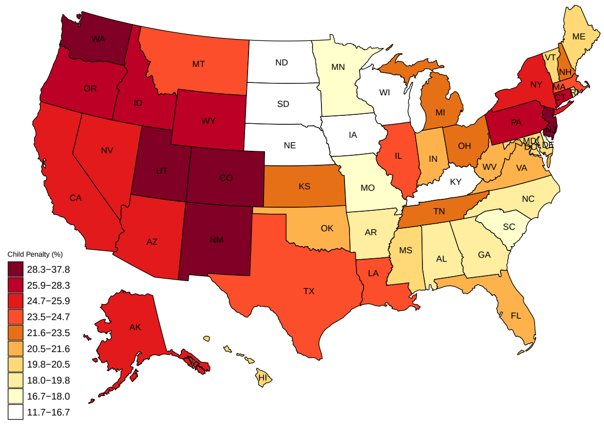
Figure 4: Event Studies of First Child Birth in Selected States



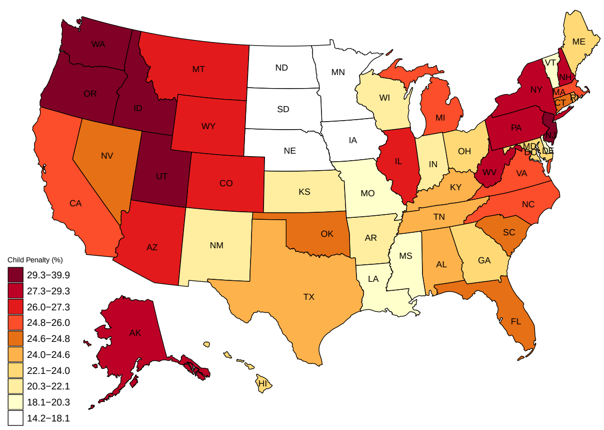
Notes: This figure shows event studies of first child birth for five US states and each of the three labor market outcomes. State-level event studies are constructed by interacting the event time dummies in equation (1) with state dummies, estimating percentage impacts of child birth on men and women at each event time ($\hat{P}_t^m$ and $\hat{P}_t^w$) 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 (1) 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.9-A.11.

FIGURE 5: HEATMAPS OF CHILD PENALTIES

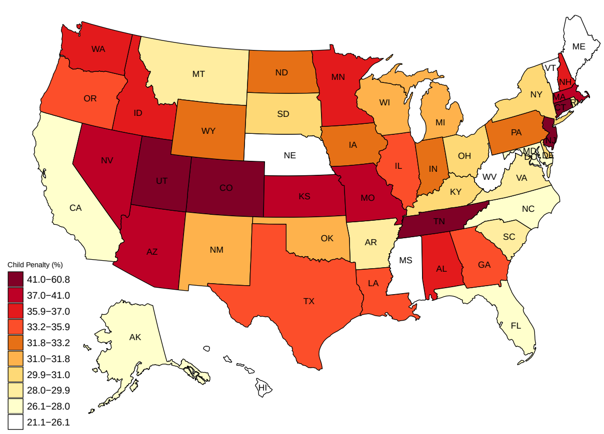
A. Annual Employment



B. Weekly Employment

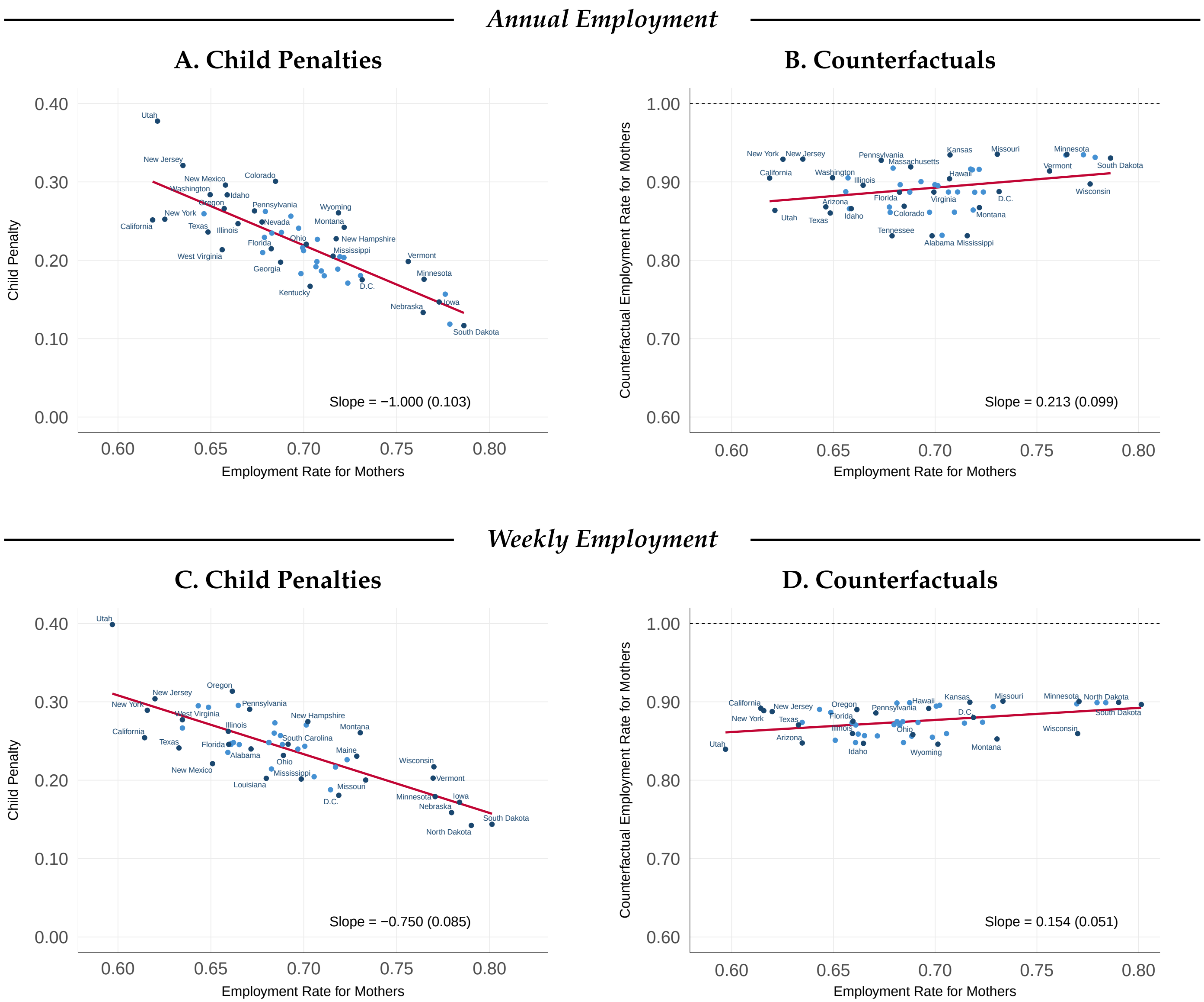


C. Earnings



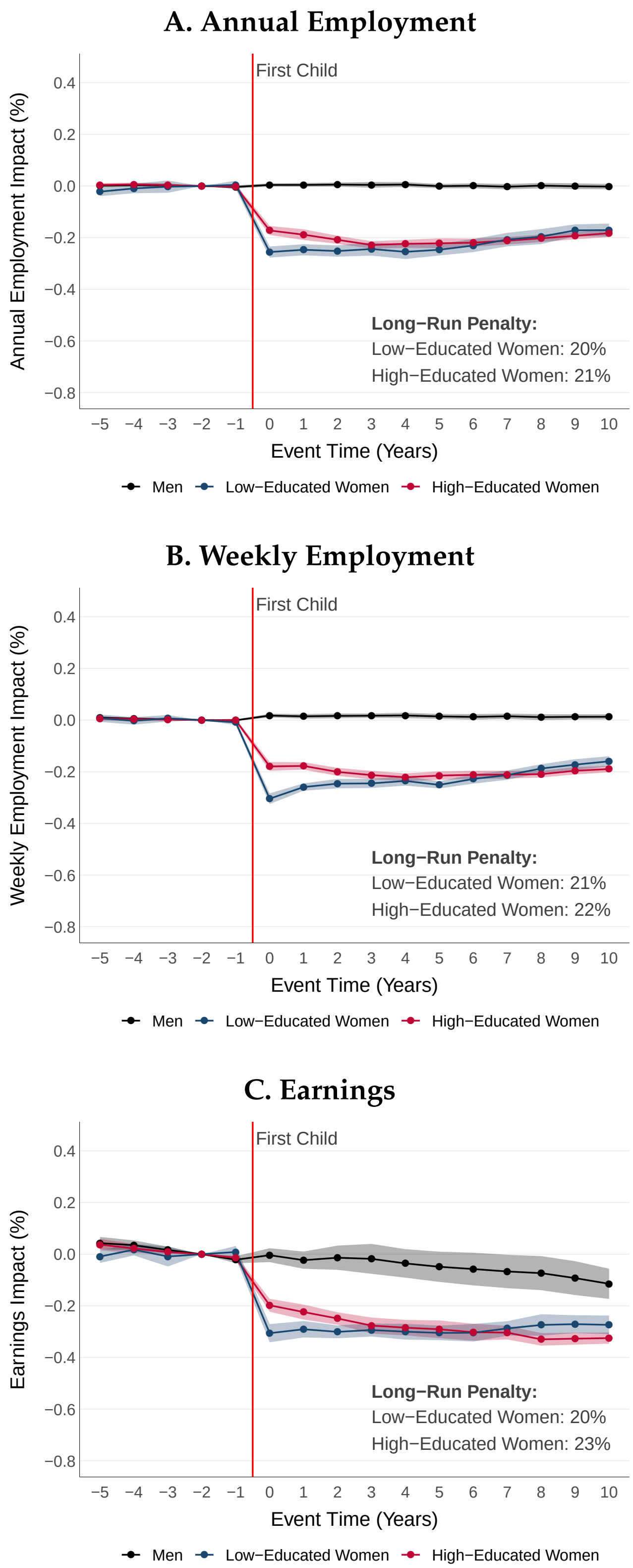
Notes: This figure summarizes the results from the state-level event studies of child birth (shown in Figures A.9-A.11 of the appendix) in heatmaps. In these maps, states are divided into deciles of the child penalty (as defined in equation 3), with darker colors implying larger child penalties.

FIGURE 6: CHILD PENALTIES AND COUNTERFACTUAL EMPLOYMENT RATES ACROSS STATES



Notes: This figure provides scatterplots of child penalties in employment (left panels) and counterfactual employment rates for mothers (right panels) against observed employment rates for mothers across states. The counterfactual employment rate for mothers is calculated as their average employment rate over event times 0-10 absent the impact of children, i.e. the average predicted outcome from equation (1) when omitting the contribution of the event time coefficients. In the right panels, the upper bound on the counterfactual employment rate — an employment rate of one — is marked by the dashed horizontal line. The relationship between child penalties and observed employment rates is strongly negative, whereas the relationship between counterfactual employment rates (adding back the child penalty) and observed employment rates is relatively flat.

Figure 7: Event Studies of First Child Birth by Education



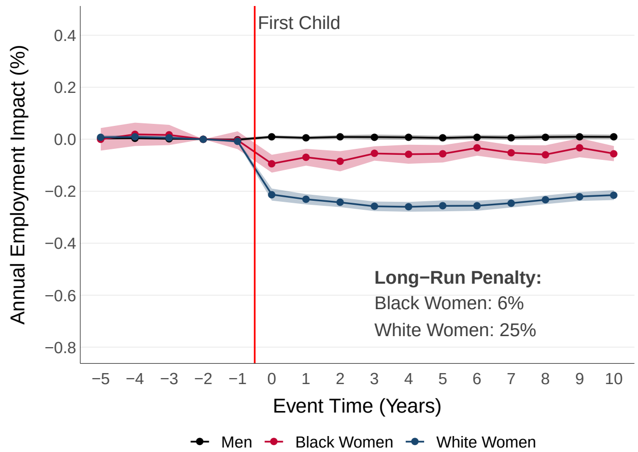
Notes: This figure presents event studies of first child birth by female education level. The figure is constructed by running specification (1) separately for low-educated women (high school degree or less) and high-educated women (college degree or more). The sample of men is not split by education as child birth is a non-event for them. 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. In all three outcomes, long-run child penalties on low- and high-educated women are very similar (and the small difference is statistically insignificant), inconsistent with specialization based on comparative advantage.

FIGURE 8: EVENT STUDIES OF FIRST CHILD BIRTH BY MARITAL STATUS

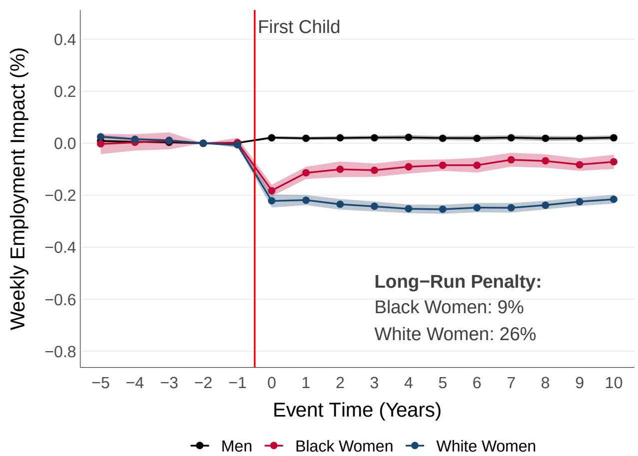
Notes: This figure presents event studies of first child birth by female marital status. The figure is constructed by running specification (1) separately for married and single women, where single is defined as never married, separated, divorced, or widowed. The sample of men is not split by marital status as child birth is a non-event for them. 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. In all three outcomes, the long-run child penalty on single mothers is much smaller than on married mothers.

FIGURE 9: EVENT STUDIES OF FIRST CHILD BIRTH BY RACE

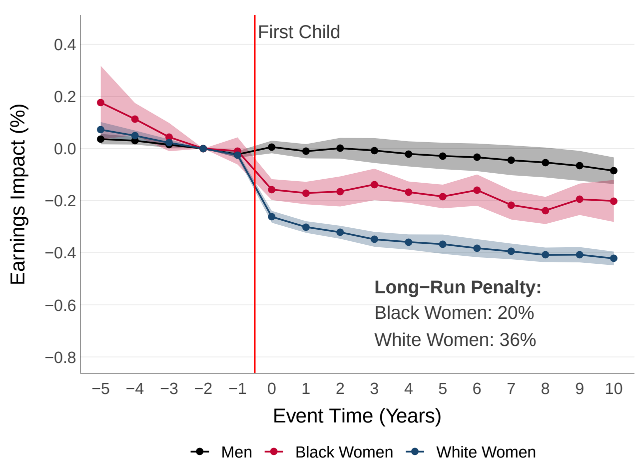
A. Annual Employment



B. Weekly Employment

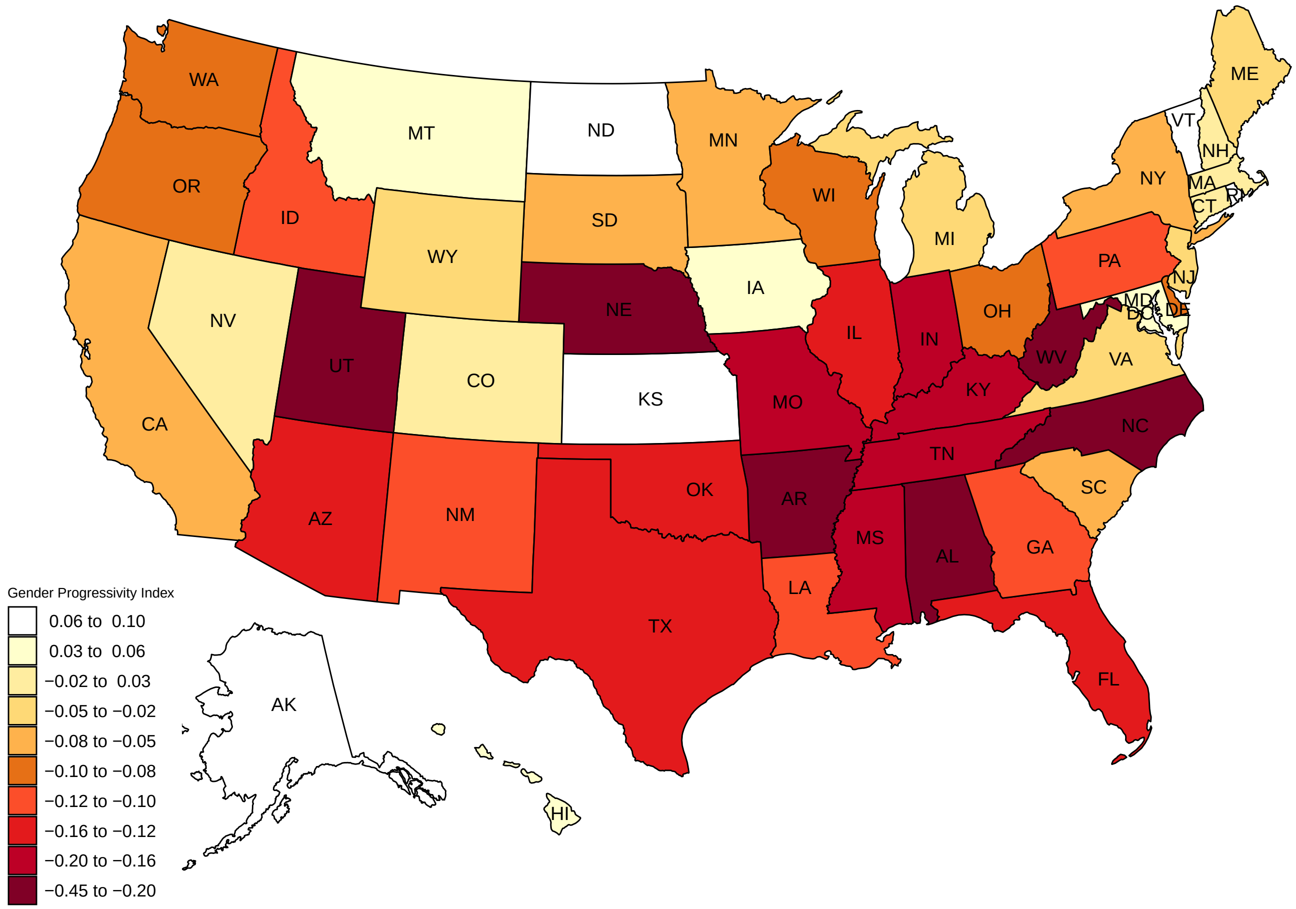


C. Earnings



Notes: This figure presents event studies of first child birth by female race. The figure is constructed by running specification (1) separately for black and white women. The sample of men is not split by race as child birth is a non-event for them. 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. In all three outcomes, the long-run child penalty on black mothers is much smaller than on white mothers.

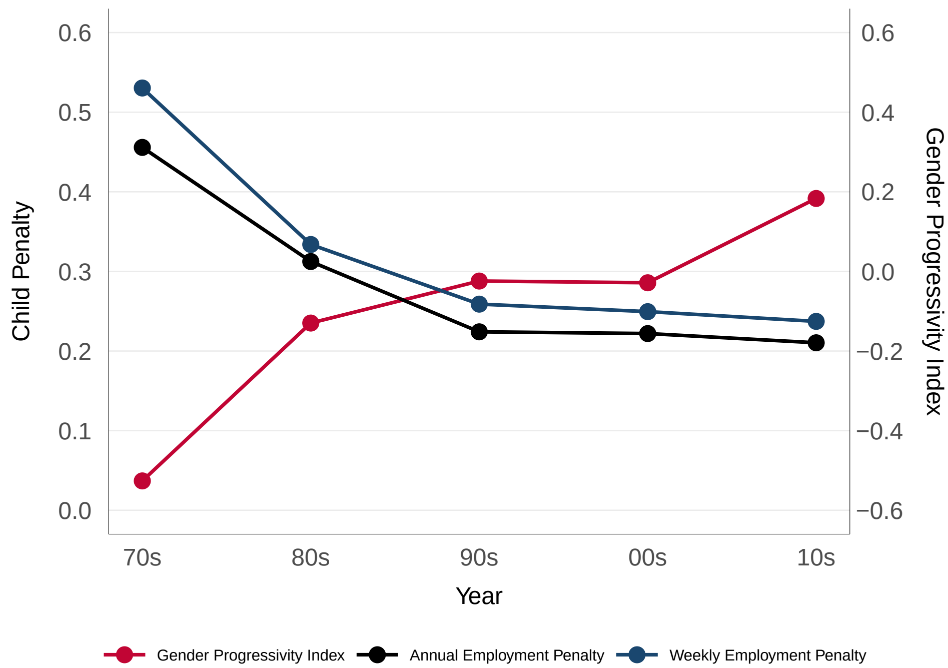
FIGURE 10: HEATMAP OF GENDER NORMS



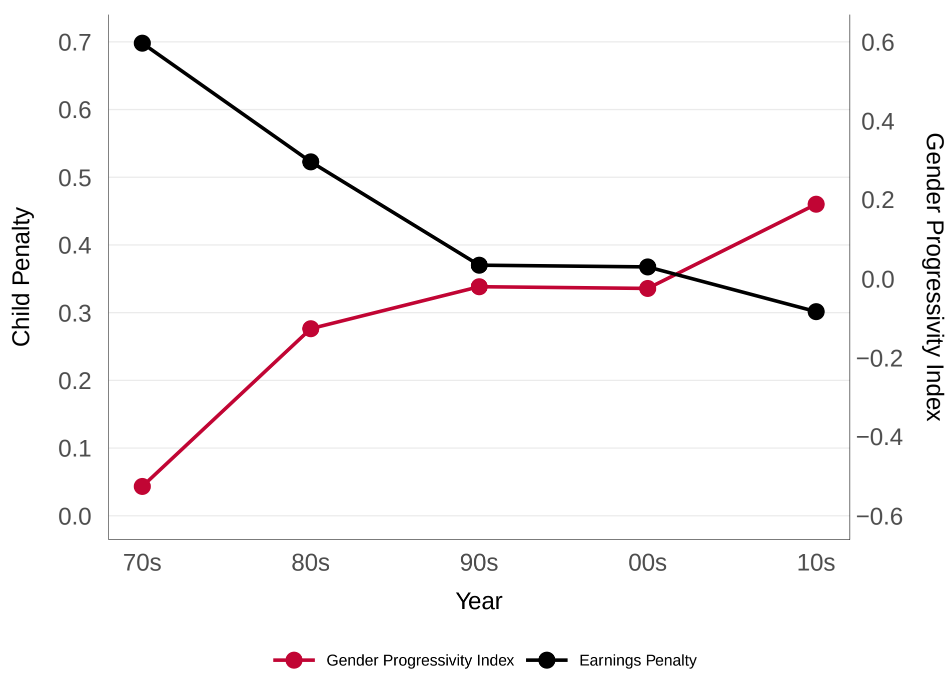
Notes: This figure presents a heatmap of gender norms using GSS data from 1972-2018. States are divided into deciles of a Gender Progressivity Index (GPI). This index is calculated as the average standardized response to GSS questions that elicit attitudes towards gender roles in families with children. The standardization ensures that the index has mean zero and standard deviation one. Three GSS questions that are consistently measured over time are included. A higher value of GPI corresponds to a more gender progressive norm, i.e. attitudes more favorable towards shared responsibility for child care and market work between men and women. In the map, lighter colors represent more progressive norms and darker colors represent more conservative norms.

FIGURE 11: CHILD PENALTIES VS GENDER PROGRESSIVITY OVER TIME

A. Employment



B. Earnings

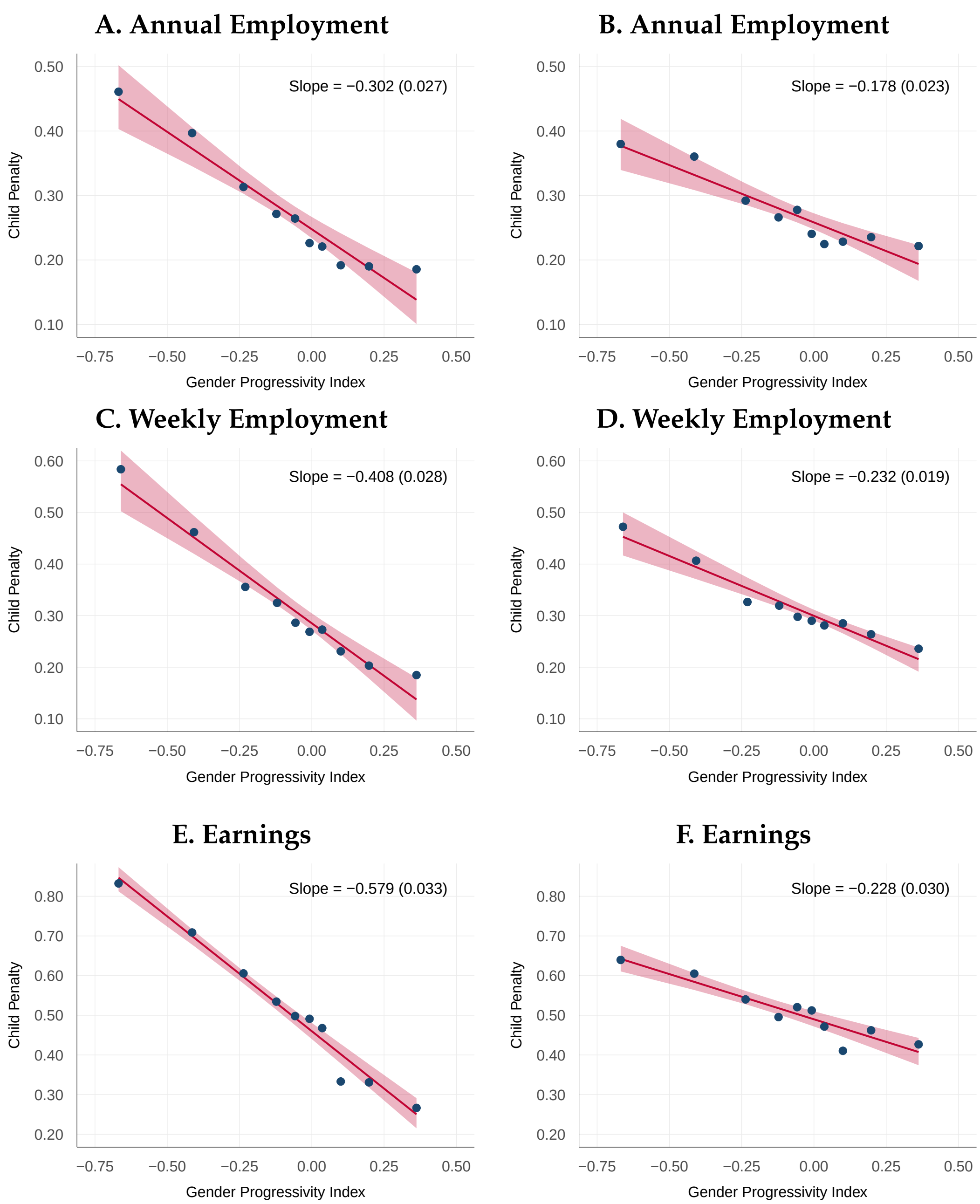


Notes: This figure plots the evolution of child penalties and gender progressivity over the last 50 years. Employment penalties are shown in Panel A and earnings penalties in Panel B. The construction of the Gender Progressivity Index (GPI) is described in the notes to the preceding figure. The GPI time 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 12: CHILD PENALTIES VS GENDER PROGRESSIVITY ACROSS STATES AND TIME

BINSCATTER:
STATE FIXED EFFECTS

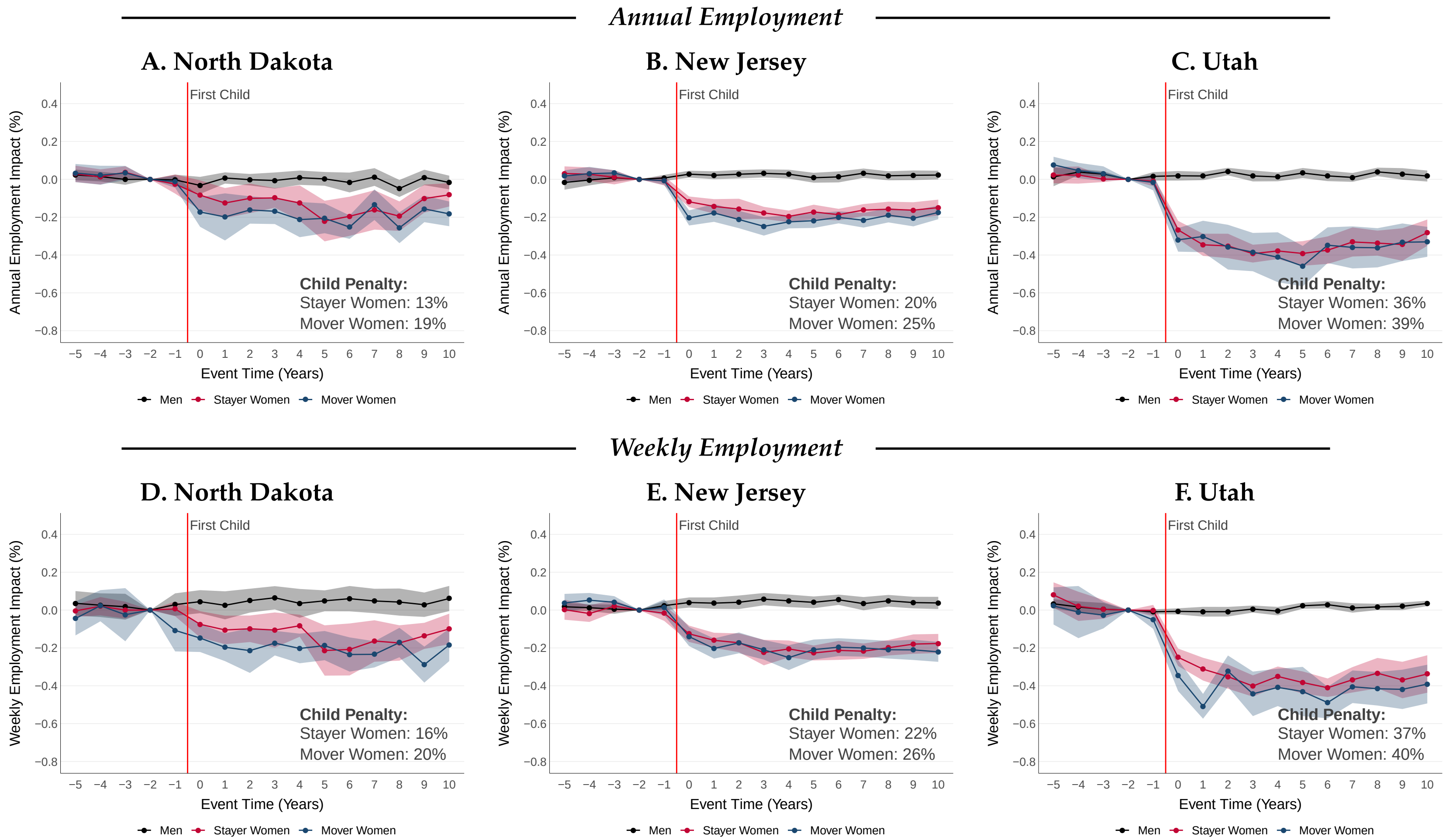
BINSCATTER:
STATE FIXED EFFECTS & DEMOGRAPHICS



Notes: This figure provides binscatters of child penalties vs gender progressivity across states and time. The analysis is based on equation (4), 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 13: EPIDEMIOLOGICAL STUDY OF US MOVERS

EVENT STUDIES OF FIRST CHILD BIRTH FOR MOVERS VS STAYERS IN SELECTED STATES

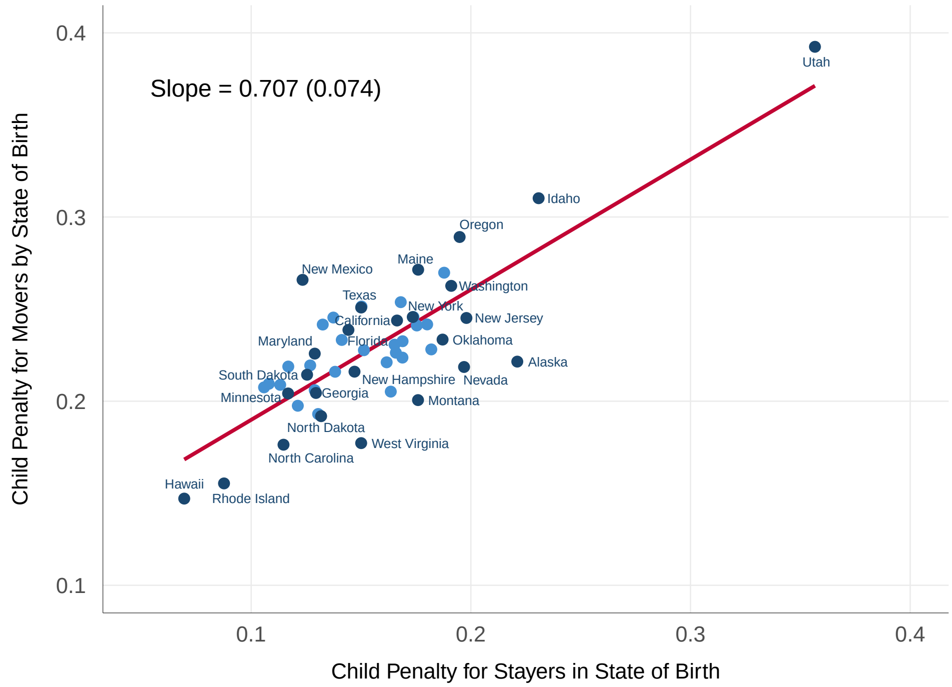


Notes: This figure presents event studies of first child birth for movers and stayers born in different states. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. To construct the figure, specification (1) 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.16-A.17.

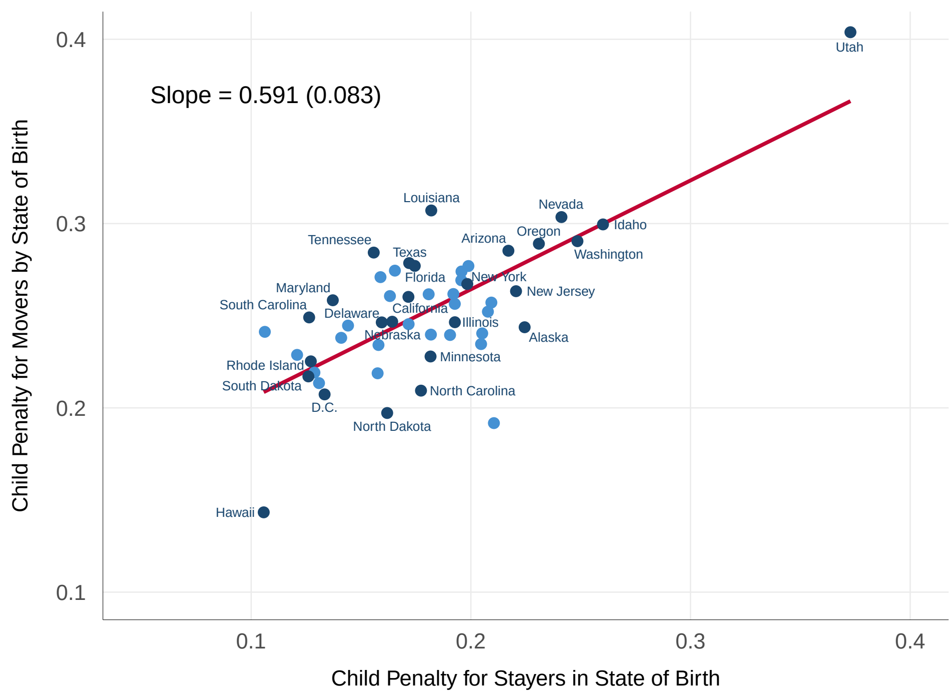
Figure 14: Epidemiological Study of US Movers

Child Penalties for Movers vs Stayers by State of Birth

A. Annual Employment



B. Weekly Employment

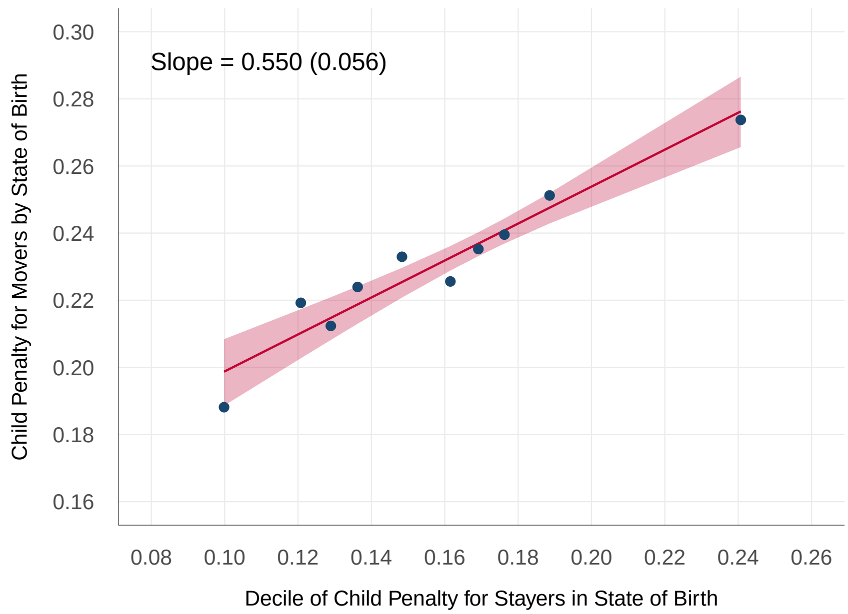


Notes: This figure provides scatterplots of the child penalty for movers against the child penalty for stayers by state of birth. Movers are defined as US-born individuals who reside in a different state than where they were born, while stayers are defined as US-born individuals who reside in the same state as where they were born. Results are shown for annual employment (Panel A) and weekly employment (Panel B). The event studies used to calculate child penalties for movers and stayers in each outcome and in each state are presented in Appendix Figures A.16-A.17. The sample is based on ACS data from 2000-2019, which contains information on both state of residence and state of birth.

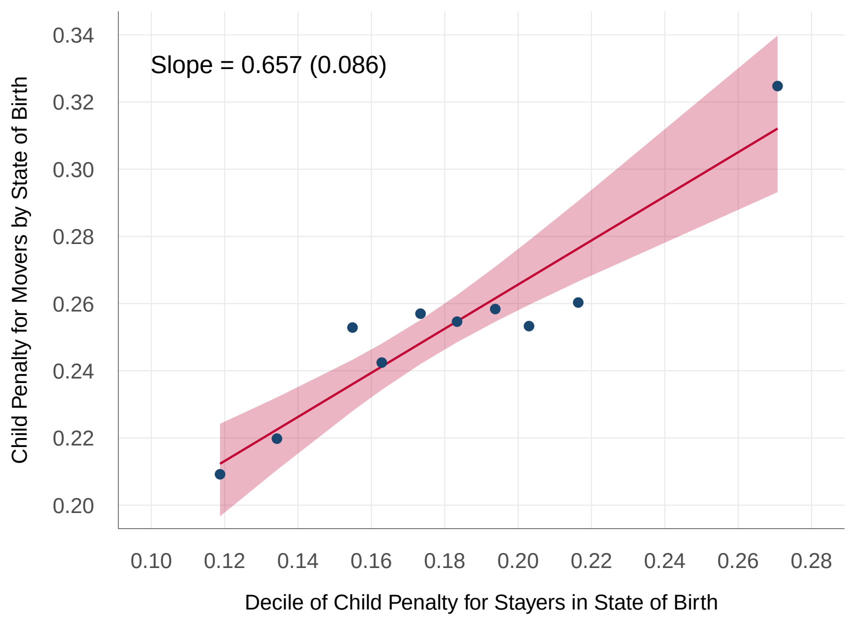
FIGURE 15: EPIDEMIOLOGICAL STUDY OF US MOVERS

CHILD PENALTIES FOR MOVERS VS STAYERS BY DECILE OF STATE OF BIRTH
CONTROLS FOR STATE OF RESIDENCE

A. Annual Employment



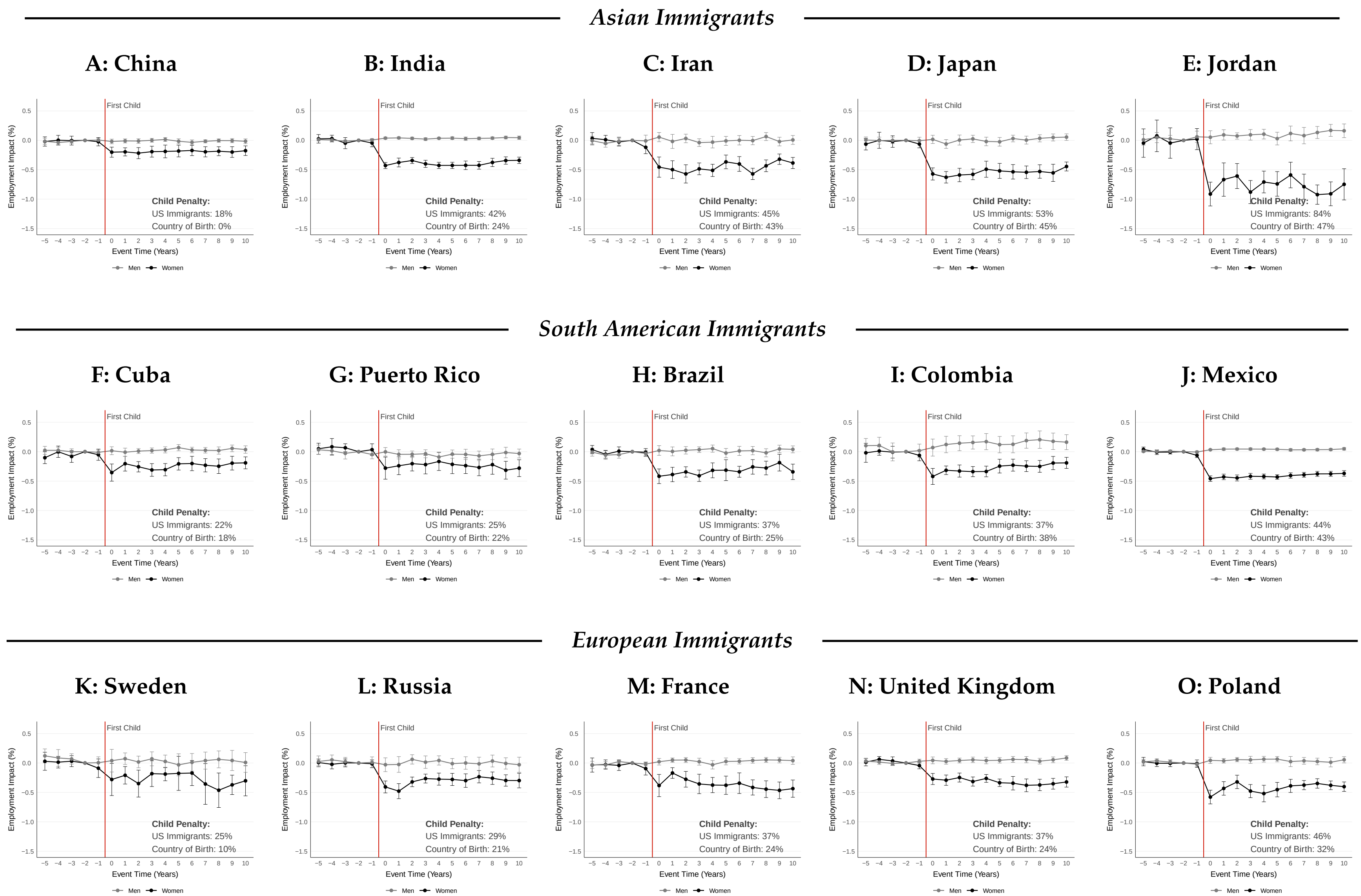
B. Weekly Employment



Notes: This figure presents binscatters of the child penalty for movers by decile of the child penalty for stayers in their state of birth, controlling for state of residence. To construct the figure, the child penalty for movers by state of birth is regressed on child penalty for stayers and controls for the fraction of movers residing in different deciles of stayer penalties. The mover penalties are then residualized by the estimated effect of the residence controls and plotted against stayer penalties by state of birth. When plotting the residualized mover penalties, the average effect of the controls is added to the residuals to make the levels of the penalties comparable to those in the preceding figure. Controlling for differences in residence choices across movers from different states does not qualitatively alter the findings from the preceding figure, suggesting that selection on state of residence is not a threat to interpretation.

Figure 16: Epidemiological Study of Foreign Immigrants

Event Studies of First Child Birth for Immigrants by Country of Birth

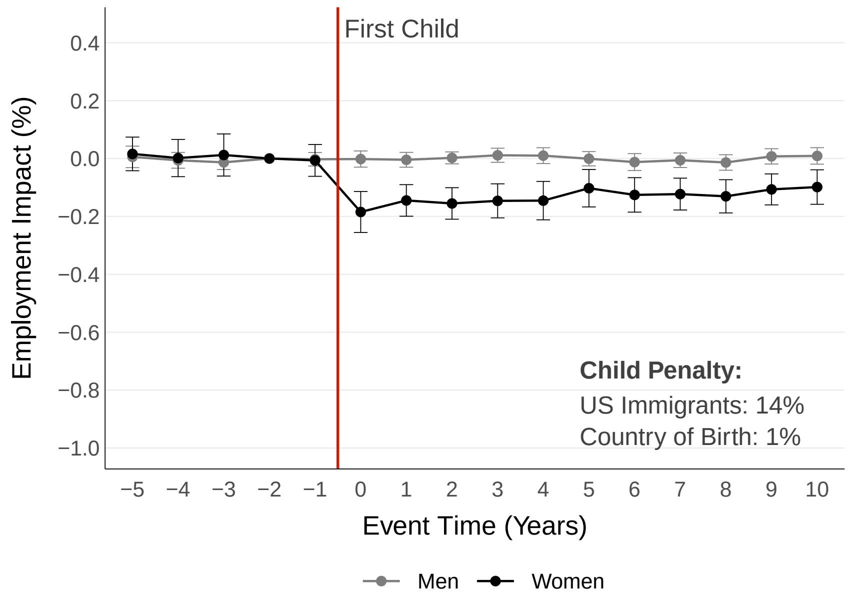


Notes: This figure presents event studies of first child birth for foreign-born immigrants by country of birth. Results are shown for selected countries in Asia (top row), South America (middle row), and Europe (bottom row), while results for a much larger set of countries are shown in Appendix Figure A.18. 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 Mariante 2022), ordering panels by the child penalty in country of birth. The event studies for US immigrants from different countries are estimated by interacting the event time dummies in equation (1) with country-of-birth dummies. 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.

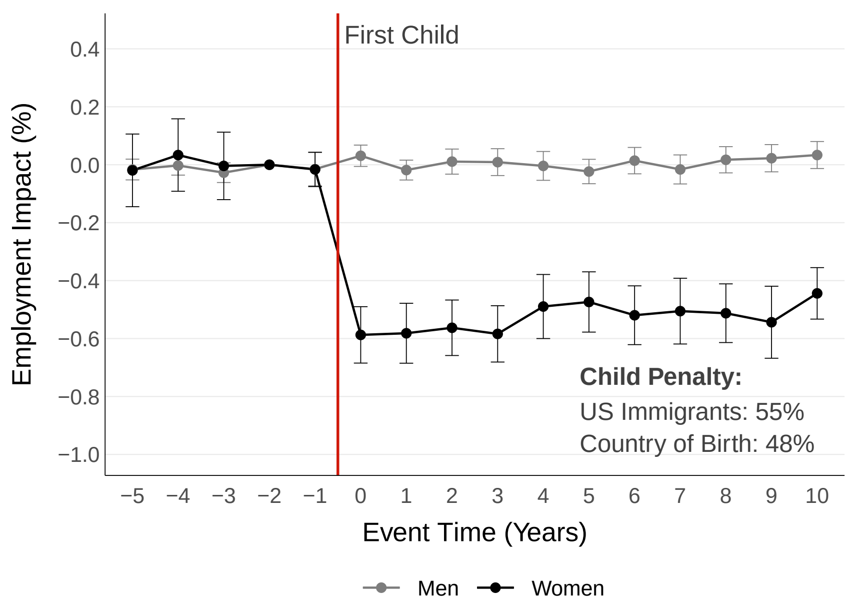
FIGURE 17: EPIDEMIOLOGICAL STUDY OF FOREIGN IMMIGRANTS

EVENT STUDIES OF FIRST CHILD BIRTH FOR IMMIGRANTS BY DECILE OF BIRTH-COUNTRY PENALTIES

A. Bottom Decile of Child Penalty in Country of Birth



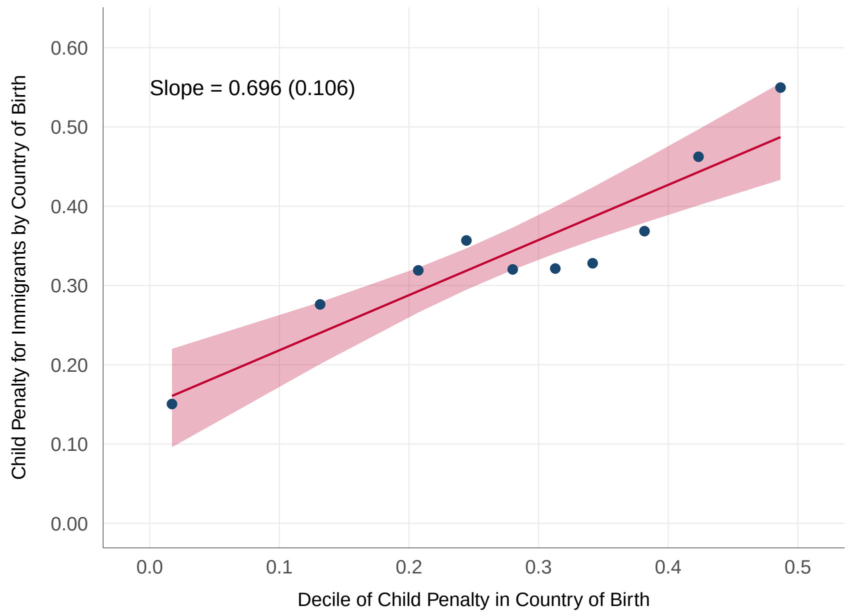
B. Top Decile of Child Penalty in Country of Birth



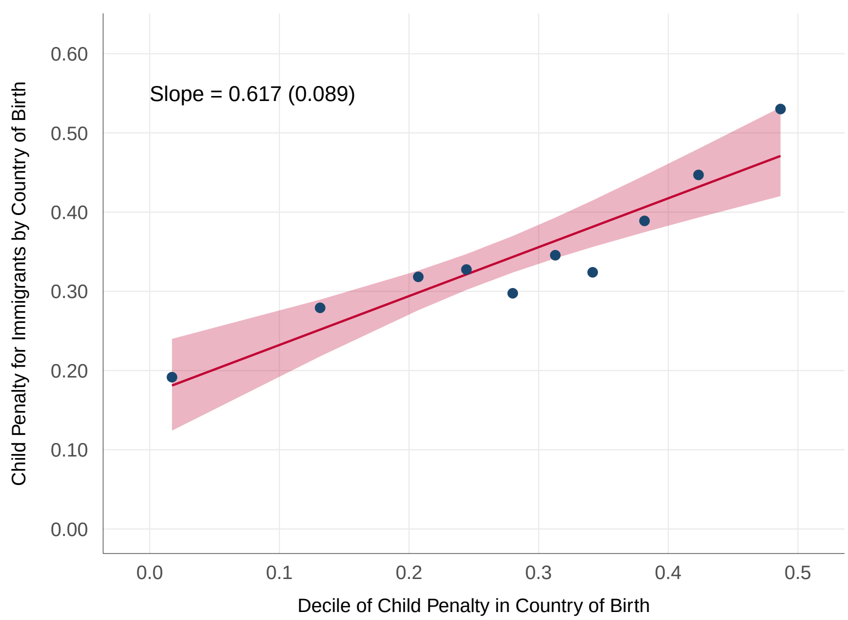
Notes: This figure presents event studies of first child birth for foreign-born immigrants in the bottom and top deciles of birth-country penalties. Countries are assigned to deciles using the child penalty estimates in Kleven, Landais, and Mariante (2022) for the sample of 66 countries shown in Appendix Figure A.18. The figure is constructed by running the event study specification (1) separately for each decile, graphing the percentage impacts on men and women at each event time t (as defined in equation (2)). 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 18: Epidemiological Study of Foreign Immigrants
Child Penalties for Immigrants by Decile of Birth-Country Penalties

A. Raw Data



B. With Controls

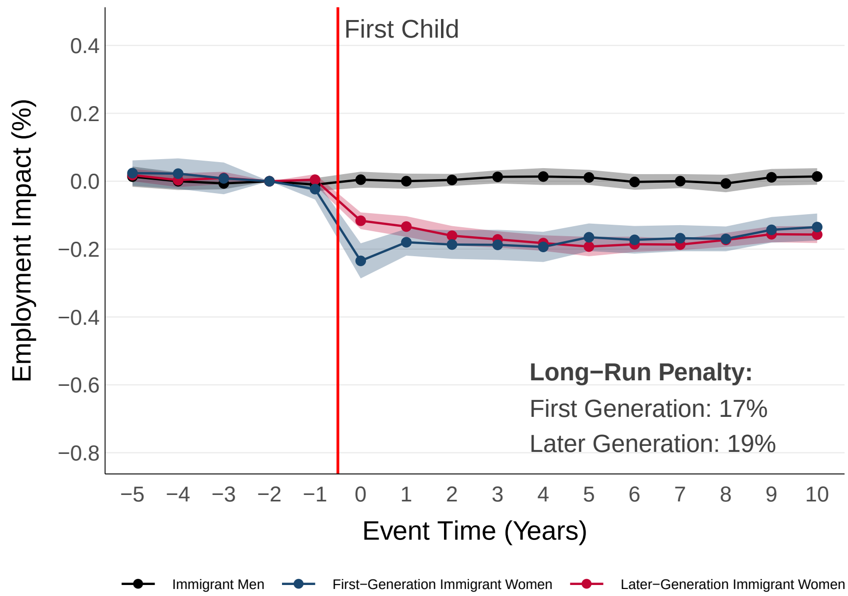


Notes: This figure presents binscatters of child penalties for foreign-born immigrants by decile of the child penalty in country of birth. Appendix Figure A.19 provides country-level scatters, which show the country-level penalties used to construct the decile-level penalties plotted here. Panel A shows 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. 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 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.

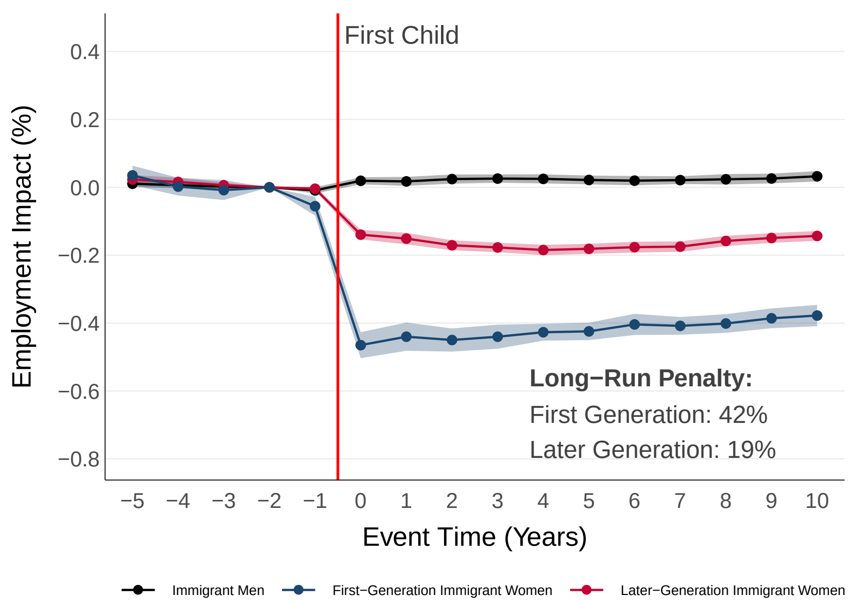
FIGURE 19: CULTURAL ASSIMILATION OF IMMIGRANTS

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



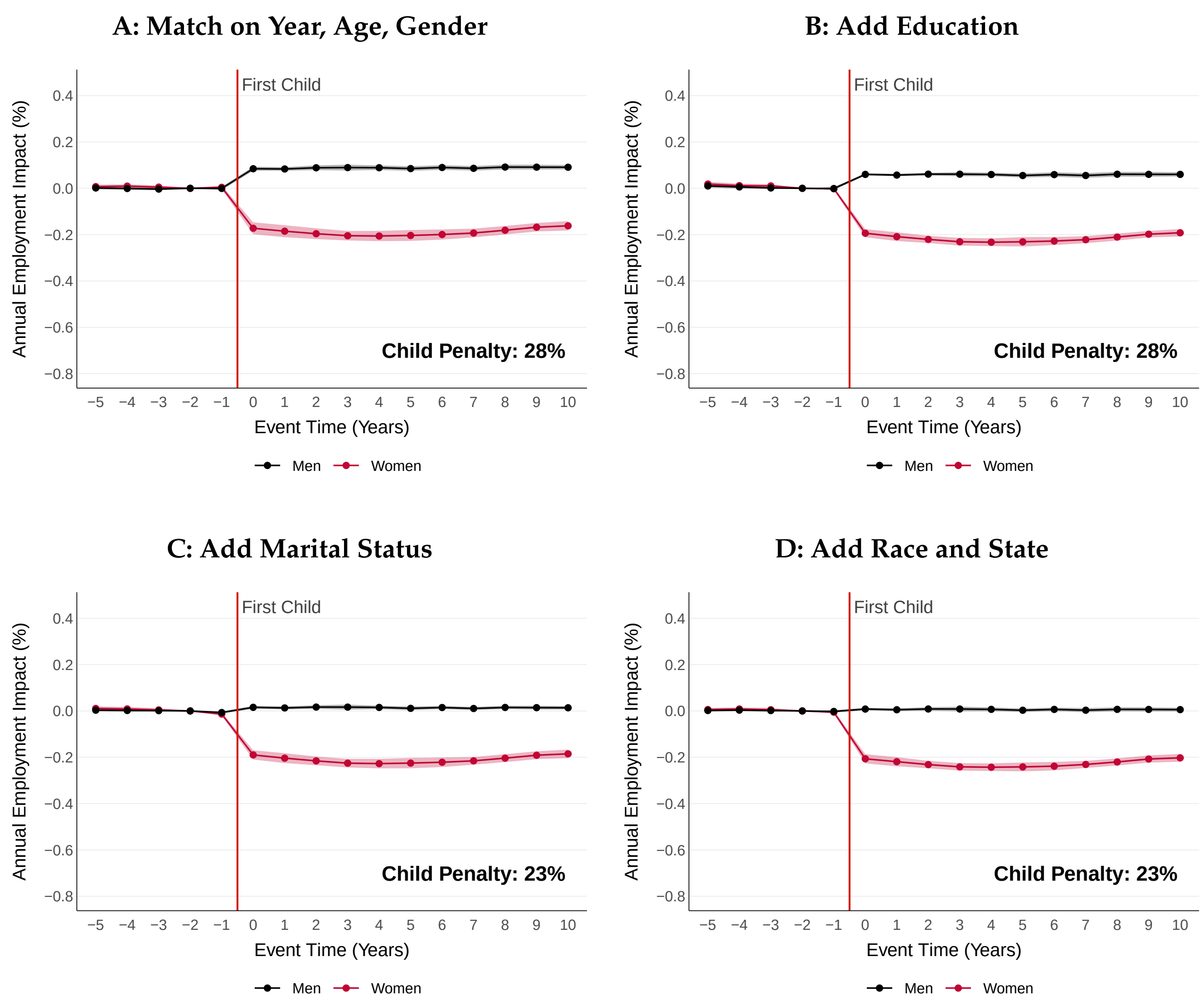
Notes: This figure presents event studies of first child birth for first-generation and later-generation immigrants by quartile of the child penalty in country of origin. First-generation immigrants are defined as foreign-born US residents, while later-generation immigrants are defined as US-born residents who report foreign ancestry. The analysis is based on the 66 countries shown in Appendix Figure A.18, dividing countries into quartiles of the child penalty using the estimates in Kleven, Landais, and Mariante (2022). The figure is constructed by running the event study specification (1) for first- and later-generation immigrant women separately (within the bottom and top quartile of origin-country penalty, respectively). Each panel displays the long-run child penalty (over event times 5-10) for US 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.

Online Appendix (Not for Publication)

A Supplementary Figures and Tables

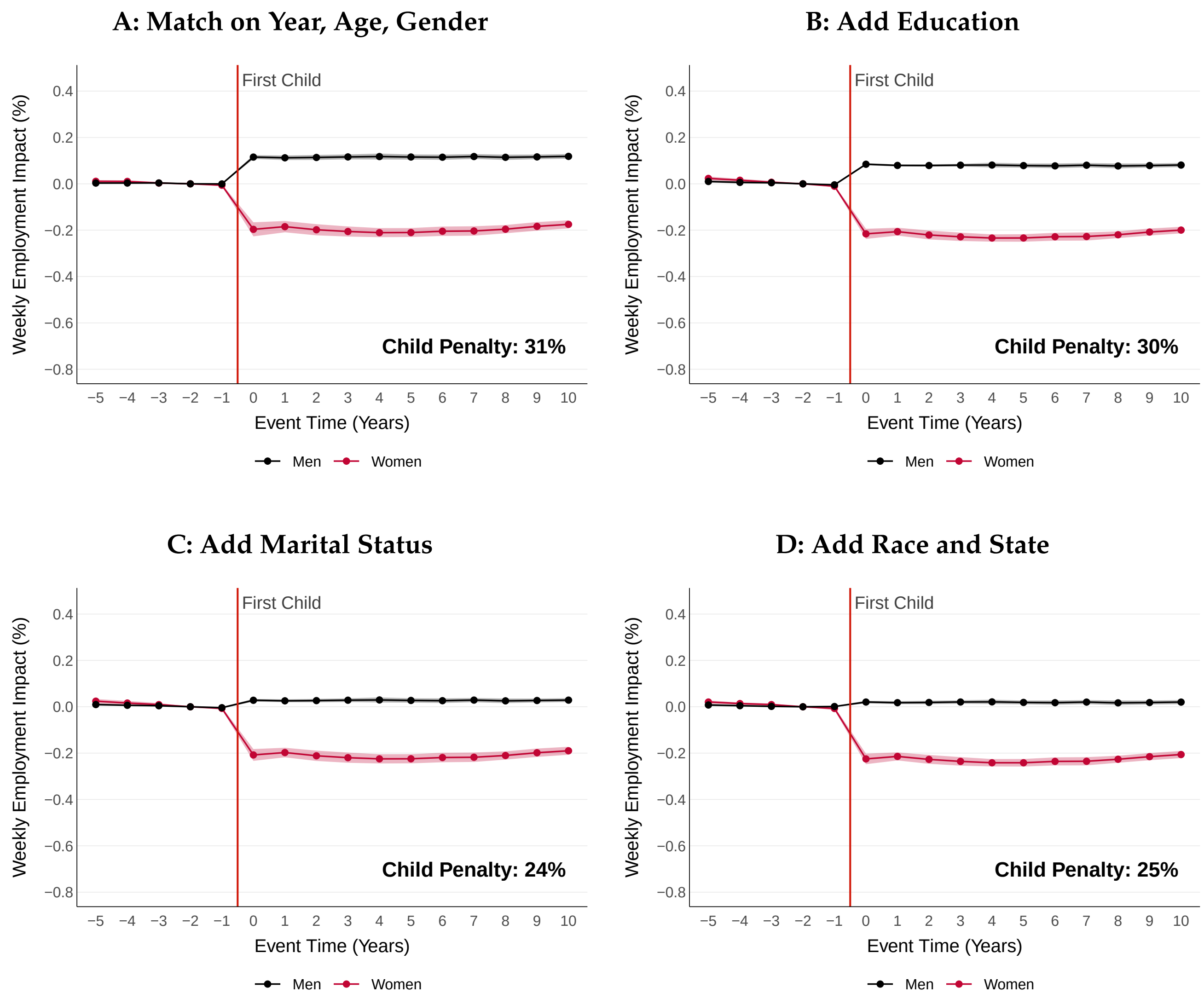
Figure A.1: Pseudo-Event Studies Under Different Matching Specifications

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 $t = -1$ and $t = 0$ 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.

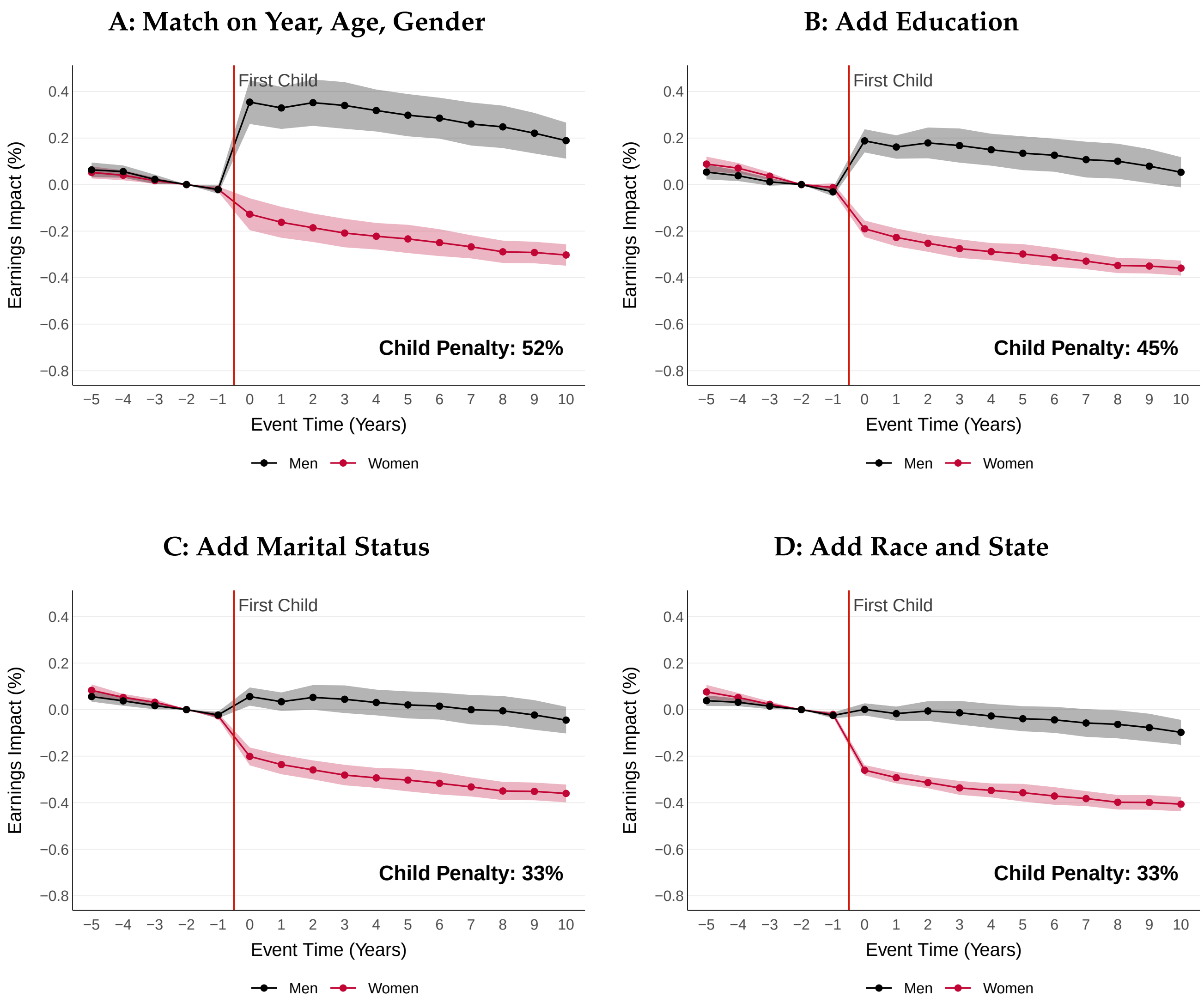
FIGURE A.2: PSEUDO-EVENT STUDIES UNDER DIFFERENT MATCHING SPECIFICATIONS
WEEKLY EMPLOYMENT



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 $t = -1$ and $t = 0$ 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

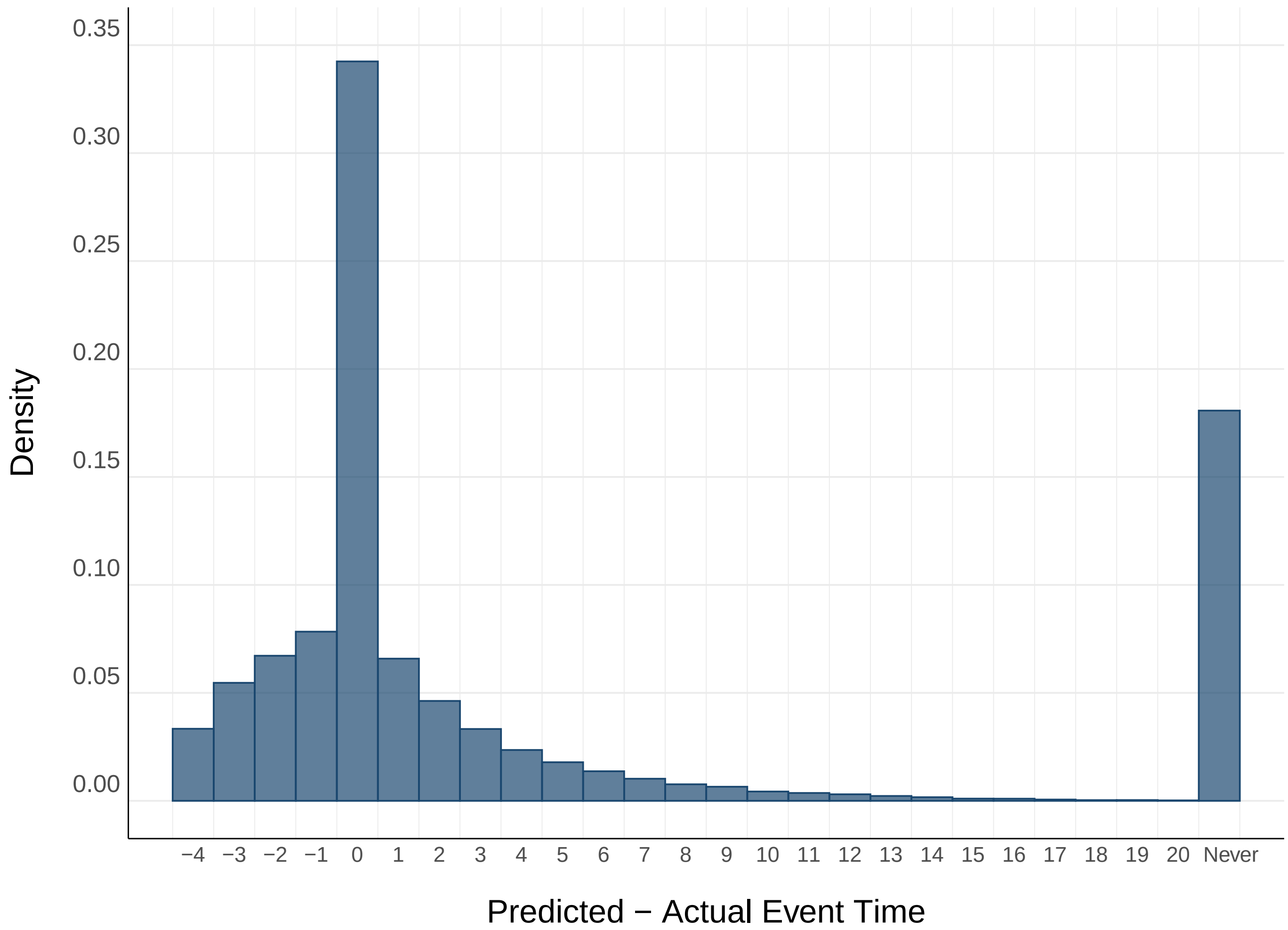
Earnings



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 $t = -1$ and $t = 0$ 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

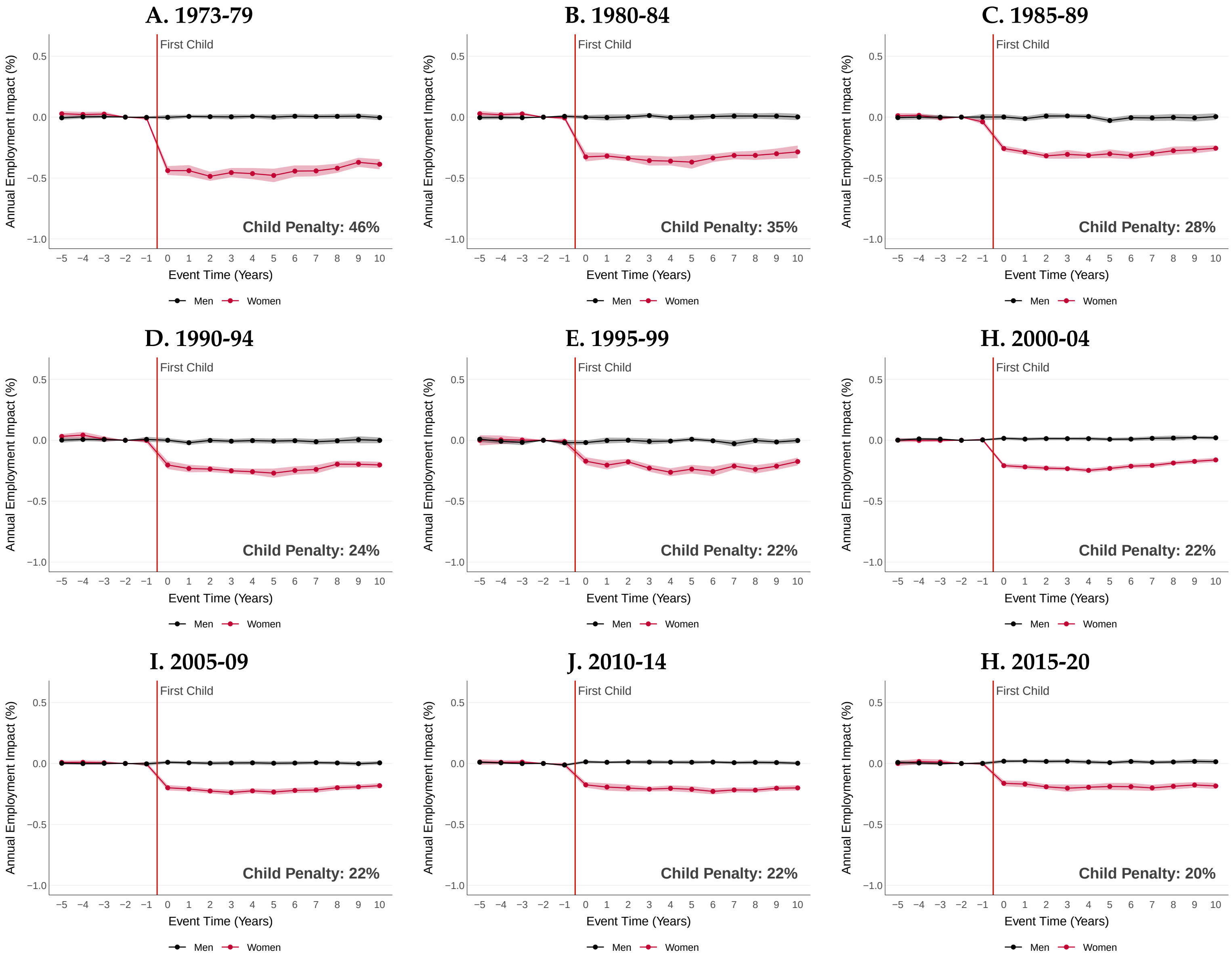
Predicted vs Actual Event Times Among Childless People



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 labelled "never" includes matched individuals (assigned to event times between -5 and -1) who never have children.

Figure A.5: Event Studies of First Child Birth Over Time

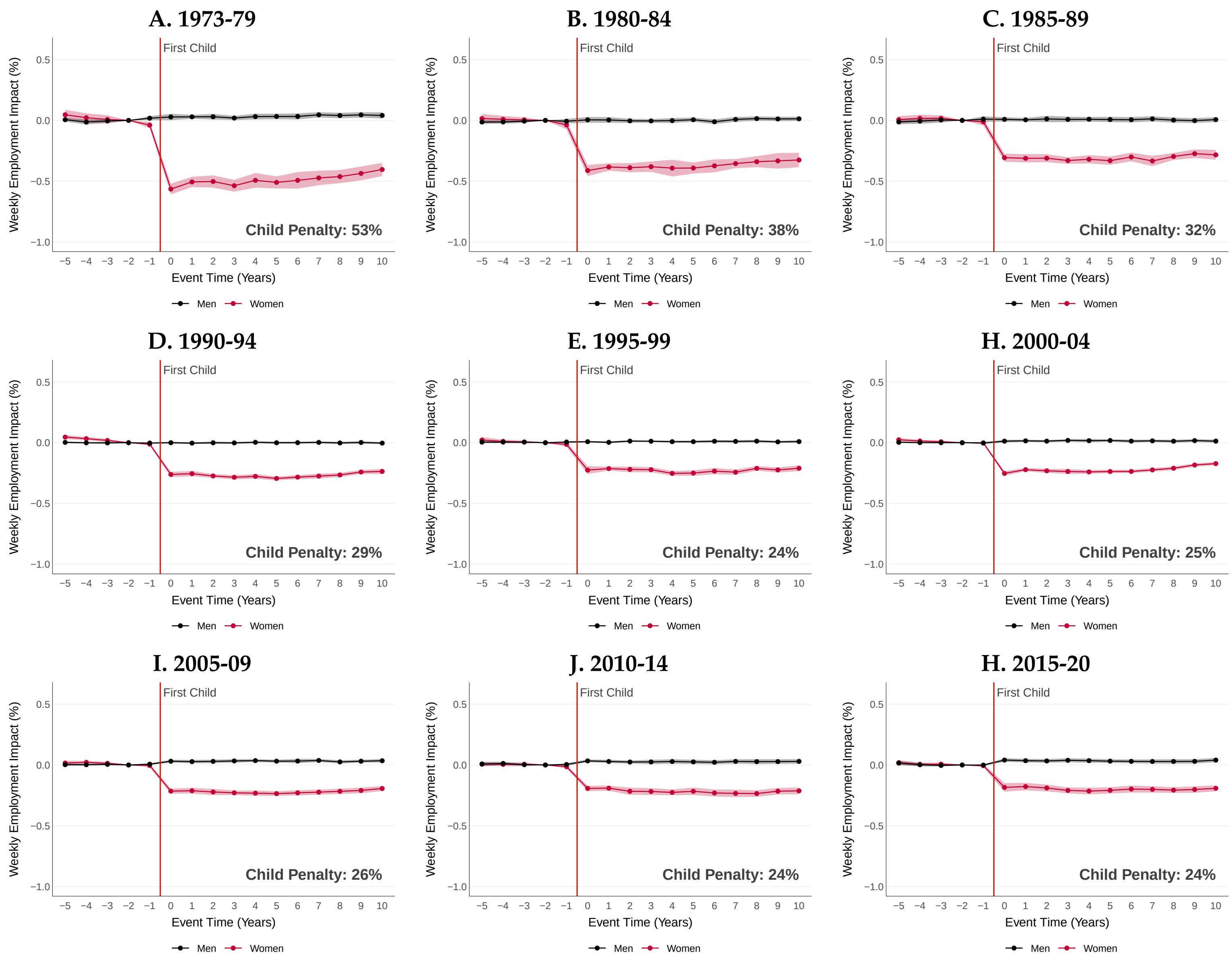
Annual Employment



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 (1) 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 surrogate 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 3) for the time period in question. The 95% confidence intervals are based on robust standard errors.

Figure A.6: Event Studies of First Child Birth Over Time

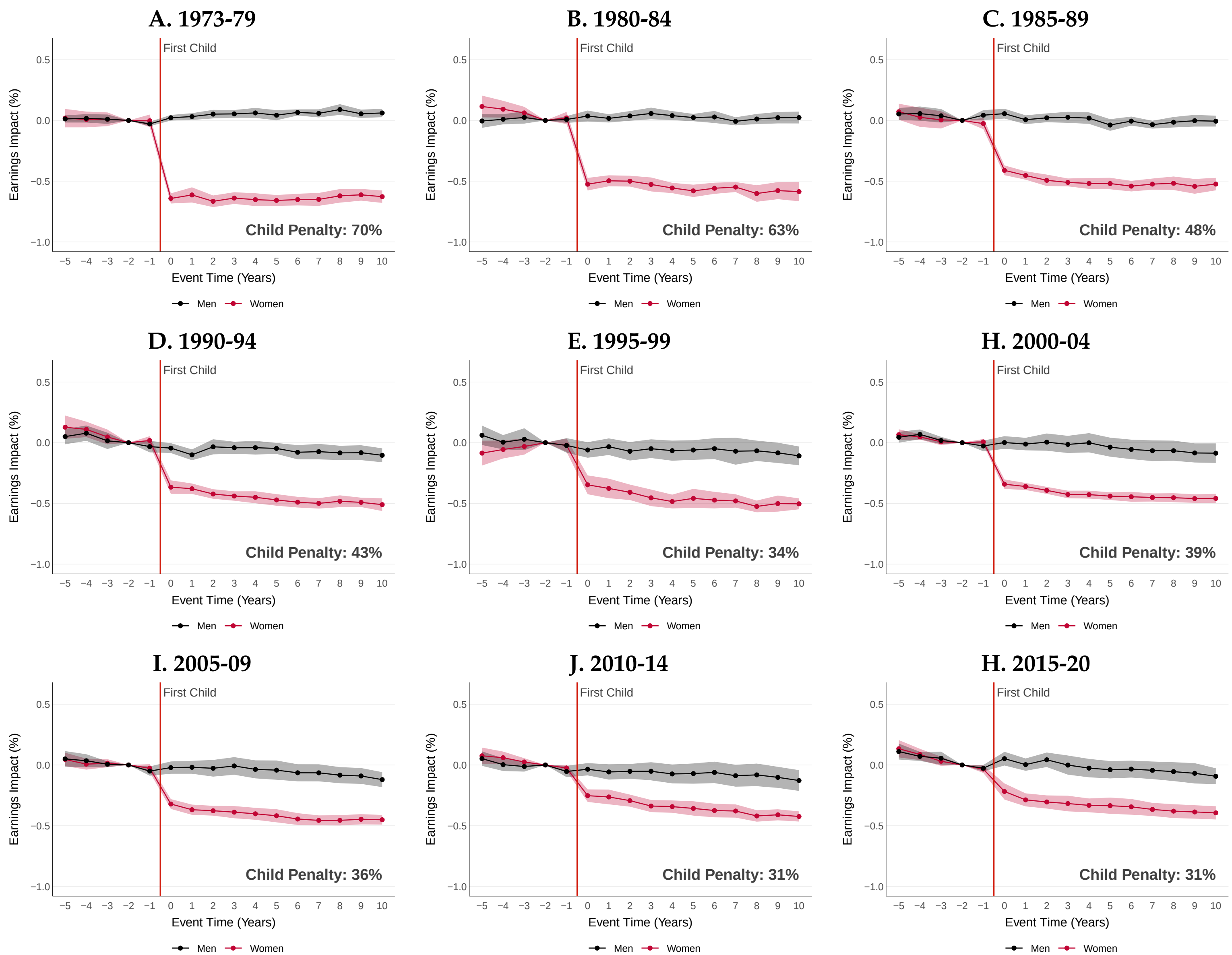
Weekly Employment



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 (1) 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 surrogate 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 3) for the time period in question. The 95% confidence intervals are based on robust standard errors.

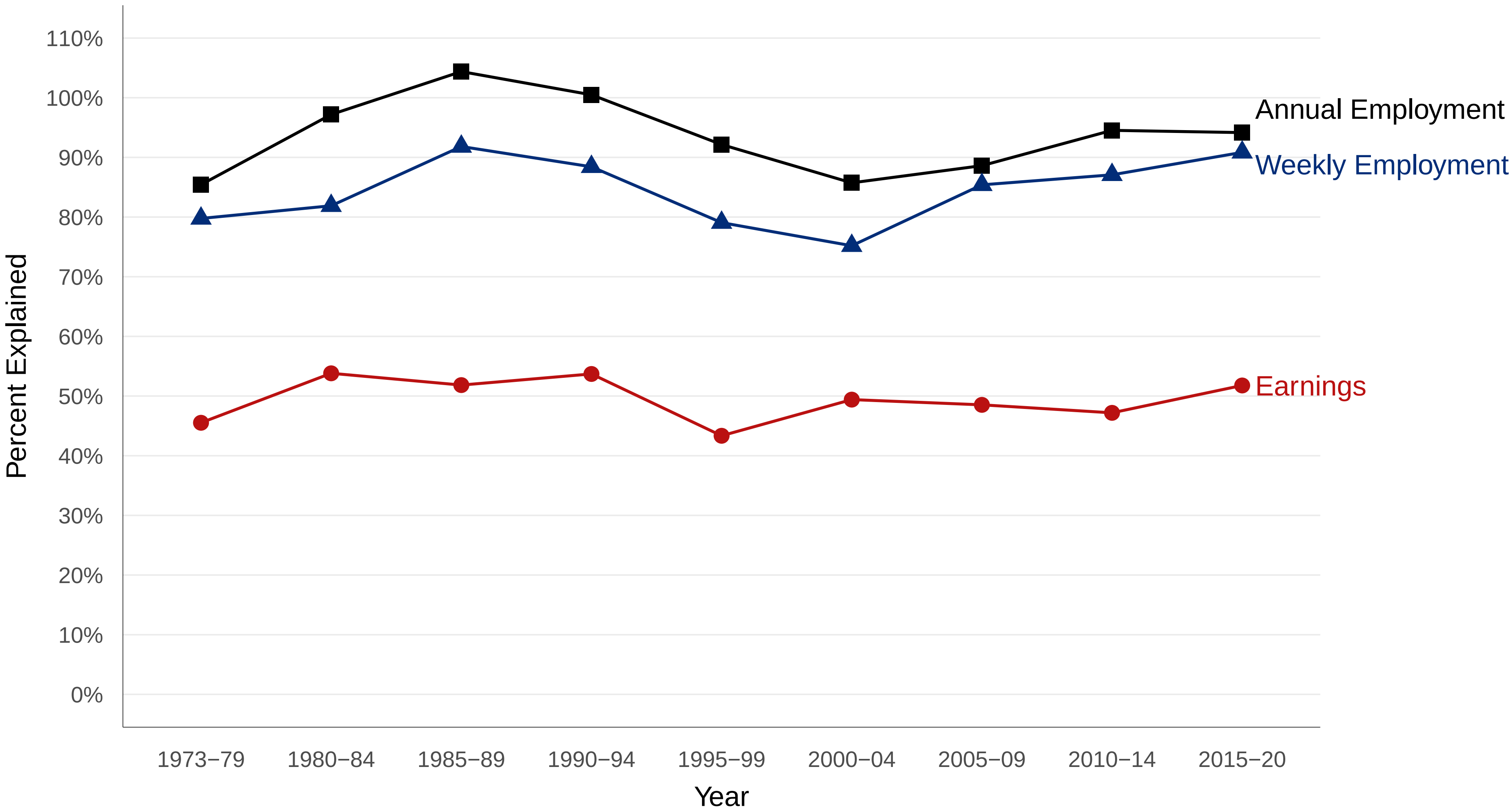
Figure A.7: Event Studies of First Child Birth Over Time

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 (1) 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 surrogate 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 3) for the time period in question. The 95% confidence intervals are based on robust standard errors.

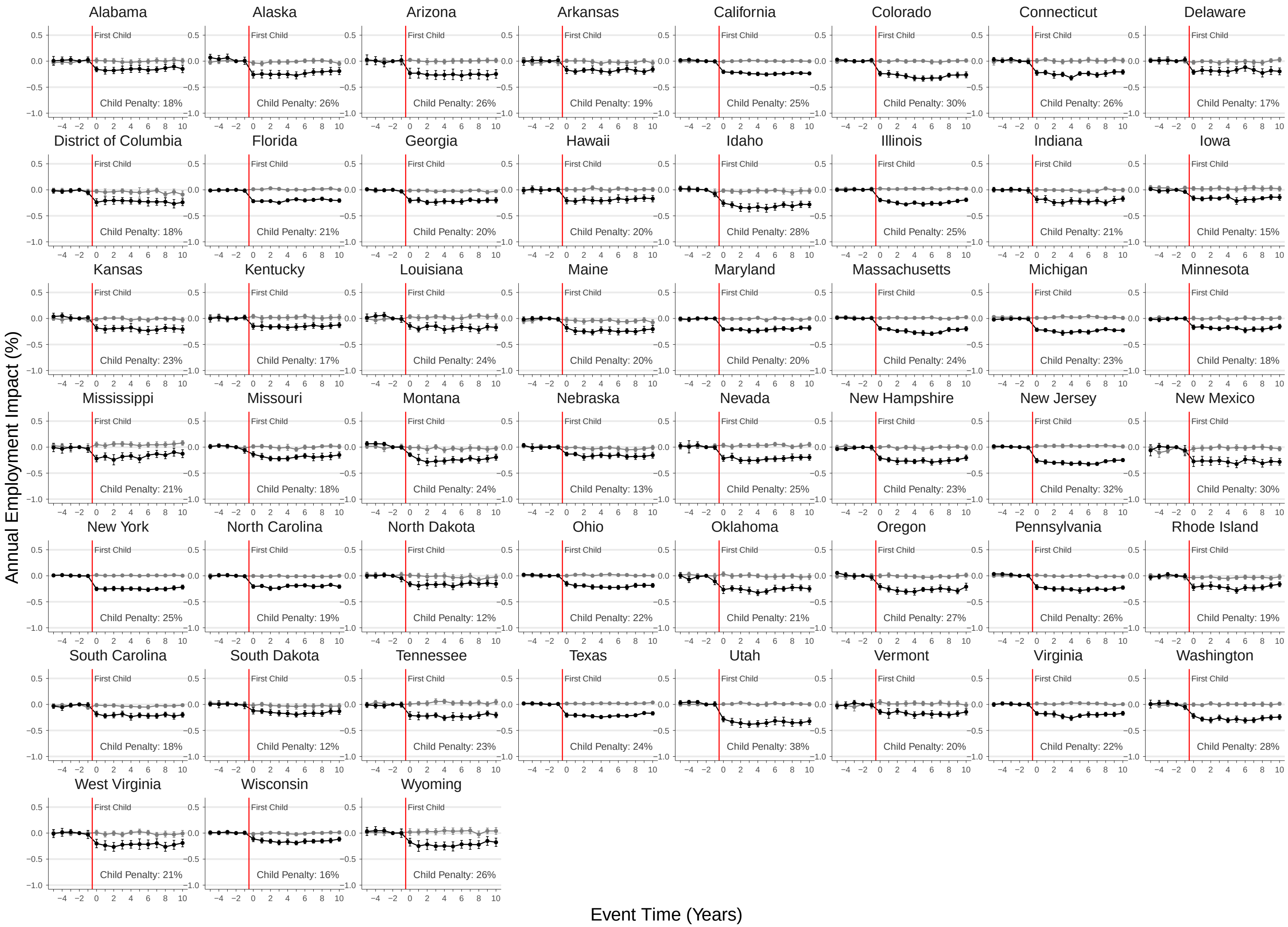
FIGURE A.8: FRACTION OF RAW GENDER GAPS EXPLAINED BY CHILD PENALTIES



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 3.

Figure A.9: 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 (1) with state dummies, estimating percentage impacts of child birth on men and women at each event time ($\hat{P}_t^m$ and $\hat{P}_t^w$) 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 (1) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

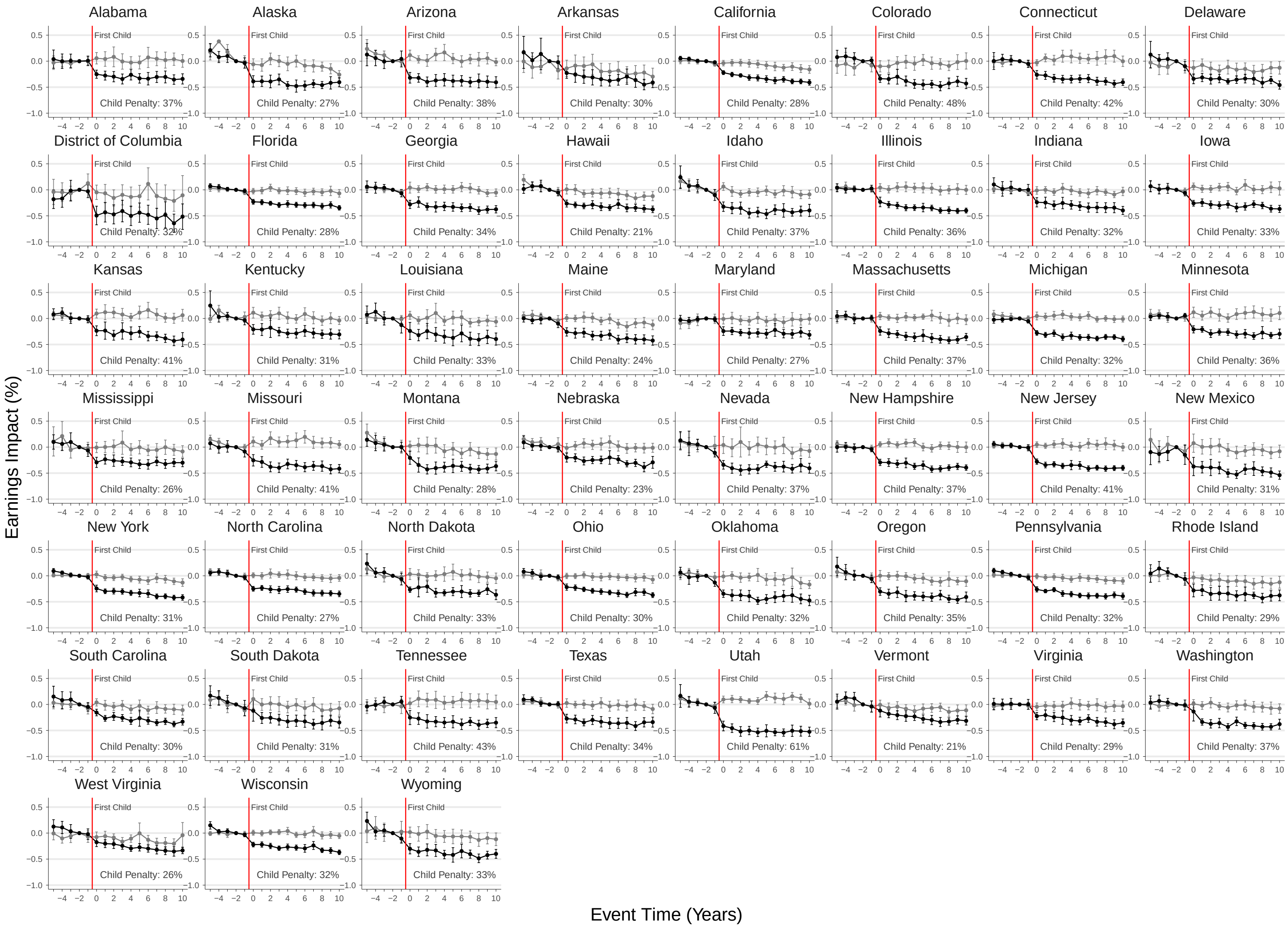
Figure A.10: Event Studies of First Child Birth Across States

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 (1) with state dummies, estimating percentage impacts of child birth on men and women at each event time ($\hat{P}_t^m$ and $\hat{P}_t^w$) 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 (1) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

Figure A.11: Event Studies of First Child Birth Across States

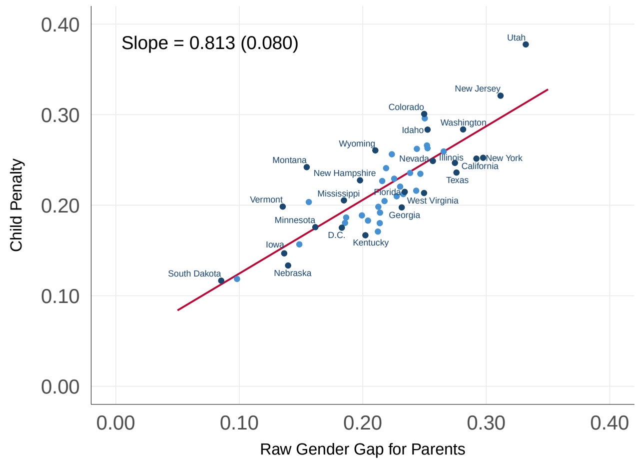
Earnings



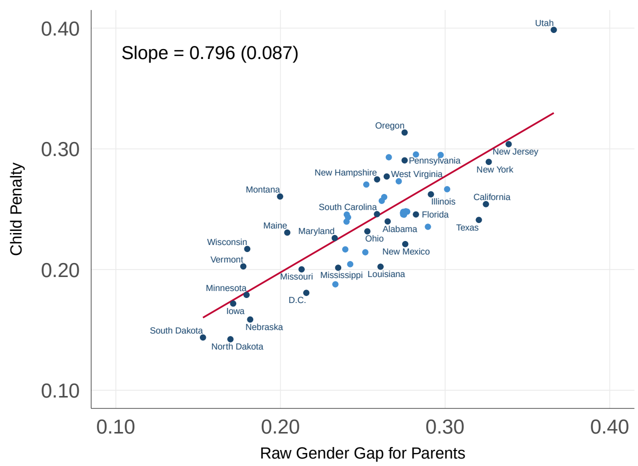
Notes: This figure shows event studies of first child birth in earnings for each of the 51 US states (including the federal district of D.C.). State-level event studies are constructed by interacting the event time dummies in equation (1) with state dummies, estimating percentage impacts of child birth on men and women at each event time ($\hat{P}_t^m$ and $\hat{P}_t^w$) 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 (1) are estimated at the level of census divisions. The 95% confidence intervals are based on robust standard errors.

FIGURE A.12: CHILD PENALTIES VS RAW GENDER GAPS ACROSS STATES

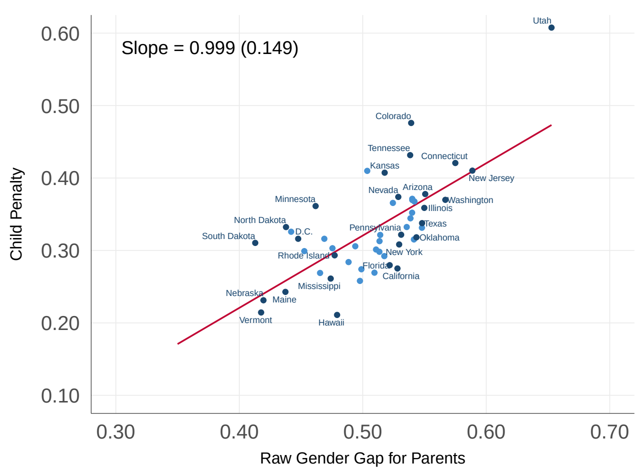
A. Annual Employment



B. Weekly Employment



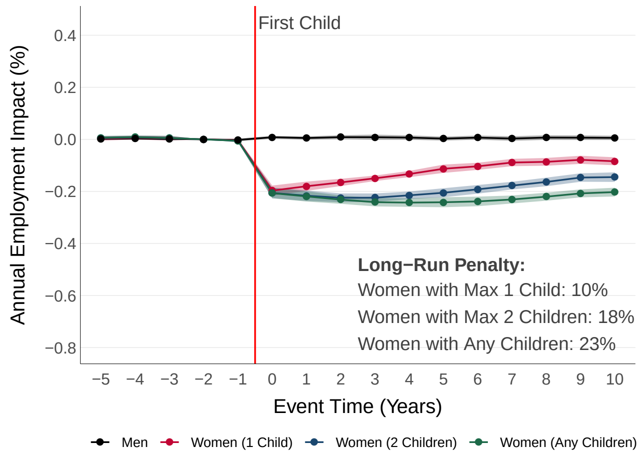
C. Earnings



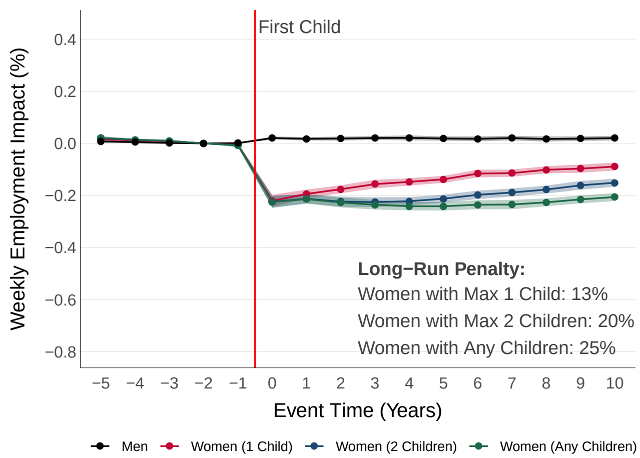
Notes: This figure provides scatterplots of child penalties against raw gender gaps for parents across states. Results are shown for each of the three labor market outcomes: Annual employment, weekly employment, and earnings. The raw gender gap is defined as the percentage difference between men and women with children, and the child penalty estimates for each outcome and state are shown in Figures A.9-A.11.

Figure A.13: Event Studies of First Child Birth by Fertility

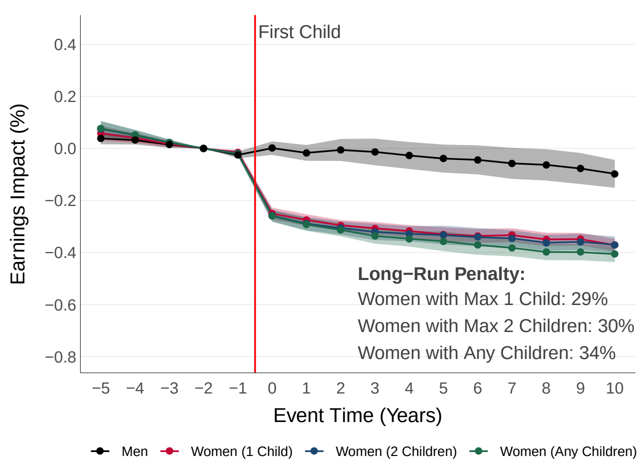
A. Annual Employment



B. Weekly Employment

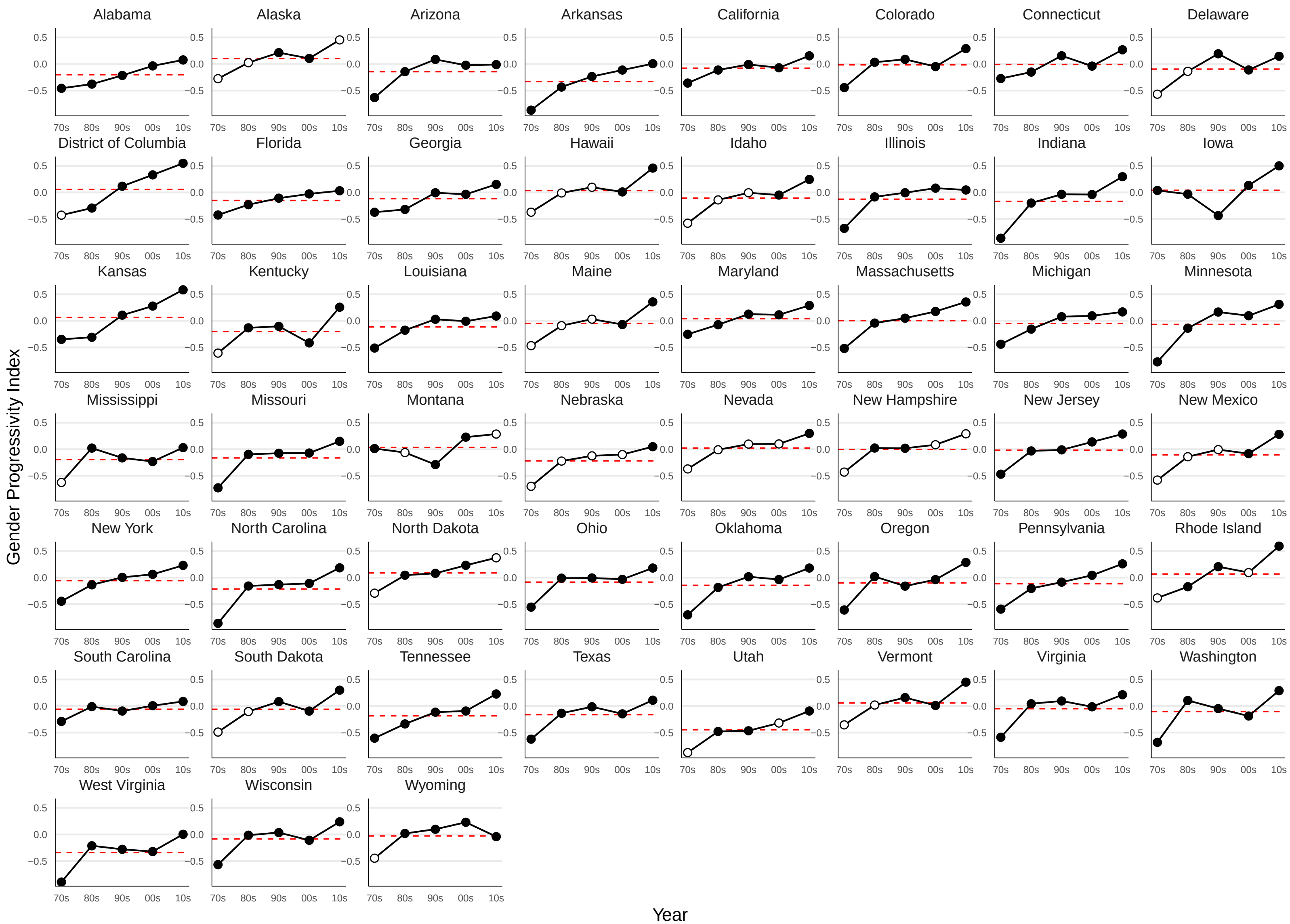


C. Earnings



Notes: This figure presents event studies of first child birth by female fertility. The figure is constructed by running specification (1) separately for different samples of women, conditioning on the maximum number of children at any given event time. Results are shown for women with at most one child, with at most 2 children, and with any number of children (baseline specification) for each of the three labor market outcomes. The long-run child penalty (over event times 5-10) is displayed for each outcome. Child penalties are increasing in the number of children and converge to the baseline specification that allows for any number of children. The 95% confidence intervals are based on robust standard errors.

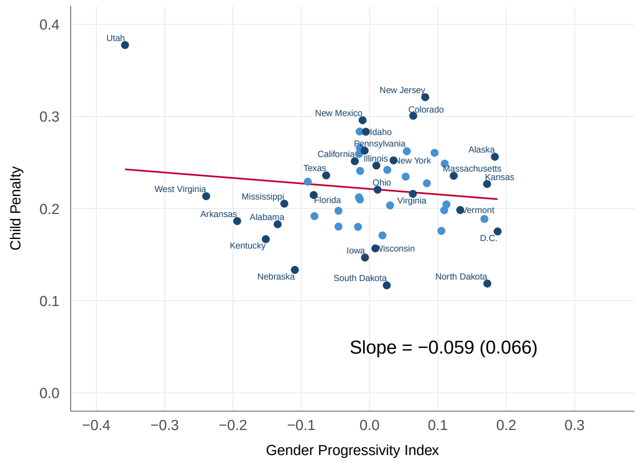
Figure A.14: Gender Progressivity Index by State and Time



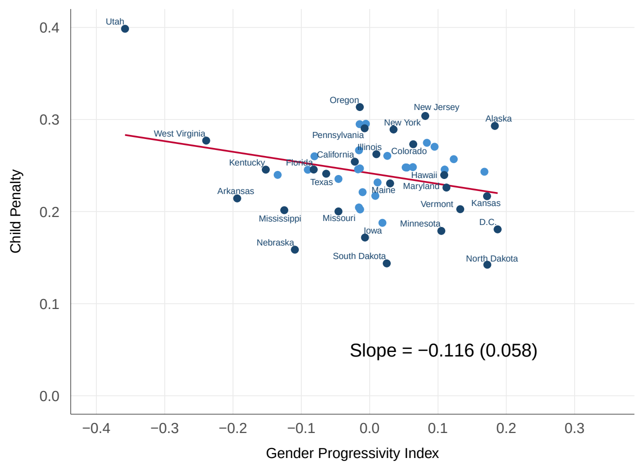
Notes: This figure presents time series of the Gender Progressivity Index (GPI) in each state over the last five decades. Using GSS data from 1972-2018, the index is calculated as the average standardized response to questions that elicit attitudes towards gender roles in families with children. The 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.15: Child Penalties vs Gender Progressivity Across States

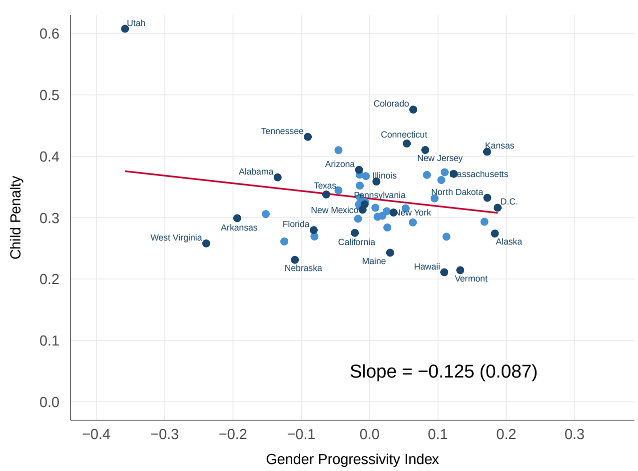
A. Annual Employment



B. Weekly Employment



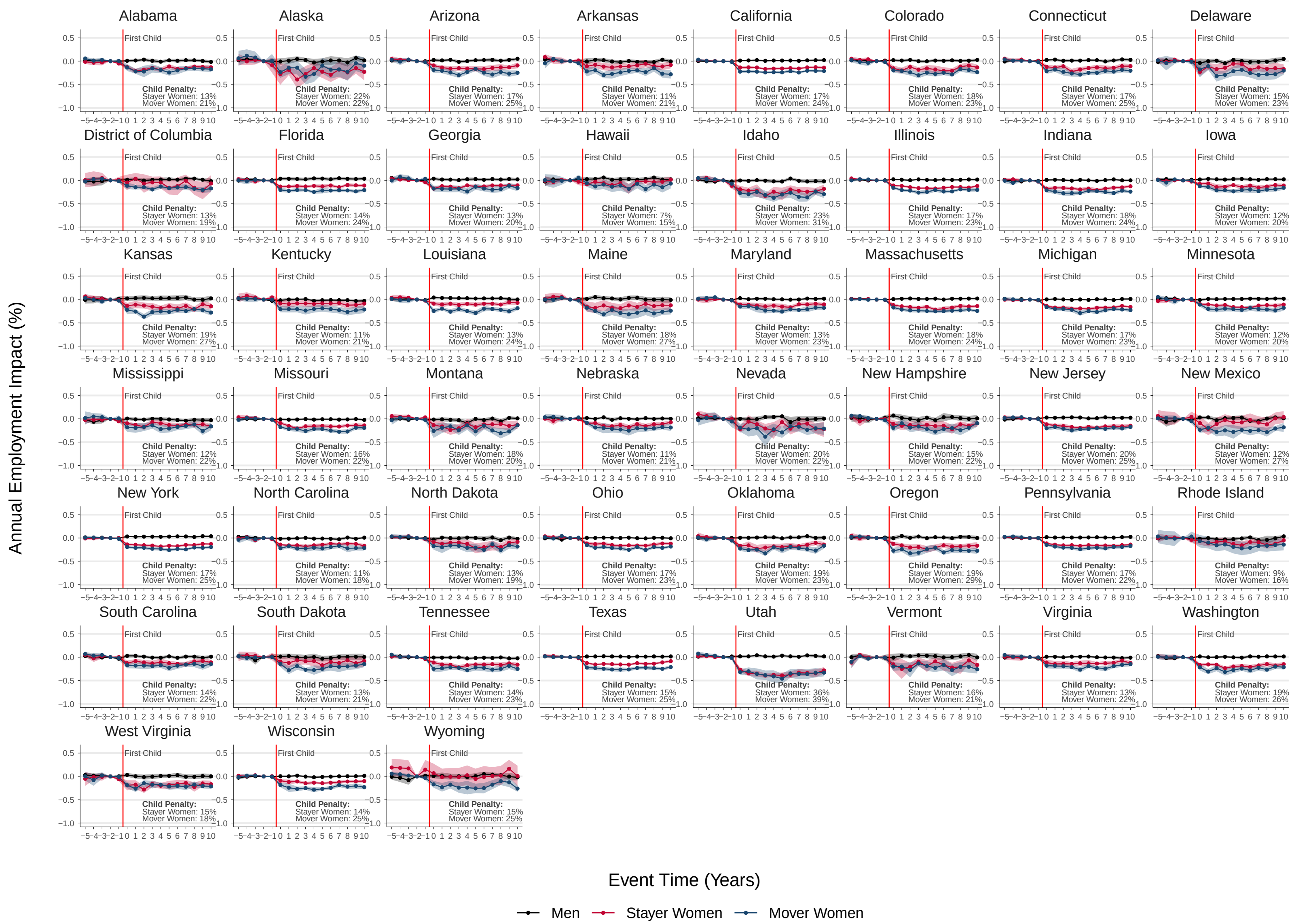
C. Earnings



Notes: This figure shows scatterplots of child penalties vs gender progressivity across states. The child penalty estimates are those presented in the heatmap of Figure 5, with all of the underlying event studies provided in Figures A.9-A.11. The construction of the Gender Progressivity Index (GPI) at the state-decade level is described in the notes to previous figures. In this figure, the GPI for each state is a weighted average of the state-decade GPIs, where each decade is weighted by its share of births in the pooled CPS and ACS data used to estimate child penalties. This ensures that the child penalty estimates and gender norms across states are directly comparable in terms of their average timing.

FIGURE A.16: EVENT STUDIES OF FIRST CHILD BIRTH FOR MOVERS VS STAYERS BY STATE OF BIRTH

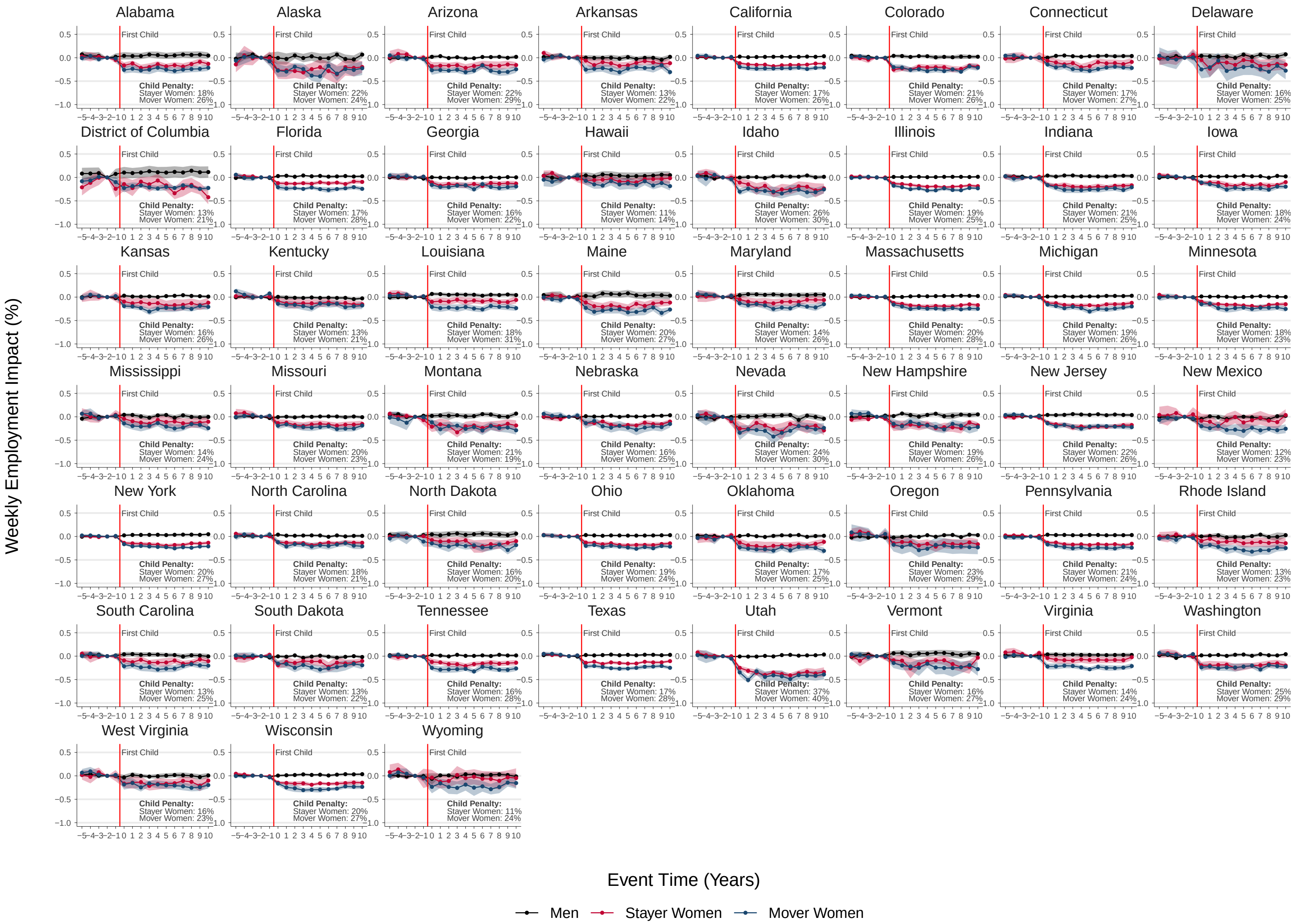
ANNUAL EMPLOYMENT



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 (1) 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.17: 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 (1) 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.

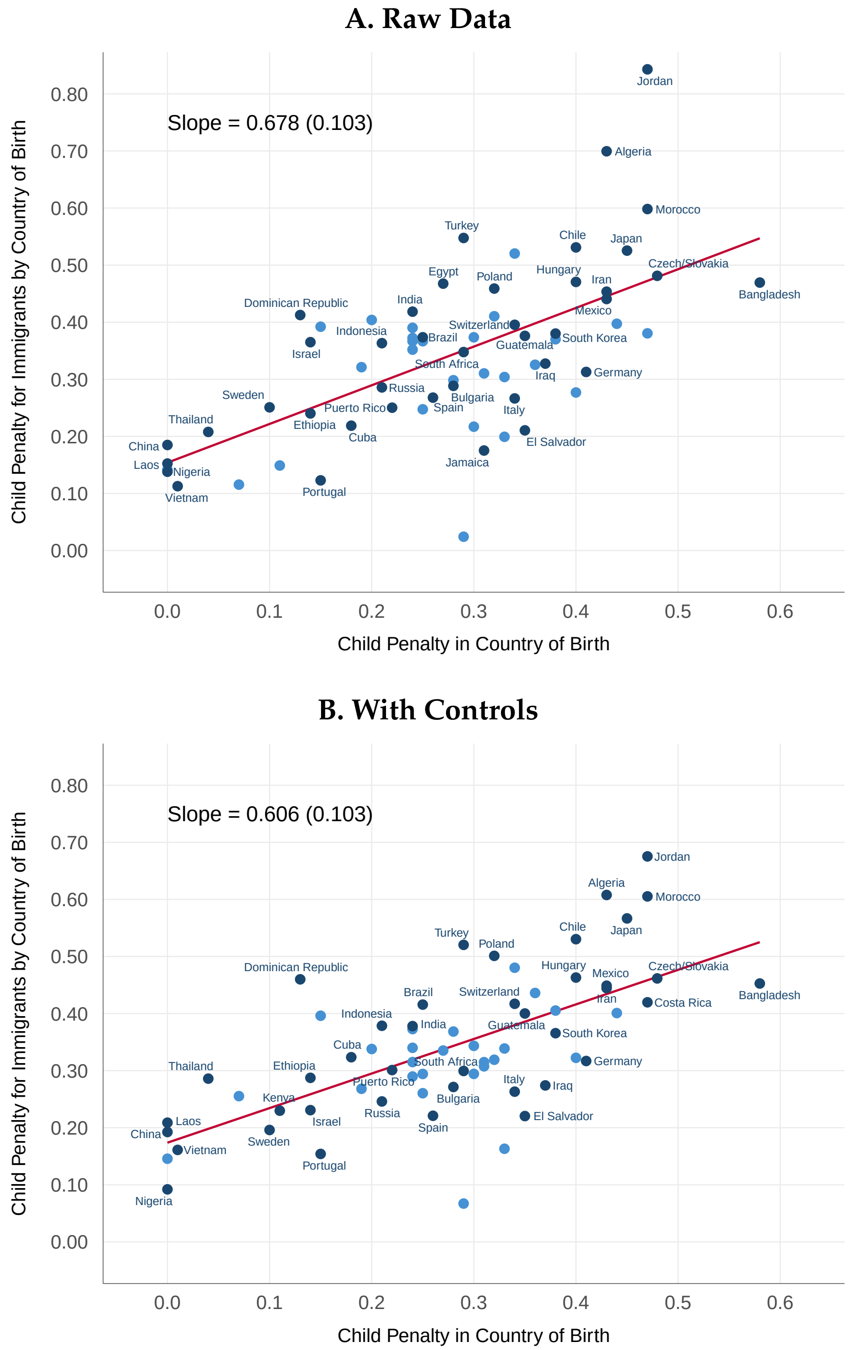
Figure A.18: Epidemiological Study of Foreign Immigrants

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 Mariante 2022). The event studies have been estimated by interacting the event time dummies in equation (1) with country-of-birth dummies. 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.19: EPIDEMIOLOGICAL STUDY OF FOREIGN IMMIGRANTS

CHILD PENALTIES FOR IMMIGRANTS VS CHILD PENALTIES IN COUNTRIES OF BIRTH

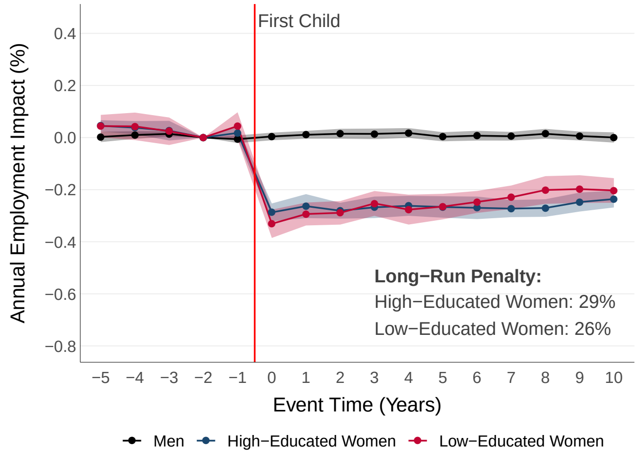


Notes: This figure presents scatterplots of child penalties for foreign-born immigrants against child penalties in country of birth. The underlying event studies for US immigrants are shown in Appendix Figure A.18 and the child penalties in country of birth are taken from Kleven, Landais, and Mariante (2022). 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.

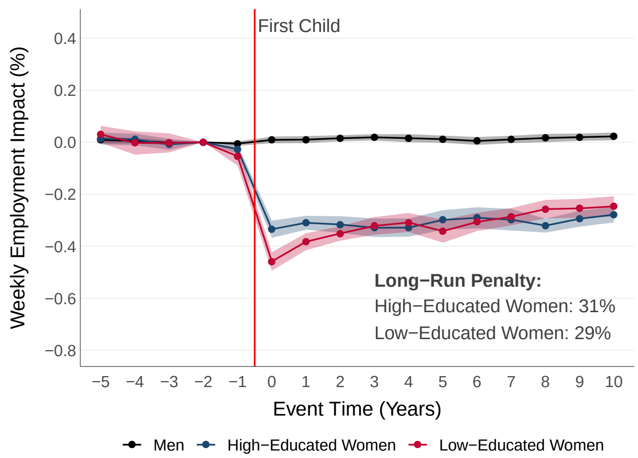
Figure A.20: Event Studies of First Child Birth by Education

Foreign Immigrants

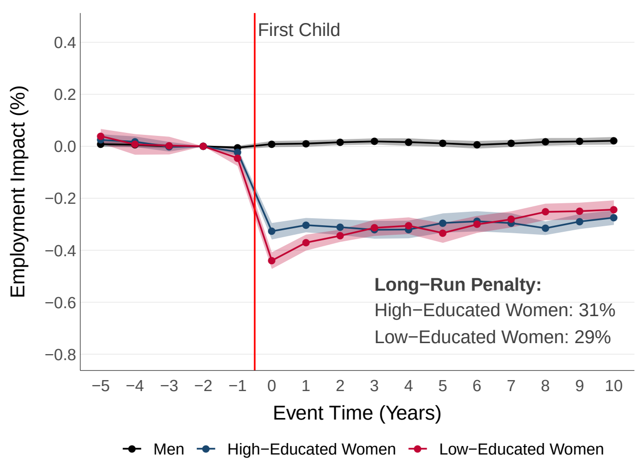
A. Annual Employment



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 Figure 7 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.