BACKGROUND NOTE 4

Nonparametric Pseudo-Event Study Designs: Estimating the Motherhood Earnings Penalty in El Salvador

Abstract

We propose a nonparametric approach to pseudo-event studies to examine motherhood earnings penalties. By combining nonparametric matching and pseudo-event studies, we address limitations highlighted in the literature, such as the lack of common support between men and women between periods and over time, and treatment assignment heterogeneity within cells of observables using inverse probability weighting. Our methodology also provides a more efficient pseudo-event estimator compared to the extant approach. We demonstrate the advantages of our methodology using data from El Salvador, where previous studies have found a comparatively high motherhood earnings penalty. Contrary to these findings, we find that the motherhood earnings penalty is lower than what standard estimates suggest.

JEL codes: D63, J13, J16, J22, J31

Keywords: Motherhood, labor outcomes, nonparametric, pseudo-panels, event study

I.

Introduction

Having children has important implications for various individual-level outcomes, especially for women. Empirical evidence suggests that the birth of the first child leads to a decline in labor earnings for women, a phenomenon referred to as the motherhood earnings penalty —which also extends to other labor market outcomes. In this note, we propose a methodology to estimate this earnings penalty using repeated cross sections, employing a combination of nonparametric matching and a pseudo-event study design, building on recent research by Kleven (2023), and apply it to El Salvador.

Our methodology presents advantages over current approaches. First, analyzing the earnings penalty relies on the Blinder-Oaxaca decomposition, which estimates the earnings penalty net of the effect of sociodemographic characteristics using a form of inverse probability weighting. However, Blinder-Oaxaca often underestimates the earnings penalty by not properly restricting comparisons to the common support of observable characteristics of men and women. Furthermore, as markets and societies change over time, the distribution of observable characteristics may also change, leading to a lack of common support when comparing cross sections over time.

Second, while the gender wage gap can be interpreted as an average treatment effect, the controls included in these parametric regressions are often post-treatment, as almost all observable characteristics used as controls are often a function of gender and birth choices. Our methodology circumvents these issues to some extent through the use of nonparametric matching: there is evidence that weighting by the inverse of the propensity score for the entire treatment history as a function of time-varying confounders, at the cell level, can improve 'all-else-equal' comparisons without introducing post-treatment bias to recover more structural estimates. Our methodology allows us to compute the difference in means within cells in the distribution of observables as an average of directed controlled effects restricted to the common support. Further, it allows us to address distributional variations due to differences in the distribution of observables between men and women as well as changes in sampling frames over time through proper reweighting. However, the size of the common support decreases with the number of observable characteristics included in the nonparametric matching procedure, that is, the curse of dimensionality. Thus, the researcher must balance internal and external validity concerns.

Finally, our matching estimates are additively separable, allowing for the decomposition of the 'naïve' motherhood earnings penalty into its component terms, including its causal effect. This enables researchers to investigate the heterogeneity of the penalty across the distribution of observables, with the standard caveat of the curse of dimensionality.

We apply our methodology to El Salvador for several reasons. First, despite notorious research on the impact of motherhood on women's labor outcomes over the past few decades, our understanding of the motherhood earnings penalty in developing countries is still limited. For instance, in Latin America and the Caribbean, previous studies encompass some of their most developed members (for example, Argentina, Brazil, Chile, Colombia, Mexico, and Peru), but the estimated motherhood penalty shows high levels of variability due to estimation strategies and data quality. More recently,

—

1. Oficina Nacional de Estadística y Censos (ONEC, 2023a).

a report applied pseudo-event study design methodology developed by Kleven (2023) and estimated significant earnings penalties for motherhood as well as relatively similar penalties across countries. Curiously, their estimates differ substantially from similar estimates to other reports by about 10 percentage points in some cases. The problem is that the samples used by the latter differ from those used by the former, affecting findings non-trivially.

Second, the effects of motherhood on labor market outcomes in developing countries may differ from those in developed countries, given higher levels of labor market informality, lower coverage of childcare, less generous leave, and more conservative social norms. For instance, the share of informal employment (not entitled to retirement benefits linked to their jobs) is 8 percent in Uruguay, 13 percent in Chile, 46 percent in Peru, and 57 percent in Mexico. According to household surveys for El Salvador, the level of labor informality is 56 percent. Likewise, the factors contributing to this earnings gap are likely to differ from those found in developed countries. The size of labor market responses to motherhood is larger in societies with more conservative social norms or weaker work-life balance policies. Therefore, it is important to carry out case studies to understand the idiosyncrasies that may be driving these earnings penalties given the local socioeconomic context, even considering the importance of comparative assessments.

Third, a recent report shows some peculiar results for El Salvador compared to all other countries, making it an interesting case in and of itself. The effect of motherhood on earnings is the smallest of all countries in the sample by a substantial amount (–34 percent), as the next smallest effect is double the one estimated for El Salvador (–12 percent). It is difficult to understand why this penalty is so low in El Salvador compared to other countries within the same region, not to mention that this penalty is comparable to some developed Western European countries and the United States and Canada. However, El Salvador ranks 16 out of 20 countries in the index of Human Development in Latin America, and it is similar to other Latin American countries in various gender outcomes compared to more developed countries (Appendix A).

Our results show that, in general, the earnings gap is statistically comparable to that obtained by the standard approach, except when we take into account the differential distribution of women in part-time jobs compared to men, after childbirth. Our procedure indicates that the five-year motherhood earnings penalty is about 3 percent in favor of men compared to about 26 percent when these distributional differences are not taken into account. While in the long run the penalty is higher, we find that it is about 7 percent in favor of men—about one-third of the penalty we estimate using the standard parametric approach.

Overall, our note contributes to the literature on gender earnings gaps, especially those associated with the implications of motherhood on this gap—the gender earnings penalty. On the one hand, we propose a nonparametric pseudo-event study approach to estimate the earnings gap that addresses empirical shortcomings attributable to variations in the distribution of observable characteristics over time and on treatment assignment. We show that this bias can be substantive. On the other hand, we show that our methodology may allow researchers to analyze the motherhood penalty with more depth through the nonparametric decomposition that emerges as a byproduct of our approach. Finally, we share Kleven's (2023) optimism regarding the use of pseudo-event study designs in studying relevant questions in labor economics by providing a discussion on how our nonparametric pseudo-event studies can be used to investigate other socioeconomic outcomes.

II.

Methodology

Estimating the impact of fertility decisions on earning outcomes between men and women—that is, the child penalty for women—is a difficult task. The decision to have children and adjust preferences and labor market choices is endogenous to the circumstances of the individual and the household as well as the gender roles at play. For instance, some women may opt for less education or family-friendly career tracks, knowing they will eventually have many children. Similarly, in households where gender roles are more traditional, women are more likely to take on lower-paid jobs that provide flexibility for child-rearing, at the expense of lower earnings. Likewise, in households where the (male) partner has a high-paying stable job are also households, it is easier to address the trade-off of lower earnings from the female spouse—regardless of whether this comes from spending less time on career advancement, accepting lower earnings jobs, or labor market discrimination. In other cases, women may need to accept lower earnings out of necessity, such as in low-income, single-parent households.

All in all, the aforementioned sources of endogeneity imply that there are both compounded self-selection effects and potentially unobservable determinants of lower earnings for women. These unobservables extend beyond gender disparities (all else being equal), insofar as the decision to have children can exacerbate gender biases. Therefore, any 'naïve' analysis of the gender penalty for having children is subject to substantial biases that are difficult to parse out.

To ground our methodology, let us assume we have panel data and that the decision to have a child for the first time is as good as exogenous. Thus, let us consider the comparison that estimates the effect of having a firstborn on earnings:

$$Y_{it} = \delta F_i \times C_{it} + X_{it} + \mu_i + \gamma_t + \varepsilon_{it}$$

where Y_{it} denotes labor earnings; C_{IT} is an indicator that takes the value of one for individual i after the birth of their first child; F_i takes the value of one if individual i is a female, zero otherwise; X_{it} is a vector of pre-child-birth observable characteristics (for example, hours of work and occupation); μ_i is a vector of individual-level fixed effects that capture time-invariant unobservable covariates; and γ_t captures any non-monotonic trends common to all individuals in the sample. The parameter of interest here is δ, which corresponds to the child penalty in terms of earnings that women experience compared to men.

To further ground our exposition, we refer to having a child as the treatment and gender as the level of treatment uptake. The standard hypothesis is that under random assignment of having a firstborn, women are much more affected by this event compared to men. Therefore, the earnings penalty for having one child for women can be estimated δ in the equation above:

$$\delta = E(Y_{it} \mid F_i = 1, C_{it} = 1) - E(Y_{it} \mid F_i = 0, C_{it} = 1)$$
$$- [E(Y_{it} \mid F_i = 1, C_{it} = 0) - E(Y_{it} \mid F_i = 0, C_{it} = 0)]$$

δ is the differential effect of having children between women and men, that is, the earnings penalty. Note how that this design is essentially a difference-in-differences (DID) design.

The literature on DID indicates that δ is identified if we have parallel trends before the treatment. This 'parallel trends' assumption implies that in the counterfactual where those who have children did not, they must have had the same trajectories in average earnings over time as those who did. The recent literature on two-way fixed effects (TWFE) has shown, however, that meeting the parallel trends assumption may be insufficient. First, the parallel trends assumption states that conditional on observable covariates, the trends between the treatment and control group are the same. However, if we observe parallel trends with one set of covariates, it does not mean we would observe the same with another. Furthermore, traditional methods are not necessarily heterogeneity robust because they do not address either the heterogeneity in treatment assignment across groups or the heterogeneity in treatment uptake between groups. This is problematic because treatment effects are essentially weighted averages from effective multiple regression weights.

These problems are exacerbated when we are using repeated cross sections, which are often used to study gender outcomes, as we do not observe the same observations over time. Therefore, observations may not even be comparable before and after the treatment, which is an essential assumption in DID designs. Indeed, let us reorganize the expression for δ above:

$$\delta = E(Y_{it} \mid F_i = 1, C_{it} = 1) - E(Y_{it} \mid F_i = 1, C_{it} = 0)$$
$$- [E(Y_{it} \mid F_i = 0, C_{it} = 1) - E(Y_{it} \mid F_i = 0, C_{it} = 0)]$$

An equivalent interpretation for δ is the differential effect of having a child for both men and women. This equivalence emerges if either the treatment or treatment uptake is exogenous as it is well known we only need an exogenous component in an interaction term to interpret it as causal. Conversely, this means that observations being analyzed must be similar before and after the event—which is satisfied by construction in panel data but not in repeated cross sections as we do not observe the same observations over time. In other words, we could observe a phenomenon akin to self-selection after the event.

To address the shortcomings, we could use an event study design, as these designs help us ensuring covariate balance before and after an event (for example, having a child), so that we know we are making 'all-else-equal' comparisons. Often this implies restricting the time bandwidth around the time of the event, 'right before and right after', as those observations far away from the time of the event are likely to be different from each other. However, we cannot control for individual-level time-invariant covariates when using cross sections.

To circumvent these problems, one could use pseudo-panels to obtain consistent estimators from cohort aggregates when we have repeated cross sections. These cohorts are equivalent to having synthetic individuals. To obtain consistent estimators from a pseudo-panel, grouping variables should not present missing values for any individual in the sample, should not vary over time, and should be exogenous and relevant. Further, it is important that the number of cohorts is large enough to avoid small sample size problems, that cohort sizes are large enough to avoid measurement error problems, and that cohorts minimize within-cohort heterogeneity while maximizing between-cohort heterogeneity. Kleven (2023) combines this idea with that of an event study design to circumvent some of the problems mentioned above in a pseudo-event study design.

For estimating the motherhood penalty, the pseudo-event study design aims to exploit an event study around the birth of the first child, indexed as event time $t = 0$. Then, each person is matched at event time $t = 0$ to a childless person n years younger and thus n years before, using a set of observable demographic characteristics. This matched observation provides a surrogate for those observations measured at $t < 0$ and similarly for $t > 0$. To do this, we need to assume that the decision time of having a child is as good as exogenous conditional on observables and conditional on earnings. The problem is that the decision to have children is not quasi-random in this sense. The decision to only have one child, or have more, depends on the child history and may potentially depend on past realizations of the outcome, as the demand for children is a dynamic process. In general, we may assume that people have children when certain conditions are met, like having a certain level of earnings or job stability —although in some cases unwanted pregnancies, losing a child, or other similar sources of variation can be considered as good as random.

While the aforementioned problems may be captured using a structural approach or addressed to some extent using the new developments in two-way-fixed-effects estimators, we believe that there is much to gain from using a pseudo-event design as it allows us to create synthetic surrogates. This follows closely the idea behind synthetic control matching. But one important caveat is that we must consider that sampling frames can change substantially over time as labor markets and demographics change. Further, men and women must be comparable at every point in time to correctly control for synthetic-individual and time-fixed effects. However, women and men have dissimilar observable characteristics, and these differences change over time. This is relevant because we may have important problems of a lack of common support both along the gender and time dimensions. Likewise, observed treatment assignment may vary with these characteristics, over time. As a result, it is hard to substantiate the assumptions that are necessary to estimate the motherhood penalty using the extant approach. Further, this concern becomes more worrisome when we consider that sampling frames adapt over the years to reflect the changing socioeconomic conditions (for example, as new census data emerge and are used to improve sampling frames for household surveys), and it is important that we adjust the sampling weights to account for variations in observables over time owing to sampling design.

Another important problem that often emerges is that when we control for sociodemographic characteristics that are affected by gender, we run the risk of inducing post-treatment bias in the parametric estimator. While one may consider using (de)mediation analysis to address this concern, there has been much debate on the merits of using (de)mediation to address these issues as opposed to using controlled direct effects, given the former's strong sequential incomparability assumptions. The latter method, in contrast, entails comparing only individuals within cells in the distribution of observable mediators but does not indicate how to aggregate these estimands when the objective is to hold numerous mediators, all else being equal.

III.

The nonparametric pseudo-event study design

We resort to a nonparametric matching technique and provide formalization and evidence for the algorithm in the context of pseudo-event study designs. The objective of nonparametric matching is to compare individuals with the exact same combination of observable characteristics through a re-weighting process that considers: (i) differences in the distribution of observable characteristics between men and women and (ii) differences in the common support of observable characteristics between these groups. We can adjust these differences in observables within a year and over time to address the concerns laid out above.

Let X be the n-dimensional vector of individuals' characteristics; let $F^M(\cdot)$ and $F^F(\cdot)$ denote the conditional cumulative distribution functions of individuals' characteristics X; and let $dF^M(\cdot)$ and $dF^F(\cdot)$ denote their corresponding probability measures. $\mu^F(S)$ denotes the probability measure of the set S under the distribution $F^F(\cdot)$, that is, $\mu^F(S) = \int dF^F(x)$ and analogously $\mu^M(S) = \int dF^M(x)$. Thus, for instance, $E[Y \mid M, X] = g^M(x)$ and $E[Y \mid F, X] = g^F(x)$, and $E[Y \mid M] = \int_{S^M} g^M(x)dF^M(x)$, and $E[Y \mid F] = \int_{S^F} g^F(x)dF^F(x)$, where S^M denotes the support of the distribution of characteristics for men and S^F the support of the distribution of characteristics for women.

We express the earnings penalty for having one child for women as

$$\delta = \int_{S^M} g^M(x, c=1)dF^M(x, c=1) - \int_{S^F} g^F(x,c=1)dF^F(x, c=1)$$
$$- \left[\int_{S^M} g^M(x, c=0)dF^M(x, c=0) - \int_{S^F} g^F(x,c=0)dF^F(x, c=0) \right]$$

where $t \in x$. Let us denote the previous expression by

$$\delta = \delta_1 - \delta_0.$$

For each $\delta_{c'}$, the support of the distribution of characteristics for women, S^F, is different from the support of the distribution of characteristics for men, S^M. Each integral is split over its respective domain into two parts: within the intersection and out of the common support.

$$\delta_c = \left[\int_{\overline{S^F} \cap S^M} g^M(x, c)dF^M(x, c) + \int_{S^M \cap S^F} g^M(x, c)dF^M(x, c) \right]$$

Since the measures $dF^M(\cdot)$ and $dF^F(\cdot)$ are identically zero out of their respective supports (by definition), the domains outside of the common support can be extended to $\overline{S^F}$ and $\overline{S^M}$ respectively, without affecting their corresponding values. Also, every integral can be adequately rescaled to obtain expressions involving expected values of $g^M(x)$ and $g^{MF}(x)$, conditional on their respective partitioned domains, as shown below:

$$\delta_c = \int_{S^M \cap S^F} g^M \left[g^M(x, c) - g^F(x, c) \right] \frac{dF^F(x, c)}{\mu^F(S^M)}$$
$$+ \int_{S^M \cap S^F} g^M(x, c) \left[\frac{dF^M}{\mu^M(S^F)} - \frac{dF^F}{\mu^F(S^M)} \right](x, c)$$
$$+ \left[\int_{\overline{S^F}} g^M(x, c) \frac{dF^M(x, c)}{\mu^M(S^F)} - \int_{S^F} g^M(x, c) \frac{dF^M(x, c)}{\mu^M(S^F)} \right] \mu^M(\overline{S^F})$$
$$+ \left[\int_{S^M} g^F(x) \frac{dF^F(x, c)}{\mu^F(S^M)} - \int_{\overline{S^M}} g^F(x, c) \frac{dF^F(x, c)}{\mu^F(S^F)} \right] \mu^F(\overline{S^M})$$

which we denote by

$$\delta_K = \delta_k^R + \delta_k^X + \delta_k^M + \delta_k^F .$$

$\delta_k^M (\delta_k^F)$ is the part of the gender gap that would disappear if there are no men (no women) with combinations of characteristics that remain entirely unmatched by women (men); δ_k^X is the part of the gap that can be explained by differences in the distribution of characteristics of men and women over the common support; δ_k^R corresponds to the part which cannot be attributed to differences in individuals' characteristics, and thus it is the part of the gap that remains after controlling for differences in the distribution of observable characteristics, restricting ourselves to the common support.

Let us focus on the latter component described above and rewrite the earnings penalty of interest, focusing on the remainder part:

$$\delta^R = \delta_1^R - \delta_0^R,$$

which is equivalent to

$$\delta^{\mathrm{R}} = \int_{S^M \cap S^F} \left[g^M(x; c=1) - g^F(x, c=1) \right] \frac{dF^F(x; c=1)}{\mu^F(S^M)} - \int_{S^M \cap S^F} \left[g^M(x; c=0) - g^F(x; c=0) \right] \frac{dF^F(x; c=0)}{\mu^F(S^M)}$$

As pointed out above, the distribution of observations in the common support of observable characteristics can change over time due to sociodemographic changes. Hence, the regression above may not compare the same individuals over time, which is essential for estimating δ^R. To address this, we proceed as follows: First, we define a base category whereby there is an event at time $t = 0$, characterized by being interviewed in the same birth year as the firstborn. Thus if $t < 0$ individuals do not have a child and if $t \geq 0$, they have at least one child. Second, the child penalty is defined as the average effect of having children on women relative to men over a specified event time horizon, in this case, controlling for the distribution of observable characteristics and restricting ourselves to the common support:

$$\delta^{\mathrm{R}} = \int_{S^M \cap S^F} \left[g^M(x, t \geq 0) - g^F(x, t \geq 0) \right] \frac{dF^F(x, t \geq 0)}{\mu^F(S^M)} - \int_{S^M \cap S^F} \left[g^M(x, t < 0) - g^F(x, t < 0) \right] \frac{dF^F(x, t < 0)}{\mu^F(S^M)}$$

This is similar to the approach by Kleven (2023), but it accounts for common-support differences in the distribution of observables between males and females.

Third, we perform an additional nonparametric decomposition along the event using time t = 0 as the base category, allowing us to create more comparable groups over time. However, we also include the year of birth in the set of covariates to follow synthetic individuals over time. Thus, let us define $\theta^{t=0}(S) = \int \frac{dF^F(x,b;\ t=0)}{\mu^F(S^M)}$ and $\theta^{t=0}(S) = \int \frac{dF^F(x,b;\ t=0)}{\mu^F(S^M)}$ the probability measure of the set S after matching, where $b \in B$ denotes the year of birth. We obtain:

$$
\begin{aligned}
\delta_R = & \int_{S^{t\neq0}\cap S^{t=0}} \left[\delta^{R,\,t\neq0}(x,b) - \delta^{R,\,t=0}(x,b)\right] \frac{dF^{t=0}(x,b)}{\theta^{t=0}(S^{t\neq0})} \\
& + \int_{S^{t\neq0}\cap S^{t=0}} \delta^{R,\,t\neq0}(x,b) \left[\frac{dF^{t\neq0}}{\theta^{t\neq0}(S^{t=0})} - \frac{dF^{t=0}}{\theta^{t=0}(S^{t\neq0})}\right](x,b) \\
& + \left[\int_{\overline{S^{t=0}}} \delta^{R,\,t\neq0}(x,b) \frac{dF^{t\neq0}(x,b)}{\theta^{t\neq0}(S^{t=0})} - \int_{S^F} \delta^{R,\,t\neq0}(x,b) \frac{dF^{t\neq0}(x,b)}{\theta^{t\neq0}(S^{t=0})}\right] \theta^{t\neq0}(\overline{S^{t=0}}) \\
& + \left[\int_{S^{t\neq0}} \delta^{R,\,t=0}(x,b) \frac{dF^{t=0}(x,b)}{\theta^{t=0}(S^{t\neq0})} - \int_{\overline{S^{t\neq0}}} \delta^{R,\,t=0}(x,b) \frac{dF^{t=0}(z)}{\theta^{t=0}(S^{F})}\right] \theta^{t=0}(\overline{S^{t\neq0}})
\end{aligned}
$$

which we denote by

$$\delta^R = \delta^{R,R} + \delta^{R,X} + \delta^{R,t\neq0} + \delta^{R,t=0},$$

where $\delta R,^{t\neq0}$ ($\delta R,^{t=0}$) is the part of the penalty attributable to the event where we cannot find observations that exhibit the same combination of observable characteristics compare to the baseline, overt time; $\delta^{R,X}$ is the part of the penalty that can be explained by differences in the distribution of characteristics of over the common support across time; $\delta^{R,R}$ corresponds to the penalty that remains after controlling for differences in the distribution of observable characteristics among men and women in the same birth cohort, restricting ourselves to the common support, over time—reflecting thus the pseudo-panel approach.

Using a set of repeated cross sections for a given country, the matching algorithm described above is as follows:

Step 1: Matching within period *t*:

1.1 Select one female from the sample at time t (without replacement).

1.2 Select all the men who have the same observable characteristics X_{it} as the female previously selected, including their age/birth cohort and having a child or not.

1.3 With all the individuals selected in step 2, construct a synthetic individual whose earnings are essentially the average of all of them and match her to the original female.

1.4 Put the observations of both individuals (the synthetic male and the female) in their respective new samples of matched individuals.

1.5 Repeat steps 1 through 4 until exhausting the original female sample.

Step 2: Matching between time t = 0 and time t + s with s ≠ 0:

2.1. Select one individual from the sample at time t = 0 (without replacement).

2.2. Select all individuals who have the same observable characteristics Xit, including birth cohort and gender in t = 0 as the individual previously selected, including their gender and birth cohort.

2.3. With all the individuals selected in step 2, construct a synthetic individual whose earnings are the average of all of them and match him to the individual in time t = 0.

2.4 Put the observations of both individuals (the synthetic male and the female) in their respective new samples of matched individuals.

2.5 Repeat steps 1 through 4 until exhausting the original sample at time t = 0.

Step 3: Estimation of the motherhood penalty: 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.

This algorithm allows us to decompose gaps accounting for gender differences in the support of observable characteristics. It also provides a weighted average of controlled directed effects restricted to each cell in the distribution of observables that balances treatment assignment within cells. The limitation of this algorithm is that by virtue of being nonparametric it is subject to the course of dimensionality—which entails that the common support falls with the number of covariates used in the matching process. Therefore, it is important to balance external concerns with internal validity by showing the robustness of the estimations, adding one observable characteristic at a time. Also, the identity of the treatment and control groups matter given the reweighting procedure. Hence, while we can additively decompose the earnings penalty, the result does not satisfy anonymity.

IV.

Estimating the pseudo-event study design

Next, we restrict ourselves to $\delta^{R,R}$ as it accounts for differences in the distribution of observables. What matters most in obtaining this term is the probability measure

$$w(S) = \int \frac{dF^{t=0}(x, b)}{\theta^{t=0}(S^{t=0})}$$

because it allows us to perform a doubly robust estimation of the differential penalty.

This expression only has positive support whenever males and females exhibit the same distribution of observable characteristics both within a period and over time. Thus, let us define the following pseudo-panel regression equation:

$$w_{igt}Y_{igt} = \left[\delta^{R,R}F_{igt} \times I(t \geq 0)_{igt} + \mu_g + \gamma_t + \psi_{gt} + \varepsilon_{igt}\right]w_{igt}$$

where Y_{igt} is the outcome for individual i that belongs to the cohort $g = \{F, M\} \times B$ at event time t and w_{igt} is the inverse-probability weight defined by the probability measure $w(\cdot)$; F_{igt} is a dummy variable that takes the value of one if individual i is a female, zero otherwise; $I(t \geq 0)_{igt}$ denotes the event whereby any individual with at least one child is considered treated; μ_g, γ_t, ψ_{gt} capture time-invariant covariates for each cohort and non-monotonic trends, including changes in survey design over time; ε_{igt} is the idiosyncratic error term. Standard errors are clustered at the household level following the design-based approach.

The estimated value of $\delta^{R,R}$ in the regression above corresponds to the differential child penalty between men and women accounting for differences in the distribution of observable characteristics and restricting ourselves to the common support, allowing us to make more 'all-else-equal' comparisons. In this sense, we obtain a nonparametric doubly robust inverse-probability-weighted estimator which takes into account differences in observable characteristics both between groups and within periods, as well as over time.

El Salvador

Previous studies have shown that El Salvador has a low motherhood penalty in comparison to other similar countries although it has similar gender outcomes as other developing countries, especially in Latin America (see Appendix A). Despite the demographic gender-relevant composition of El Salvador being similar to most Latin American countries (for example, share of women by age), it is more comparable to other Central American countries such as Guatemala, Honduras, and Nicaragua, with regard to labor-relevant gender outcomes. For instance, these countries have low female participation rates (below 50 percent), low shares of women in the labor force (40 percent or less), and a low share of salaried women (at about 50 percent). Further, they also exhibit a similar labor market composition with around 20 percent of women employed in industry, and around 72 percent in the service sector. Hence, El Salvador is not a unique case in the region, but it has a uniquely low motherhood wage gap.

However, there are significant disparities between men and women regarding labor market outcomes in El Salvador compared to other Latin American countries. Women's participation in the labor market during the past 20 years has been consistently low, among the lowest in Latin America and the Caribbean, with less than 45 percent of working-age women (15 years and older) participating in the labor market, while around 76 percent of men participate in the labor market. There are also important gender differences in types of occupations (salaried, self-employment, and employers).

Women's inactivity is influenced by demographic factors and caring for dependents within the household compared to men. For example, among inactive working-age women, 65 percent cite domestic and care work within the home as the primary reason for being out of the labor market, compared to only 2 percent of men. In contrast, men cite other reasons for their inactivity, such as attending school (31.4 percent), disability (23.7 percent), and retirement (11 percent).

Labor informality is also more frequent among women and has increased in recent years. For instance, using social security as a proxy for informality, the data show that although by 2008, both men and women had the same informality rate, by 2020 the gender gap had reached 6 percentage points.

We also observe a persistent wage gap as the average hourly was US$1.44 (PPP) for men and US$1.32 (PPP) for women, whereas in 2022 these values were US$2.25 and US$2.30 (PPP), respectively. Women are also more likely to be poor. Using the international poverty line of US$6.85 per day at PPP, with 2017 as the base year, statistics show that more than half of the poor are women—53.4percent women and 46.6 percent men; Also, female-headed households represent 35.8 percent of poor households living with income. Moreover, the poverty rate for households headed by women with at least one child under six years of age increased by 4.3 percentage points after fiscal policy, from 38.4 percent to 42.7 percent.

V.

Data

Our main data come from the national household survey for El Salvador, the 'Encuesta de Hogares de Propósitos Múltiples' (EHPM). In our analysis, we include 22 rounds between 2000 and 2023. Importantly, the sampling design of the EHPM is updated every five years. For 2000–2002, the sampling frame is based on the 1971 census; for 2003–2007 this is the 1992 census; and from 2008 onward the 2007 census is used. These survey rounds are nationally representative. The surveys for the period of analysis are harmonized by the Socioeconomic Database for Latin America and the Caribbean (SEDLAS). Across the surveys we use, the sample is representative at the national, regional, and the autonomous municipality levels.

VI.

Results

Figure 4.1a displays the average firstborn earnings penalty for both men and women given our nonparametric pseudo-event study design in percentage points; the x-axis is the running variable which corresponds to the years that have passed since the birth of the first child. In this case, we only use the year of birth of both the respondent and his/her first child as matching variables. Hence, our comparisons are restricted to the common support on the basis of these observable characteristics (that is, 72 percent of our sample). Despite this, our results still represent about 6.5 million observations.

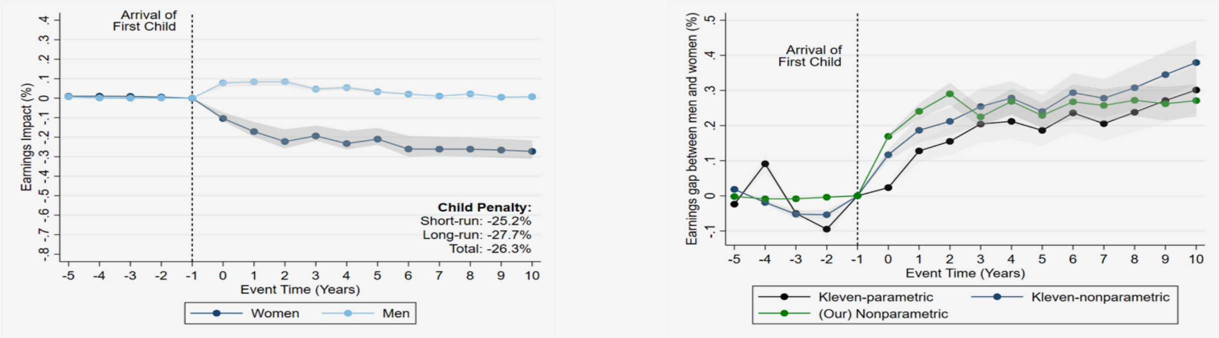
Figure 4.1a provides suggestive evidence for the parallel trends assumption as we do not observe a statistically significant difference between the wage penalties between men and women, on average. After the event of having the first child (t=0), we observe a divergence of the penalties against women. In fact, we observe that the wage penalty is about 20 percent at the onset and grows over time, reaching about 25 percent by year 5 during the main child-rearing years and then stabilizes. Overall, we observe that the motherhood earnings penalty in the firstborn's first five years of age is 25 percent, and for his/her first 10 years, it is about 28 percent, although these gaps are statistically equal (Table 4B.2).

Figure 4.1b compares the motherhood penalties obtained from both the nonparametric methodology herein and the parametric approach developed by Kleven both without restricting and restricted to the common support. Overall, we observe that the differential penalty is larger using the nonparametric approach, but these differences are not statistically significant for our short-run and long-run estimates (Table 4B.2), albeit the estimates do seem larger for the nonparametric approach between periods zero and three. Furthermore, our estimates are also more efficient owing to the higher functional flexibility inherent in the nonparametric procedure. However, we will show below that differential self-selection into part-time work explains much of the differential earnings penalty.

FIGURE 4.1. EFFECTS OF THE FIRST CHILDBIRTH ON LABOR EARNINGS

a. Non parametric

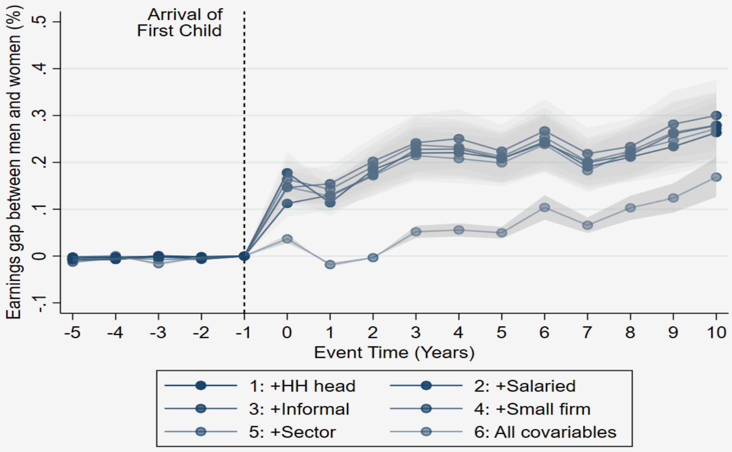
b. Parametric vs nonparametric



Note: Panel b reports the standardized estimates of the effects of the first childbirth on labor earnings for men and women, separately using the methodology proposed by Kleven (2023) with a Propensity Score Matching procedure and restricted to the common support of observable characteristics. Controls include the year of the interview and the respondent's birth year.

Next, we add other observable characteristics to increase comparability within cells of observables (Figure 4.2). We add one observable characteristic at a time in the following order: education and area of residence, is a household head, is a salaried worker, is an informal worker, works in a small firm, sector, and is a part-time worker (see also Table 4B.3). Overall, we observe a strong level of stability in the motherhood earnings penalty, except when adding whether the worker is a part-time worker or not, in which case the differential earnings penalty is much lower: about –3 percent for the short-run and –7 percent for the long-run estimates. This provides support for the idea that one of the mechanisms by which motherhood earnings penalties arise is by differential self-selection into part-time work, which is much higher for women with children (Table 4B.1).

FIGURE 4.2. EFFECTS OF THE FIRST CHILDBIRTH ON LABOR EARNINGS, ADDING COVARIATES WITHOUT REPLACEMENT



Note: The standardized coefficient measures children's impact as a percentage of the counterfactual outcome absent children relative to the year before the first childbirth. Controls include year, age in years. The motherhood effect reported is the average motherhood effect from t = 1 to t = 5 (short run) and from t = 6 to t = 10 (long run).

V.

Discussion

Motherhood earnings penalties—that is, the impact of parenthood on women's earnings relative to men—account for a substantial part of the remaining gender inequality in developed countries as well as in Latin America. Thus, eliminating gender inequality is virtually synonymous with eliminating differential child penalties. Herein, we contribute to the endeavor of measuring and analyzing these penalties by proposing a nonparametric pseudo-event study approach that takes into account differences in the distribution of observable characteristics between men and women, both within a period and over time. In this way, our methodology lays out a pseudo-panel regression approach that increases comparability using cross-sectional data, allowing researchers to make more 'all-else-equal' comparisons with observational data. Further, our methodology also provides more efficient estimators.

We investigate the case of El Salvador, finding that the differential earnings penalty for the first child is much lower than using the extant parametric approach. We find that this result is driven by a differential self-selection of women into part-time jobs, as women with children are more likely to take on part-time jobs compared to their peers regardless of gender. This result echoes with the argument that the last chapter of gender convergence must involve changes in labor market structure and job flexibility, especially for women facing fertility choices.

References

Atal, J., H. Ñopo, and N. Winder. 2009. New Century, Old Disparities: gender and Ethnic Wage Gaps in Latin America.

Berniell, I., L. Berniell, D. De la Mata, M. Edo, and M. Marchionni. 2021a. “Gender Gaps in Labor Informality: The Motherhood Effect.” Journal of Development Economics 150: 102599.

Berniell, I., L. Berniell, D. De la Mata, M. Edo, Y. Fawaz, M. P. Machado, and M. Marchionni. 2021b. “Motherhood and the Allocation of Talent.” Working Paper No. 270, CEDLAS.

Berniell, I., L. Berniell, D. de la Mata, M. Edo, and M. Marchionni. 2023. “Motherhood and Flexible Jobs: Evidence from Latin American Countries.” World Development 167: 106225.

Blau, F. D., and L. M. Kahn. 2017. “The Gender Wage Gap: Extent, Trends, and Explanations.” Journal of Economic Literature 55 (3): 789–865.

Callaway, B., and P. H. C. Sant’Anna. 2020. “Difference-in-Differences with Multiple Time Periods.” Journal of Econometrics. https://doi.org/10.1016/j.jeconom.2020.12.001

Clarke, D. 2018. “Children and Their Parents: A Review of Fertility and Causality.” Journal of Economic Surveys 32 (2): 518–540.

Cukrowska-Torzewska, E., and A. Lovasz. 2020. “The Role of Parenthood in Shaping the Gender Wage Gap–A Comparative Analysis of 26 European Countries.” Social Science Research 85: 102355.

Deaton, A. 1985. “Panel Data from Time Series of Cross-Sections.” Journal of Econometrics 30: 109–126.

De Chaisemartin, C., and X. D’Haultfceuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” American Economics Review 110 (9): 2964–2996.

Doepke, M., A. Hannusch, F. Kindermann, and M. Tertilt. 2023. The economics of fertility: A new era. In Handbook of the Economics of the Family (Vol. 1, No. 1, pp. 151-254). North-Holland.

Gaebler, J., W. Cai, G. Basse, R. Shroff, S. Goel, and J. Hill. 2022. A causal framework for observational studies of discrimination. Statistics and public policy, 9(1):26–48.

Gamboa, L. F., and B. Zuluaga. 2013. “Is There a Motherhood Penalty? Decomposing the Family Wage Gap in Colombia.” Journal of Family and Economic Issues 34: 421–434.

Hausman, C., and D. S. Rapson. 2018. “Regression Discontinuity in Time: Considerations for Empirical Applications.” Annual Review of Resource Economics 10, 533–552.

Imai, K., L. Keele, and T. Yamamoto. 2010. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects.

Kleven, H. 2023. “The Geography of Child Penalties and Gender Norms: Evidence from the United States.” Technical Report, National Bureau of Economic Research.

Kleven, H., C. Landais, and G. Leite-Mariante. 2023. The Child Penalty Atlas. National Bureau of Economic Research.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller. 2019. "Child Penalties across Countries: Evidence and Explanations." In AEA Papers and Proceedings, volume 109, 122–126. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kline, R. B. 2011. Principles and Practice of Structural Equation Modeling (3rd edition). Guilford Press.

Piras, C., and L. Ripani. 2005. The Effects of Motherhood on Wages and Labor Force Participation: Evidence from Bolivia, Brazil, Ecuador and Peru, Volume 109. Washington, DC: Inter-American Development Bank.

Ñopo, H. 2008. "Matching as a Tool to Decompose Wage Gaps." The Review of Economics

and Statistics 90 (2): 290–299.

Ñopo, H., and A. Hoyos. 2010. "Evolution of Gender Gaps in Latin American at the Turn of the Twentieth Century: An Addendum to "New Century, Old Disparities." IDB Working Papers Series No. 176.

McKenzie, D. 2004. "Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels." Journal of Econometrics 120 (2): 235–262.

Marchionni, M., and J. Pedrazzi. 2023. "The Last Hurdle? Unyielding Motherhood Effects in the Context of Declining Gender Inequality in Latin America." Working Paper 321, CEDLAS.

Piras, C., and L. Ripani. 2005. The Effects of Motherhood on Earnings and Labor Force Participation: Evidence from Bolivia, Brazil, Ecuador and Peru, Volume 109. Washington, DC: Inter-American Development Bank.

Robins, James M., Miguel A. Hernan, and Babette A. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." Epidemiology 11 (5): 550–60.

Robayo-Abril, M., A. M. Tribin, & J. A. Oliva. 2023. Fiscal Policy as a Tool for Gender Equity in El Salvador.

Sloczynski, T. 2015. "The Oaxaca-Blinder Unexplained Component as a Treatment Effects Estimator." Oxford Bulletin of Economics and Statistics 77 (4): 0305–9049.

Sun, L., and S. Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." Journal of Econometrics 225: 175–199.

Verbeek, M., and T.Nijman. 1993. "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections," Journal of Econometrics, Elsevier, vol. 59(1-2) 125-136.

Verbeek, M. 2008. "A guide to modern econometrics". 3th edition. Willey, New York.

Villanueva, A., and K. H. Lin. 2020. "Motherhood Wage Penalties in Latin America: The Significance of Labor Informality." Social Forces 99 (1): 59–85.

Villanueva, A., and K. H. Lin. 2020. "Motherhood Earnings Penalties in Latin America: The Significance of Labor Informality." Social Forces 99 (1): 59–85.

Appendix A:

Case Selection

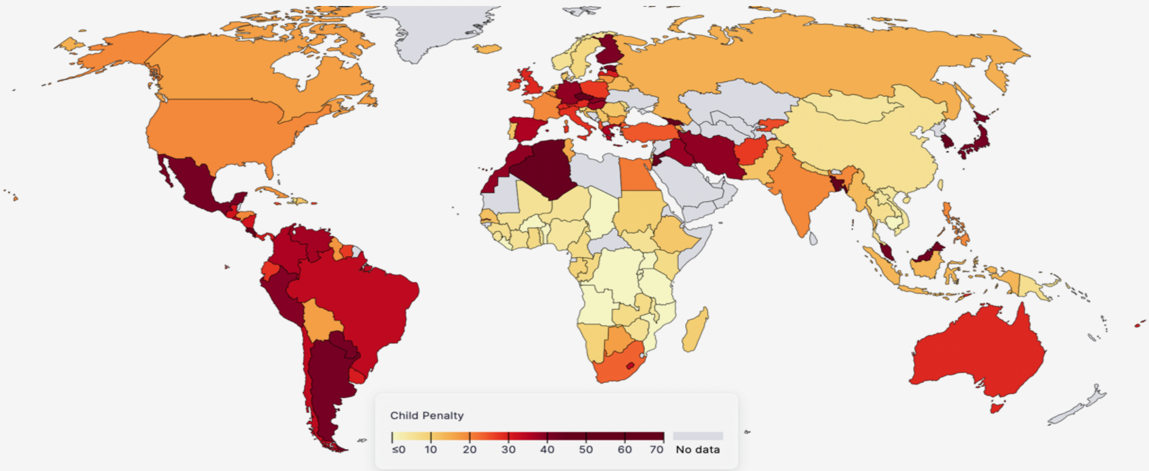
We select El Salvador as our case study because the motherhood earnings penalty is comparatively lower compared to other countries in the Latin American Region in the Marchionni and Pedrazzi (2023) analysis (Table 4A.1) and because its motherhood penalty is comparable to some developed Western European countries in Kleven et al. (2023), as shown in Figure 4A.1. This case is even more interesting once we consider that El Salvador is especially similar to other Latin American countries and some Asian, Middle Eastern, and African countries with regard to gender outcomes, as we show below.

TABLE 4A.1. MOTHERHOOD EARNINGS PENALTY IN LATIN AMERICA

Country	Marchioni and Pedrazzi (2023)	Kleuven et al. (2023)	Difference
El Salvador	11.8	38	−26.2
Paraguay	23.7	44	−20.3
Panama	24	35	−11
Ecuador	25.6	33	−7.4
Honduras	26.6	25	1.6
Uruguay	26.9	35	−8.1
Bolivia	28	23	5
Costa Rica	30.8	48	−17.2
Colombia	30.8	38	−7.2
Chile	33.2	37	−3.8
Mexico	33.9	44	−10.1
Argentina	34.8	44	−9.2
Brazil	40.3	37	3.3
Peru	40.4	42	−1.6

Source: Marchionni and Pedrazzi 2023.
Note: Motherhood earnings penalty estimates are computed using household survey data from SEDLAC.

FIGURE 4A.1. MOTHERHOOD EARNINGS PENALTY IN THE WORLD



Source: Kleven et al. 2023.
Note: Motherhood earnings penalty estimates are computed using household survey data from Demographic Health Surveys.

To identify those countries that are likely to display similar gender-relevant outcomes compared to El Salvador, we use numerous indicators gleaned from the World Bank's statistical indicators. To select the indicators, we use various variables associated with gender sociodemographic outcomes regarding education, the labor market, and others and perform an automated sub-selection of indicators to reduce the amount of missing data which is common in these data. We define our period of analysis as 2000–2020, to be consistent with the period of study of our analysis. The variables selected from our analysis are shown in Table 4.A3.

We use the indicators in Table 4.A3 to perform a cluster analysis to find comparable countries to El Salvador on the identified dimensions. For this endeavor, we choose Ward's linkage using the L2 norm to measure dissimilarity between countries in a multidimensional space. We also use Duda–Hart Je(2)/Je(1) and pseudo-T2 to define the optimal number of clusters—to minimize dissimilarity within groups and maximize it between groups. We find that the largest Je(2)/Je(1) stopping-rule value is 0.76, corresponding to 3 groups, with a pseudo-T2 of 9.2—which is among the smallest using 14 potential clustering groups (Table 4.A1). In this regard, a combination of a high Je(2)/Je(1) with a low pseudo-T2 indicates more distinct clusters, while increasing within-cluster homogeneity.

We find that El Salvador is classified with other Latin American countries such as Guatemala, Honduras, Mexico, and Nicaragua as well as with Malaysia, Maldives, Mauritius, Sri Lanka, and Tonga. We further validate our data-driven grouping by looking at the means in every indicator for the income-level aggregates compared to those values for El Salvador. All in all, we find that El Salvador is similar to lower-middle-income countries.

TABLE 4A.2. CLUSTER ANALYSIS

Clusters	Je(2)/Je(1)	Pseudo R-squared
1	0.6831	56.6
2	0.7558	22.95
3	0.7652	9.21
4	0.7389	17.32
5	0.673	9.23
6	0.7213	15.07
7	0.7448	10.62
8	0.4872	7.37
9	0.6996	8.59
10	0.6599	8.53
11	0.6512	5.36
12	0.6289	4.72
13	0.6981	8.65
14	0.3818	6.48

TABLE 4A.3. COUNTRY AVERAGES FOR SELECTED INDICATORS

Indicator	El Salvador	Guatemala	Honduras	Malaysia	Maldives	Mauritius	Mexico	Nicaragua	Sri Lanka	Tonga	Low and lower middle income	Lower middle income	Upper middle income
Employment in agriculture, female (% of female employment) (modeled ILO estimate)	4.21	13.23	9.23	9.17	4.85	7.30	4.50	8.36	36.75	7.51	10.51	16.89	6.67
	(0.19)	(0.59)	(0.36)	(0.46)	(0.49)	(0.49)	(0.21)	(0.23)	(1.31)	(0.64)	(0.67)	(1.34)	(0.27)
Employment in industry, female (% of female employment) (modeled ILO estimate)	19.62	20.06	21.78	22.41	24.10	24.45	18.11	15.73	26.49	48.59	24.13	21.02	27.53
	(0.47)	(0.71)	(0.56)	(0.78)	(1.17)	(1.86)	(0.43)	(0.43)	(0.37)	(1.63)	(0.69)	(0.47)	(1.22)
Employment in services, female (% of female employment) (modeled ILO estimate)	76.17	66.71	68.99	68.42	71.04	68.24	77.39	75.92	36.76	43.90	65.36	62.09	65.80
	(0.45)	(1.24)	(0.44)	(1.21)	(1.61)	(2.27)	(0.62)	(0.35)	(1.36)	(0.99)	(1.00)	(1.69)	(1.31)
Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)	45.64	40.83	42.78	46.36	38.36	42.07	42.23	43.47	34.82	43.38	41.99	40.47	42.48
	(0.15)	(0.40)	(0.75)	(0.78)	(0.41)	(0.63)	(0.48)	(1.00)	(0.22)	(0.14)	(0.29)	(0.51)	(0.35)
Labor force participation rate, female (% of female population ages 15–64) (modeled ILO estimate)	48.87	42.36	44.03	49.30	39.54	47.11	45.07	45.44	38.27	45.79	44.58	42.52	45.36
	(0.19)	(0.40)	(0.78)	(0.91)	(0.48)	(0.87)	(0.58)	(1.10)	(0.24)	(0.09)	(0.32)	(0.50)	(0.44)
Labor force, female (% of total labor force)	40.80	33.80	34.11	36.32	30.92	36.38	36.28	35.26	33.50	39.41	35.68	34.17	35.86
	(0.16)	(0.21)	(0.35)	(0.34)	(0.56)	(0.50)	(0.31)	(0.60)	(0.15)	(0.19)	(0.23)	(0.35)	(0.33)
Population, female (% of total population)	52.15	50.41	49.44	48.76	45.37	50.32	51.05	50.72	51.15	49.68	49.90	50.43	49.04
	(0.03)	(0.01)	(0.00)	(0.05)	(0.58)	(0.04)	(0.00)	(0.01)	(0.12)	(0.10)	(0.14)	(0.10)	(0.23)
Survival to age 65, female (% of cohort)	81.06	77.49	79.20	85.16	87.64	83.98	82.74	79.41	85.61	77.00	81.93	80.43	83.30
	(0.28)	(0.38)	(0.53)	(0.28)	(0.99)	(0.40)	(0.22)	(0.81)	(0.73)	(0.06)	(0.30)	(0.42)	(0.42)
Unemployment, female (% of female labor force) (modeled ILO estimate)	3.87	3.47	6.32	3.55	8.04	12.07	4.27	6.40	8.90	3.42	6.03	6.27	6.27
	(0.13)	(0.05)	(0.43)	(0.05)	(0.78)	(0.41)	(0.15)	(0.35)	(0.56)	(0.41)	(0.23)	(0.28)	(0.39)
Wage and salaried workers, female (% of female employment) (modeled ILO estimate)	51.92	45.97	44.49	76.33	59.96	84.56	65.51	50.77	56.12	42.08	57.77	49.34	65.69
	(0.25)	(1.23)	(0.45)	(0.52)	(1.92)	(0.34)	(0.38)	(0.26)	(0.51)	(0.48)	(0.98)	(0.57)	(1.52)

Note: Standard errors in parentheses.

TABLE 4A.4. SELECTED INDICATORS

Indicator Name	Description
Employment in agriculture, female (% of female employment) (modeled ILO estimate)	Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job or working-time arrangement. The agriculture sector consists of activities in agriculture, hunting, forestry, and fishing, in accordance with division 1 (ISIC 2) or categories A-B (ISIC 3) or category A (ISIC 4).
Employment in industry, female (% of female employment) (modeled ILO estimate)	Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job or working-time arrangement. The industry sector consists of mining and quarrying, manufacturing, construction, and public utilities (electricity, gas, and water), in accordance with divisions 2-5 (ISIC 2) or categories C-F (ISIC 3) or categories B-F (ISIC 4).
Employment in services, female (% of female employment) (modeled ILO estimate)	Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job or working-time arrangement. The services sector consists of wholesale and retail trade; restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services, in accordance with divisions 6-9 (ISIC 2) or categories G-Q (ISIC 3) or categories G-U (ISIC 4).
Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)	Labor force participation rate is the proportion of the population ages 15 and older that is economically active: all people who supply labor for the production of goods and services during a specified period.
Labor force participation rate, female (% of female population ages 15–64) (modeled ILO estimate)	Labor force participation rate is the proportion of the population ages 15–64 that is economically active: all people who supply labor for the production of goods and services during a specified period.
Labor force, female (% of total labor force)	Female labor force as a percentage of the total shows the extent to which women are active in the labor force. The labor force comprises people ages 15 and older who supply labor for the production of goods and services during a specified period.
Population, female (% of total population)	Female population is the percentage of the population that is female. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.
Survival to age 65, female (% of cohort)	Survival to age 65 refers to the percentage of a cohort of newborn infants that would survive to age 65, if subject to age-specific mortality rates of the specified year.
Unemployment, female (% of female labor force) (modeled ILO estimate)	Unemployment refers to the share of the labor force that is without work but available for and seeking employment.
Wage and salaried workers, female (% of female employment) (modeled ILO estimate)	Wage and salaried workers (employees) are those who hold the type of jobs defined as ‘paid employment jobs’, where the incumbents hold explicit (written or oral) or implicit employment contracts that give them a basic remuneration that is not directly dependent upon the revenue of the unit for which they work.

Note: Standard errors in parentheses.

Appendix B:

Empirical appendix

TABLE 4B.1. DESCRIPTIVE STATISTICS

	Men		Women	
	No child	Child	No child	Child
Labor market outcomes				
Employed (%)	89.15	94.06	58.36	47.03
Informal (%)	44.88	42.19	56.37	55.48
Salaried (%)	71.35	72.92	58.46	57.32
Small firm (5 workers or less) (%)	41.12	43.61	30.64	31.10
Labor income per hour	1.89	2.06	1.87	2.18
Hours worked	45.99	46.95	44.20	35.96
Demographic characteristics				
Age in years	35	34	35	31
Household head (%)	94.10	92.99	36.43	22.42
Number of members in main household	2.02	3.77	2.23	3.63
Urban (%)	66.30	58.49	68.52	61.35
Educational attainment				
Never attended (%)	8.63	6.46	11.96	5.79
Incomplete primary education (%)	37.52	37.06	38.14	35.16
Complete primary education (%)	14.22	16.03	10.62	13.40
Incomplete secondary education (%)	7.38	8.75	7.17	11.10
Complete secondary education (%)	16.78	19.47	15.45	20.77
Incomplete tertiary education (%)	8.46	6.17	8.73	7.59
Complete tertiary education (%)	7.00	6.06	7.93	6.19
Sector				
Agriculture, hunting and forestry (%)	19.72	20.92	2.89	3.26
Fishing (%)	1.56	1.36	0.20	0.16
Mining and quarrying (%)	0.10	0.15	0.01	0.01
Manufacturing (%)	14.70	15.76	17.57	18.50
Electricity, gas and water supply (%)	0.76	0.98	0.34	0.27
Construction (%)	11.26	11.19	0.53	0.35
Wholesale and retail trade (%)	18.92	18.25	30.44	29.88
Hotels and restaurants (%)	3.17	2.58	11.86	13.45
Transport, storage, and communications (%)	7.89	8.52	1.38	1.53
Financial intermediation (%)	1.28	1.06	1.83	2.10
Real estate, renting, and business activities (%)	6.10	5.44	3.93	3.05
Public administration and defense (%)	5.54	6.13	3.21	3.37
Education (%)	2.38	1.90	5.43	5.42
Health and social work (%)	1.63	1.80	4.53	5.06
Other community, social, and personal service activities (%)	4.08	3.09	6.03	6.59
Activities of private households as employers (%)	0.89	0.83	9.73	6.99
Extraterritorial organizations and bodies (%)	0.02	0.01	0.08	0.01
Number of observations	30,145	59,891	24,239	58,655

Source: EHPM El Salvador 2000–2022.

Note: This table compares labor market and demographic variables for women and men observed with and without the first child in the pseudo-panel data.

TABLE 4B.2. EFFECTS OF THE FIRST CHILDBIRTH ON LABOR EARNINGS

	(1)	(3)	(4)
Short term: *t=0 to t=5*	−0.162	−0.229	−0.252
	(0.059)	(0.055)	(0.049)
Long term: *t=6 to t=10*	−0.267	−0.336	−0.277
	(0.200)	(0.063)	(0.045)
Total: *t=0 to t=10*	−0.210	−0.277	−0.263
	(0.123)	(0.059)	(0.047)
Matching procedure	Kleven (Parametric)	Kleven (parametric + cell common support)	Nonparametric
Percentage of men in the common support (%)	98.65	73.51	73.51
Percentage of women in the common support (%)	97.18	70.84	70.84

Note: Columns 1 and 2 report the standardized estimates of the effects of the first childbirth on labor earnings for men and women, separately using the methodology proposed by Kleven (2023) with a Propensity Score Matching procedure and restricted to the common support of observable characteristics, respectively. Controls include the year of the interview and the respondent's birth year.

TABLE 4B.3. EFFECT OF FIRST CHILDBIRTH ON LABOR EARNINGS, ADDING COVARIATES WITHOUT REPLACEMENT

	(1)	(2)	(3)	(4)	(5)	(6)
Short term: *t=0 to t=5*	−0.187	−0.179	−0.203	−0.196	−0.178	−0.028
	(0.050)	(0.048)	(0.054)	(0.051)	(0.030)	(0.016)
Long term: *t=6 to t=10*	−0.228	−0.239	−0.260	−0.244	−0.232	−0.113
	(0.052)	(0.050)	(0.054)	(0.053)	(0.031)	(0.015)
Total: *t=0 to t=10*	−0.206	−0.207	−0.229	−0.218	−0.202	−0.067
	(0.051)	(0.049)	(0.054)	(0.052)	(0.031)	(0.016)
Additional matching variables	+HH. Head	+Salaried	+Informal	+Small firm	+Sector	+Part-time
Percentage of men in the common support (%)	70.9	68.0	65.1	55.7	47.9	42.1
Percentage of women in the common support (%)	61.4	61.2	57.9	45.3	48.0	35.2

Note: Columns 1 and 2 report the standardized estimates of the effects of the first childbirth on labor earnings for men and women, separately using the methodology proposed by Kleven (2023) with a Propensity Score Matching procedure and restricted to the common support of observable characteristics, respectively. Controls include the year of the interview and the respondent's birth year.