DISCUSSION PAPER SERIES

Fertility as a Driver of Maternal Employment

Julia Schmieder

DISCUSSION PAPER SERIES

Fertility as a Driver of Maternal Employment

Julia Schmieder
DIW Berlin and IZA

ABSTRACT

Fertility as a Driver of Maternal Employment*

Based on findings from high-income countries, typically economists hypothesize that having more children unambiguously decreases the time mothers spend in the labor market. Few studies on lower-income countries, in which low household wealth, informal child care, and informal employment opportunities prevail, find mixed results. Using Mexican census data, I find a positive effect of an instrument-induced increase in fertility on maternal employment driven by an increase in informal work. The presence of grandparents and low wealth appear to be important. Econometric approaches that allow extrapolating from this complier-specific effect indicate that the response in informal employment is non-negative for the entire sample.

JEL Classification:	J13, J16, J22, J46
Keywords:	fertility, female labor supply, middle-income countries, informality

Corresponding author:
Julia Schmieder
DIW Berlin
Mohrenstraße 58
10117 Berlin
Germany
E-mail: jschmieder@diw.de

1 Introduction

An extensive and longstanding literature on the effects of children on household behavior emphasizes their costs in terms of maternal time inputs. Empirical research from high-income countries fuels this by exclusively finding a negative causal relationship between family size and maternal labor supply.[1] Hence, economists often hypothesize that having more children, especially young ones, clearly decreases the time mothers spend in the labor market. However, from a theoretical point of view, the sign of the relationship between family size and maternal labor supply is ambiguous. Under reasonable assumptions, a woman is predicted to decrease her leisure time after increasing her fertility. However, some women will increase labor supply and focus on monetary investments while others will reduce labor supply and focus on time investments in order to increase child quality. Besides the monetary inputs in child quality, there are some fixed costs of children that decrease household income, which makes increases in labor supply more valuable by increasing the marginal utility of income.[2]

This paper provides new insights about the effect of fertility on maternal labor supply in a setting with a low level of average household wealth, informal child care provision within extended households, and a high prevalence of informal work. These are all conditions that are typically more prevalent in middle- and low-income countries than in high-income countries; all playing a potentially interesting role in the relationship between fertility and maternal employment. First, little income implies that households already have a high marginal utility of income. Budgetary consequences of children might be more severe in these households and they can have a larger impact on the mother's propensity to take up employment in order to provide for additional resources. Second, if there are other household members who provide some child care, this will lower the return to maternal time relative to monetary investments into the children. Hence, they increase the incentives to augment labor supply and focus on monetary inputs. Third, the informal labor market might also provide employment opportunities to mothers: informal employment might be easier to arrange with child care schedules, which are often half-day, it might provide the possibility to take care of the children while working, and fixed costs of work might be lower in informal than in formal employment.

In order to establish a causal link between the two variables, an instrumental variable approach exploiting parental preferences for mixed-sex siblings, as outlined in Angrist and Evans (1998), is used. My empirical analysis uses data on a sample of about 500,000 Mexican women. The setting provides two attractive features. First, Mexican census data from 2010 allows for observing detailed labor market outcomes in both the informal and formal sectors as well as for identifying the complex household constellations for a large sample of mothers. Second, Mexican parents show a preference for a mixed-sex sibling composition of their children while

---

[1]See Appendix Table A1 for an overview of estimates from high-income countries.

[2]Angrist and Evans (1996) and Heath (2017) provide theoretical models of how children affect a woman's time allocation, which allow to analyze and understand these different mechanisms.

at the same time there are no strong preferences and discriminatory practices toward one sex.

My results confirm the presence of mixed-sex sibling preferences: Families with two children of the same sex are more likely to have another child compared to those with children of different sexes. A considerable and significant negative correlation between having more than two children and female labor supply is reversed when the potential endogeneity of fertility is accounted for with the proposed instrumental strategy. I find an increase in maternal employment in the informal sector at the extensive margin. At the same time, there is no effect on the probability of working in the formal sector. This is in line with suggestive evidence that informal jobs provide greater flexibility to care for children compared to formal jobs for women in my sample. I further show that having more than two children increases the probability of living with extended family that might provide childcare. In particular, the probability of living with maternal grandparents rises. Moreover, I find suggestive evidence that the positive employment effect is mainly driven by women who live in households with relatively low household wealth, which have a particularly high value for labor income as family size increases.

These employment effects contrast with findings for other low- and middle-income countries. This raises a number of questions: Can potential violations of the identification assumptions challenge the findings? Are there heterogeneous effects across the working hours distribution for the group of compliers? And what can be learned from the estimated effects that pertain to the group of compliers, i.e., mothers who had more children than they otherwise would have due to the sex composition of their first two children? I provide several pieces of evidence in order to shed lights on these questions.

First, I carefully assess the concerns that are discussed with respect to the internal validity of the instrument (f.e., Wolpin and Rosenzweig (2000), Lee (2008)) using several additional data sources. None point to a violation of the identification assumptions. In addition, I take advantage of a formal test procedure for the validity of the identification assumptions proposed by Huber and Mellace (2015). Applying this test, I cannot reject the null hypothesis that the identification assumptions are valid. Second, I estimate local quantile treatment effects which correspond to the effects of having more than two children on the maternal working hours distribution of the complier subpopulation. My findings suggest that fertility increases maternal labor supply in the formal sector at the intensive margin by shifting a few quantiles from around 20 to 40 hours, while mainly having an impact at the extensive margin in the informal sector. Third, I assess the question of generalizability beyond the group of mothers who comply with the instrument to always- and never-takers. The proportion of compliers in my sample is estimated to be around 3 to 4% and I show that the average complier seems to be different from the average non-complier both in terms of her observable characteristics and potential labor market outcomes. This challenges the one-to-one transferability of this local effect to non-compliers. However, by imposing further assumptions, I can bound and estimate the effect of having more than two children on the propensity to be informally employed to be non-negative for all mothers in the sample based on approaches by Kowalski (2016) and Brinch et al. (2017),

respectively.

My paper is primarily related to the large literature on the relationship between fertility and female labor supply. As illustrated in Appendix Table A1, there are numerous studies on high-income countries that unambiguously find a negative effect of an exogenous increase in family size on maternal labor supply at the extensive margin.[3] In contrast, Appendix Table A2 indicates that the few studies examining this relationship in middle- and low-income countries find mixed results.[4] Agüero and Marks (2008, 2011) find close to zero and insignificant effects of different fertility measures on the decision to work in a pooled sample of middle- and low-income countries. Cáceres-Delpiano (2012) finds a significant negative effect on the probability of working when increasing fertility beyond one child, but no significant effect when moving beyond two. Cruces and Galiani (2007) and Heath (2017) find significant negative effects of having more than two children and of the number of children, respectively, on the probability to work. I contribute to this literature by showing that mothers can be pushed into taking up employment when having a larger family. In addition, I identify multiple mechanisms that can explain my results, namely low wealth and an increasing presence of potential caretakers.[5,6]

In addition to the effect on maternal employment, there are other dimensions of maternal labor supply that are important to consider. First, I go beyond the effect at the extensive margin of maternal labor supply and investigate how the working hours distribution of compliers changes with children. Among the studies of middle- and low-income countries in Appendix Table A2, only Heath (2017) investigates the intensive margin of labor supply and finds a positive effect of having children on self-employment conditional on employment.[7] Second, I illustrate that informal compared to formal labor arrangements appear to provide more job flexibility at the cost of lower wages. Distinguishing between the two, I show that my positive extensive margin response is driven entirely by an increase in informal jobs.[8,9]

---

[3]Appendix Table A1 includes the following studies on the relationship between fertility and female labor supply that are all based on IV or quasi-experimental approaches with a focus on high-income countries only: Rosenzweig and Wolpin (1980), Bronars and Grogger (1994), Angrist and Evans (1998), Jacobsen et al. (1999), Cristia (2008), Lundborg et al. (2017), and Kleven et al. (2019).

[4]Appendix Table A2 includes the following studies on the relationship between fertility and female labor supply that are all based on IV or quasi-experimental approaches with a focus on middle- and low-income countries only: Cruces and Galiani (2007), Agüero and Marks (2008), Agüero and Marks (2011), Cáceres-Delpiano (2012), and Heath (2017).

[5]Aaronson et al. (2017) estimate the effect of fertility on maternal labor supply across income levels by comparing the effect across countries and time. They find that the effect of an increase in family size on maternal employment is not different from zero at low and negative at higher (per-capita) income levels. In contrast, by comparing households within the same setting, I can explore effects across households with different income levels that face the same institutions and labor market conditions.

[6]There are three cross-country studies, Cruces and Galiani (2007), Aaronson et al. (2017), and Bisbee et al. (2017), that investigate the relationship between fertiliy and maternal employment and also report estimates using Mexican census data. I provide an overview of these results and a discussion in Online Appendix Section B.

[7]However, her effects are conditional on employment and, thus, are only identified under relatively restrictive identification assumptions.

[8]Rodin et al. (2012) find in an interview-based study in Mexico that the need for more job flexibility to provide care forces some women to take up informal employment.

[9]Other studies in Appendix Table A2 study heterogeneities across the type of labor arrangement: Agüero and Marks (2011) show that, in low-income countries, an additional child does not impact maternal employment on

Most of the studies listed in Appendix Table A1 and A2 rely on an IV strategy that is applied in a specific context. Still, the local nature of the estimated local average treatment effects is rarely discussed, despite its importance for interpretation and the derived policy implications. One exception is Bisbee et al. (2017), who apply the samesex-instrument across a large number of countries and time periods and provide an approach to extrapolate the complier-specific local average treatment effect *across different contexts*. In contrast, my paper zooms in and investigates the potential to extrapolate the local treatment effect for compliers to non-compliers *within a context*. Hence, the two studies are highly complementary in the discussion about the local nature of average treatment effects that are estimated using the instrument in a specific context.

The remainder of the paper proceeds as follows: Section 2 provides background information on households and several aspects of the labor market in Mexico. This information is helpful for putting the results into perspective. In Section 3, I describe the data and the construction of the sample. In particular, I provide details on informal employment in the Mexican setting. Section 4 explains the empirical strategy and assesses the underlying identification assumptions. Section 5 presents my findings on the relationship between family size and parental labor market outcomes. In Section 6, I provide some evidence on potential mechanisms behind my findings. The extrapolation of the effects for compliers to non-compliers within the sample is discussed in Section 7. The final section offers some concluding remarks.

2   Background

In 2010, female labor force participation in Mexico was, at 43.2%, among the lowest in Latin America.[10] One potential explanation for this is that Mexican women face difficulties in reconciling their family and work life. In 2009, women faced an exceptionally high burden of unpaid work: They shouldered 6.2 hours per day of housework and care, which amounts to 77% of all unpaid work in their households.[11] Moreover, there is a culture of long working hours in Mexico; the country has among the highest levels of average weekly working hours in the OECD. At the same time, working schedules in the formal sector tend to be very inflexible, with only a small share of workers observed to work part-time (OECD, 2017). Both the unequally shared responsibilities at home and the inflexible formal labor market that expects long hours might force women to drop out of employment as they have more children. On the other hand, many women can rely on informal child care that is provided by family members who reside in the same household (Gong and van Soest, 2002). From 2000 to 2010, there were policies implemented that aimed to reduce the child care burden: Preschool education was declared mandatory for all children aged 3 to 5 years and a federal daycare program was launched.

---

average, but it reduces the probability of paid employment. Cáceres-Delpiano (2012) finds that only jobs with a high degree of informality react to changes in family size, but the results go into the opposite direction as mine.

[10]Data retrieved from `http://data.worldbank.org/country/mexico`, last access 5/2019.

[11]Data retrieved from `http://ilo.org/ilostat`; last access 02/2018.

The latter targets working mothers with children aged 1 to 4 from low-income households that are not covered by social security-based childcare services, i.e. those who work in the informal sector (Staab and Gerhard, 2010; Ángeles et al., 2014).

As in many other middle- and low-income countries, Mexican women face a high likelihood of working in the informal sector. For example, in 2007, around 54% of all employed Mexican women, among them many with young children, worked informally (International Labour Organization, 2009). Moreover, there are large gender gaps: women are more likely to work informally and to have informal jobs of lower quality than men (OECD, 2017). Informal work can both result from segmentation in the labor market that prevents individuals from taking jobs in the formal sector as well as from a free individual choice between formal and informal jobs resulting from a cost-benefit calculation (Perry et al., 2007). Perry et al. (2007) provide evidence that the Mexican labor market is well-integrated and that workers freely choose to be informal. So why would so many women favor informality? By working in the informal sector, individuals with a relatively small income avoid, on average, little taxes since the average tax rate in Mexico is close to zero for these households.[12] They do not pay formal social security contributions and, hence, are not entitled to receive the associated benefits. However, the latter have become relatively less attractive due to newly introduced policies that provide the traditional benefits of contributory social security to informal workers at zero or virtually no cost, such as the universal health insurance ("Seguro Popular") introduced in 2002 and the aforementioned daycare program (Conti and Ginja, 2017; Bosch and Campos-Vazquez, 2014). It might be the case that women decide to enter the informal sector since it offers more flexibility than the formal sector to balance work and household activities (Perry et al., 2007; Rodin et al., 2012; Heath, 2017). Informal employment might provide the possibility to take care of the children while working and the fixed costs of work might be lower in informal compared to formal employment (Heath, 2017). I discuss the relevance of these different hypotheses in Section 3.

The economic situation of many Mexican households with children is affected by the *Oportunidades* program.[13] In 2010, there were close to six million beneficiary families (about 20% of all households), each receiving an average 9,006 Mexican Pesos (715 US$) per year. The program provides basic health care for all members of the family, as well as means-tested educational grants and in-kind school supplies conditional on children attending school (Parker and Todd, 2017).

---

[12]I calculated the tax liability for a monthly income of 3,000 Pesos in the year 2010, which is the mean income of the women in my sample, using `http://www.fiscalia.com/modules.php?name=Calculadoras&op=isr`.

[13]The program was initiated under the name *Progresa* and changed its name in 2002.

3 Data

3.1 Data & sample selection

The analysis is based on Mexican Population and Housing Census ("Censo de Población y Vivienda") data from 2010. The original data is provided by the National Institute of Statistics and Geography (INEGI) in Mexico and was obtained via the IPUMS International database of the Minnesota Population Center (2014). The data set contains 11,938,402 observations in 2,903,640 households, which accounts for approximately 10% of the whole population.[14] It includes detailed information on each household member including his or her family relationship to the other individuals in the household, on labor market outcomes both in the formal and informal market, as well as on household characteristics such as the housing and economic situation.

The units of observation in this study are women who are between 21 and 35 years old and who have at least two children.[15] I am limited to look at this specific parity due to the empirical strategy that is introduced in Section 4. At the same time, this seems to be a reasonable margin to look at given that the reduction in the total fertility rate in Mexico is concentrated in this range, as described in Section 2.

The data set has only incomplete information about a woman's fertility history: It only includes the number of children ever born to each woman and the number of those who were still living at the time of the census. Child-specific characteristics, such as their age and sex, are only available for those who still live with their mothers at the time of the census. This information is necessary in order to implement the empirical strategy. Because of this, I match children and mothers within each household and drop mothers with children who died (6.5%)[16] or moved out of the household (8.3%).[17] I only include women whose eldest child was 18 or younger and whose second child was at least 12 months old at the time of the census (8.4%). Women who were 14 years or younger at first birth are dropped (1.5%).

The final sample consists of 505,569 women, of whom 90% are married or cohabiting at the time of the survey. Compared to the group of all women in the Mexican Census aged 21 to 50 (21 to 35), Table D1 in the Appendix shows that the females in my sample are much more likely to be married or cohabiting and they tend to have lower education. They are more concentrated

---

[14]The enumeration areas for the census are selected with a simple one-stage random sampling within three strata that are based on the size of the municipality (see http://sinegi.page.link/jWc9; last access 11/2019).

[15]This age restriction is motivated in Angrist and Evans (1998). My results do not change if I apply a minimum age of 18 or a maximum age of 40 (see Section 5).

[16]To be more precise, I exclude women who had children that were no longer living at the time of the census, considering all live births and, hence, excluding stillbirths.

[17]I also exclude females for whom at least one of the following variables is missing: Age and/or sex of the two firstborn children, employment status, working hours, and earnings. Moreover, females who had twins at the second birth (or triplets at the first birth) are excluded from the sample due to the difficulty of assigning a birth order to the second and third child.

in the 25-30 and 31-35 age range compared to all women aged 21-35. By construction, the average woman in the sample has more children than the average Mexican woman. Around 58% (49%) of the group of women aged 21 to 50 (21 to 35) have at least two children, about 32% (25%) have more than two. In line with these differences in observable characteristics, women in the sample have lower employment levels and earnings, conditional on being employed, compared to the broader groups of Mexican women. Moreover, employed women in the sample are much more likely to have an informal job.

3.2 Descriptive statistics

Table 1: Sample characteristics

Demographic characteristics		
Age (yrs)	29.39	(3.83)
Age at first birth (yrs)	19.76	(3.08)
Married	0.91	(0.29)
Literate	0.91	(0.28)
Primary completed	0.62	(0.48)
Secondary completed	0.12	(0.32)
University completed	0.03	(0.17)
Household characteristics		
Rural	0.49	(0.50)
Number of household members	5.41	(1.94)
Maternal parent present	0.10	(0.30)
Paternal parent present	0.08	(0.27)
Monthly non-mother income	3,913	(5,634)
Ownership dwelling	0.80	(0.40)
Fertility characteristics		
Children in household	2.84	(1.05)
More than 2 children	0.53	(0.50)
Same sex	0.50	(0.50)
Two boys	0.26	(0.44)
Two girls	0.24	(0.43)
Age second child (yrs)	6.44	(3.83)
Age youngest child (yrs)	3.74	(3.07)
Labor market outcomes		
Employed	0.25	(0.43)
Labor market outcomes for employed		
Informal employment	0.42	(0.49)
Hours worked per week	37.19	(20.78)
Monthly (Pesos)	2,959	(2,507)
Observations	505,569	

Note: 2010 census data from Mexico with restrictions as outlined in the text. The units of observation are women aged 21–35 with at least 2 children. Statistics depicted are means with standard deviations in parentheses. All variables are measured at the time of the survey. The labor market outcomes refer to the week before the survey.

Descriptive statistics are presented in Table 1. The women in the sample are, on average,

29 years old and gave birth to their first child at the age of 20. The majority of women (62%) completed at most primary school. Only 3% have a university degree. About half of the women live in very small municipalities with less than 2,500 inhabitants. Ten percent live with at least one of their parents in the same household, while 8% share a household with a parent of their partner. The mean monthly labor income of the household, excluding mother's income, averages 3,913 Mexican Pesos (307 US$).[18] For married and cohabiting women, the former is mainly earned by the partner, while for single women by the maternal grandfather. About 2% of the eldest children in my sample work.

Mothers in the sample live with an average of 2.8 children in the same household. About 53% live with more than two children, constituting the group I refer to as the treatment group. The first two children of 50% of the observations in the sample are of the same sex, with two firstborn males being with a proportion of 26% a bit more likely than two females. In the week before the census, 25% of women are working. Of these employed women, 42% work in an informal job, averaging 37 working hours per week with an income of 2,959 Pesos (232 US$) per month.

3.3 Informal employment in Mexico

As noted earlier, the Mexican labor market is characterized by a large share of women who are employed in the informal sector. I use my sample to learn more about informal jobs in Mexico and to discuss the hypotheses stated in Section 2.

My definition of informal employment is defined with respect to the main activity in the week before the census. It comprises own account workers (self-employment with no employees hired), day laborers, and unpaid family workers who work outside home, whereas formal employment includes business owners with employees and blue- and white-collar employees who work for a company or a government for payment.[19] Women with informal jobs are mainly working as store merchants (20.9%), workers in the cultivation of maize/beans (6.0%), domestic workers (5.8%), street vendors (5.0%), and sales workers (4.5%).

In Figure 1, I compare working hours and hourly wages between women in formal and informal employment. This is purely descriptive: women in the formal versus the informal sector might be very different from each other. Panel (a) shows that, on average, women work fewer hours in informal relative to formal employment. Furthermore, the working hours are

---

[18]Given an average exchange rate of 12.75 Mexican Pesos per US$ in 2010.

[19]According to international standards, informal employment is defined as comprising employers and own-account workers employed in their own informal sector enterprises or engaged in the production of goods exclusively for own final use, contributing family workers, and employees holding informal jobs (jobs without social protection and other job-related benefits) both in the formal and informal sector (International Labor Organization, 2013). The National Institute of Statistics and Geography in Mexico defines informal employment based on these international standards (INEGI, 2014). Unfortunately, I can neither observe whether employees are in jobs with social protection and job-related benefits nor whether enterprises are formally registered. In Appendix C I show that my informal employment definition identifies around two thirds of all informal workers from the definition based on the international standard. The workers that are identified as informal with the two different definitions have a similar distribution across occupations and working hours.

Figure 1: Distribution of working hours and wages in formal and informal employment

(a) Working hours   **(b)** Net hourly wages
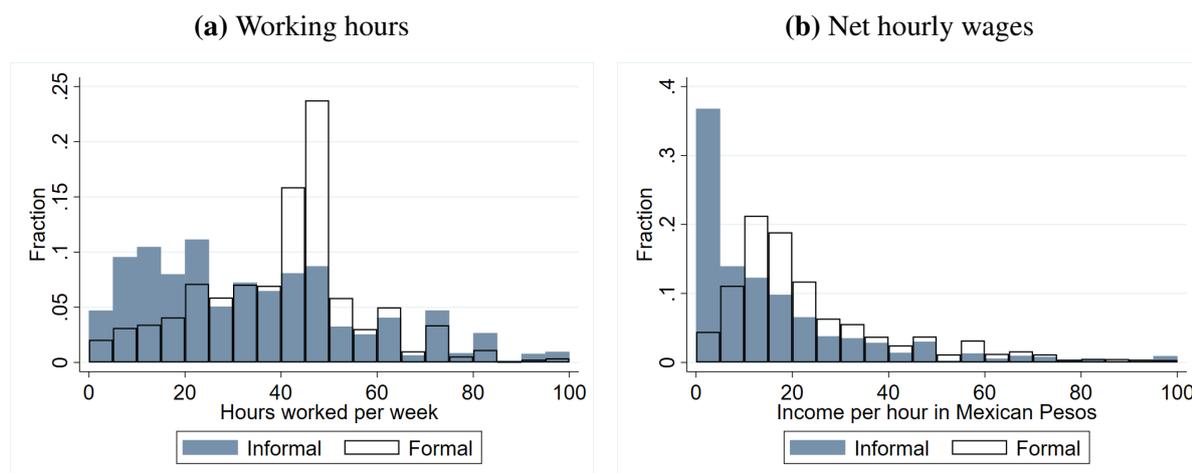
Formal: Mean: 39.7, Median: 40, SD: 17.8   Formal: Mean: 26.5, Median: 18.7, SD: 28.8
Informal: Mean: 33.7, Median: 30, SD: 23.8   Informal: Mean: 20.3, Median: 10.4, SD: 37.8

Notes: 2010 census data from Mexico with restrictions as outlined in the text. The units of observation are women aged 21–35 with at least 2 children. Informal employment is defined as in the text. The graph to the left shows the distribution of working hours and the one to the right the distribution of net hourly wages by type of employment.

much more dispersed for women in the informal sector. This observation suggests that informal employment is more flexible regarding working schedules. Moreover, women with an informal job are much more likely to work between zero and twenty hours, while around 40% of those in a formal job work full-time (40 to 50 hours). This supports the hypothesis that the fixed costs of working are higher in formal than in informal employment. These costs might drive women who want to work relatively few hours out of the formal labor force. Panel (b) indicates that formal pays a higher average hourly net wage than informal employment. Interestingly, more than a third of all informal workers earn only 0 to 5 Mexican Pesos per hour; potentially they receive some payment in kind. To sum up, Figure 1 indicates that women face a trade-off between job quality, as measured by wages, and flexibility when choosing their employment.

4 Estimation strategy

4.1 Research design

The empirical analysis focuses on the effect of an increase in family size on maternal labor market outcomes. I define the treatment indicator D_i equal to 1 if a woman has more than 2 children and 0 otherwise. Let Y_{1i} denote the outcome if i has more than two children and Y_{0i} the outcome in the absence of treatment, i.e., with two children. The outcome observed for i is $Y_i = D_i \times Y_{1i} + (1 - D_i) \times Y_{0i}$. If fertility is randomly assigned, then the difference in the mean outcomes of treated and untreated women would identify the average treatment effect, since both groups were comparable and, thereby, had similar potential outcomes independent of actual treatment assignment.

However, the number of children is potentially endogenously determined. Women decide about the number of children given earning potentials and career plans. Moreover, there probably exist other unobserved factors that influence both fertility and labor market outcomes, f.e., preferences for having children that are correlated with those for working in the labor market. Thus, the average potential outcome for females with two compared to those with three or more children would differ, even if they had the same number of children. Simply comparing the two groups confounds the effect of the treatment with other differences across these groups.

The idea of the estimation method is to use the variation from an instrument that indirectly shifts the number of children while holding other determinants of the outcome variables constant. I exploit parental preferences for a mixed-sex sibling composition to instrument for an increase in family size from two to three or more children. Introduced by Angrist and Evans (1998), this instrument is commonly used in the literature. The rationale behind this strategy is based on previous results showing that, compared to parents who have children of different sexes, parents with children of the same sex are more likely to have an additional child. At the same time, children's sex mix is assumed to be virtually randomly assigned. Z_i is hereafter defined as an indicator that is equal to 1 if the two firstborn children are of the same sex and 0 otherwise.

I use the quasi-random assignment of children's sex composition to estimate the causal effect from an increase in family size on household outcomes. In the main analysis, I estimate the following linear model:

$$D_i = \gamma Z_i + X_i'\theta + \epsilon_i \tag{1}$$

$$Y_i = \beta_i D_i + X_i'\delta + \eta_i, \tag{2}$$

where Y_i, D_i, and Z_i are defined as above. X_i contains relevant control variables, including indicators of the sex of the first and second child.[20] Since I allow the treatment effect β_i to be arbitrarily heterogeneous across individuals, I am only able to identify local effects that pertain to the subpopulation that responds to a change in the value of the instrument, the *compliers*. The target of the estimation is the average of β_i among compliers. The baseline specification uses 2SLS with first and second stage equations given by (1) and (2) to estimate this Local Average Treatment Effect (LATE). However, the outcomes of interest are limited: Employment status is binary and weekly hours worked are non-negative with a mass point at 0. The descriptive statistics in Section 3 show that a substantial fraction – 75% – of women in the sample work zero hours per week. In a model with a non-saturated set of covariates, the Conditional Expectation Function (CEF) for a limited dependent variable is typically non-linear, and it might be

---

[20]The set of control variables includes indicators of the sex of the first and second child, dummies for the age of the mother in yearly categories, dummies for her age at first birth in yearly categories, dummies for the age of the second child in yearly categories, municipality size in categories, dummies for schooling in yearly categories, indicator for literacy, indicator for indigeneity, and state fixed effects. I flexibly control for the age variables, since the outcome variables are potentially non-linear in these age variables and there may be discontinuities at particular ages.

important to use more flexible modeling strategies (Angrist, 2001). Therefore, I additionally implement an estimator for the LATE that incorporates the covariates X_i in a fully nonparametric way (Frölich, 2007). For this nonparametric IV estimator (*NP-IV*), the conditional mean outcome is estimated via a local linear estimator and the conditional mean treatment via a local logit estimator. Thereby, I avoid the functional form restrictions on the conditional expectation functions of D_i and Y_i in the linear model.

The average treatment effect on working hours might hide important heterogeneity in the labor supply behavior of households. The overall effect comprises both the impact of family size on the decision to work and on the volume of work conditional on employment. The former refers to the extensive, the latter to the intensive margin of labor supply. In order to investigate the impact of fertility on the distribution of hours worked, I compare quantiles of the potential outcome distributions among compliers. For this purpose, I apply the local quantile treatment effect (LQTE) estimator developed by Frölich and Melly (2013). The identification of LQTE does not require any functional form restrictions and relies on the same assumptions necessary to identify LATE. Thus, the framework naturally accommodates discrete outcomes and outcomes with mass points (Melly and Wüthrich, 2017).[21] In contrast, alternative models that deal with limited outcome variables require further identification assumptions and their interpretation is often not straightforward (Angrist, 2001). Similar to the *NP-IV* estimator, I can flexibly include covariates in the model in order to relax the identification assumptions.

4.2 Identification assumptions

In the following, I discuss the assumptions that are necessary for the validity of the above-described instrumental variable strategy. These assumptions are introduced by Imbens and Angrist (1994) and I refer to them as the LATE assumptions.

I must assume that the sex composition of the two firstborn children is as good as randomly assigned. This assumption might be violated if parents can influence the sex mix of the two firstborn children, for example by sex-selective abortions or discrimination in care practices. Sex ratios of children aged zero to four that are close to the biological ones speak against the former (Cruces and Galiani, 2007). Moreover, sex-selection technologies are very expensive (Dahl and Moretti, 2008) and abortions themselves carry high costs, given that access to abortion is prohibited or severely restricted in most of Mexico (OECD, 2017). It seems implausible that families have such extreme preferences for a particular sex composition of their children that they are willing to bear the burden of a sex-related abortion. In general, family institutions in Mexico do not exhibit extreme preferences for male children as observed in some Asian countries by Schultz (2008): For instance, there is no dowry custom and no systematic discrimination against girls (Cruces and Galiani, 2007).

---

[21]However, the well-behaved asymptotic distributions for the LQTE estimator rely on the continuity of the dependent variable. This implies that the inference procedures are not valid in the neighborhood in which the dependent variable has a mass point, i.e., at 0 (Melly and Wüthrich, 2017).

Table 2: Testing for the quasi-random assignment of children's sex composition

	Mean		Difference in Means	
	Mixed [$Z_i = 0$]	Same [$Z_i = 1$]	Coefficient	SE
	(1)	(2)	(3)	(4)
I. Predetermined mother characteristics				
Age (yrs)	29.408	29.381	0.027**	(0.011)
Age at first birth (yrs)	19.765	19.760	0.005	(0.009)
Age of second child (yrs)	6.440	6.443	-0.003	(0.011)
Literate	0.911	0.912	-0.001	(0.001)
Primary completed	0.622	0.624	-0.002	(0.001)
Secondary completed	0.119	0.120	-0.001	(0.001)
University completed	0.029	0.029	0.000	(0.000)
Years of schooling	7.333	7.342	-0.009	(0.011)
Indigenous	0.343	0.342	0.001	(0.001)
Locality size [2,500-14,999]	0.242	0.241	0.001	(0.001)
Locality size [15,000-99,999]	0.125	0.127	-0.002	(0.001)
Locality size [100,000+]	0.140	0.140	0.000	(0.001)
II. Sex of the two firstborn children				
First child boy	0.505	0.518	-0.013***	(0.001)
Second child boy	0.495	0.518	-0.022***	(0.001)
Observations	251,305	254,264		

Note: This table illustrates the balancing of variables by the value of the instrument. Columns (1) and (2) show the mean of the corresponding variable among households with a mixed and same sex composition of the two firstborn children, respectively. The coefficient and SE in column (3) and (4), respectively, are obtained by regressing the corresponding variable on the instrument dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel I in Table 2 illustrates the correlation between predetermined characteristics of women and their instrument status. Columns (1) and (2) give the means of the corresponding variable in households with two children of different sexes and the same sex, respectively. Columns (3) and (4) show the mean difference between the two groups and its standard error, respectively. Mothers with children of mixed sexes are remarkably similar in almost all observable characteristics to those with children of the same sex. Still, there is a small but statistically significant age difference between the two groups. In the empirical analysis, I examine whether the results are sensitive to the inclusion of a large and varying set of pre-determined control variables. If the sex composition is randomly assigned, then the estimates should not change.

My instrumental strategy exploits the sex composition of the two firstborn children, i.e., whether the sex of the first child matches the sex of the second child, rather than the sex of the first and second child. However, in the absence of manipulation, the natural sex ratio at birth is, with 105-107 male for every 100 female births, biased toward boys (Hesketh and Xing, 2006). This implies that having a male child among the two firstborn children makes it more likely that a couple has two firstborn children of the same sex, i.e., the two firstborn children's sex is systematically related to the *same sex* instrument. This is illustrated in Panel II in Table 2:

parents with children of the same sex have, on average, a significant 1.3 percentage points (2.2 percentage points) higher likelihood of a male first (second) child.

This needs to be taken into account because the sex of the first and second child is likely related to potential fertility and potential labor market outcomes. Dahl and Moretti (2008) and Ruiz and Campos-Vazquez (2013) show that having a male firstborn affects subsequent fertility behavior by reducing the likelihood of having additional children in the US and Mexico, respectively. Moreover, children's sex is related to several variables that likely have an impact on maternal labor market outcomes. First, Dahl and Moretti (2008) and Ruiz and Campos-Vazquez (2013) find a relationship between having a firstborn boy and household structure. Second, there might be differences in the monetary or time cost of raising girls versus boys (Angrist and Evans, 1998; Dahl and Moretti, 2008). Third, the likelihood of a male birth varies positively with health conditions during pregnancy (Almond and Mazumder, 2011). By including a dummy for the sex of the first and a dummy for the sex of the second child in the empirical models, I can capture these direct effects of the sex of the two firstborn children. Conditional on these controls, they should not pose a problem for my identification assumptions.

Another potential concern is that many poor households receive subsidies through the *Oportunidades* program. The amounts that can be received depend on the sex of the children in the household. Hence, there might be direct effects of children's sex on maternal employment through differences in the subsidies, for example due to an income effect or a change in maternal decision power within the household (Attanasio and Lechene, 2002). However, conditional on the sex of the two firstborn children, the sex composition does not impact the amount of subsidies received and hence, the program should not induce a violation of my identification assumptions.

To interpret the IV estimates as identifying the causal effect of fertility on parents' labor market outcomes, I must further rule out that the sex composition of the two firstborn children has a systematic effect on labor supply other than through its effect on having additional births. This might be violated if the sex composition affects child costs. Lower levels of expenditures can result from hand-me-downs (e.g., for clothing and footwear) when children have the same sex. Wolpin and Rosenzweig (2000) demonstrate that these savings exist for Indian households and that they account for a substantial fraction of the household income. Nevertheless, these results might vary significantly across countries. Bütikofer (2011), examining the Mexican Family Life Survey in 2000 and 2005, finds no statistically significant differences in the economies of scale across households with different sibling sex compositions due to clothing- and room-sharing. Even if savings related to the sex composition existed and affected female labor supply directly, their size would have to be meaningful enough to violate the exclusion restriction severely. Data on household expenditures in Mexico suggest that the fraction of household income spent on clothing and footwear for all members of the family in 2010 is, at 2.3%, very small.[22] The exclusion assumption might also fail to hold if the sex composition of

---

[22]Data is from OECD StatExtracs/Final consumption expenditure of households,

the two firstborn children directly affects the marginal utility of parents' leisure and, thus, labor supply. Conditional on the sex of the first- and second-born children, assuming separability of sex-sameness and parental leisure does not seem to be too restrictive.

The last assumption needed for a causal interpretation of the IV estimates is the monotonicity of the *same sex* instrument. Monotonicity requires that while having mixed-sex siblings compared to same-sex siblings may not affect the fertility of some women, all of those affected are assumed to be affected in the same direction. It fails, for instance, if there are mothers who have a preference for at least two children of the same sex and, thereby, choose to have a third child if the first children have different sexes. Lee (2008), for example, shows that parents with two sons are less likely to continue childbearing than parents with one son and one daughter in South Korea. Wolpin and Rosenzweig (2000) find a similar pattern in India. However, both of these countries are well known for their sex bias due to extreme son preferences. In Appendix Section D, I analyze two data sets on Mexico, the National Survey of Demographic Dynamics and the Mexican Family Life Survey, to learn about fertility preferences in my setting. First, around 3 percent of women who have more than two children, but who claim that their ideal number is two, say that this was because "they were wishing for a child of the opposite sex than the children they already had." The data does not show that any women stated that they had more children than desired because they wanted two children of the same sex. Second, among women who want to have two children but have not reached this number yet, 79.2 percent would like to have one boy and one girl, 4.0 percent two boys, and 4.7 percent two girls. The remaining are indifferent. Hence, the share of women who either have a preference for two children of different sexes or who are indifferent is one order of magnitude larger than the share of women who desire two boys or two girls only. Unfortunately, these two data sets have no information on mothers' desired sex composition if they already have at least two children. Hence, I cannot infer how the realized fertility relates to the stated preferences. Therefore, I provide complimentary evidence from mothers in the Demographic and Health Survey data on Colombia who share similar characteristics to Mexican mothers. Among women whose first two children are a boy and a girl, 0.77 percent have three children or more and retrospectively declare that their ideal sex composition would have been two boys and no girl or no boy and two girls.[23] In comparison, among women whose first two children are of the same sex, 12.7 percent have at least three children and retrospectively declare that their ideal sex composition would have been one boy and one girl. Relating these shares to the share of women who have a preference for two boys and two girls and for a mixed sex composition, respectively, it does not seem to be the case that the realized fertility is more responsive to unmet preferences of the former group. To sum up, the share of women who have fertility preferences that might induce them to act as defiers with respect to my instrument is very small, especially in comparison to the share of women with preferences who might induce them to act as compliers. Moreover, it

---

`http://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE5`; last access 09/2017.

[23] This is similar to the analysis of De Chaisemartin (2017) for Peru.

does not seem to be the case that the realized fertility is more responsive to unmet preferences of women with two boys and two girls compared to those with mixed sex preferences.[24]

The underlying assumptions behind the causal interpretation of my results have testable implications. This is exploited in a test developed by Huber and Mellace (2015).[25] The intuition behind the test is that jointly assuming independence, exclusion, and monotonicity, the mean potential outcomes of always-takers under treatment and of never-takers under non-treatment are both point-identified and bounded. The fact that the point-identified moments have to lie within the respective bounds provides testable restrictions. A rejection of the null hypothesis points to the violation of independence, exclusion, monotonicity, or any combination of them. However, it is important to emphasize that non-rejection does not imply the validity of the assumptions. There exist violations of these three assumptions that do not cause violations of the testable restrictions and, hence, the test has no power to detect them. Taking these caveats into consideration, I apply Huber and Mellace (2015)'s test to my large data set. In Appendix Table D3 I show the test results for the full sample as well as for subsamples defined by covariates such as the sex of the first and the second child, education, rural status, and their combinations. For the full sample, I obtain a p-value close to one. In general, I cannot reject the hypothesis that the instrument satisfies the assumptions for validity at any conventional significance level.[26]

To sum up, the discussion based on empirical observations for the Mexican context and the formal test procedure provide no evidence that the *same sex*-instrument is invalid.

5 Empirical results

To estimate the effect of fertility on household labor market outcomes, I follow the instrumental strategy described in Section 4. I begin by examining the first stage relationship in Table 3. Column (1) suggests that there is a statistically significant relationship between fertility and the samesex-indicator: Women with two children of the same sex are 3.3 percentage points more likely to have an additional child relative to those with children of distinct sexes. This coefficient does not change if I add a large set of control variables in column (2). The significantly negative effects of *boy1* and *boy2* indicate that women stop having more children in response to having a boy. The model in column (3) splits up the $same\ sex$ instrument in two indicators that are equal to one if the first two children are boys or girls, labeled *twoboys* and *twogirls*, respectively. The estimated coefficients suggest that females increase childbearing both in response to the birth

---

[24]Angrist et al. (1996) show that even if non-monotonicity does not hold, it is still possible to capture the local treatment effect of compliers if defiers and compliers have the same treatment effects. This is not too unreasonable in this setting, because selection into being a defier or a complier is due to parents' sex preferences and not due to gains from treatment (De Chaisemartin, 2017).

[25]See Machado et al. (2019) for an alternative testing procedure on the validity of the LATE assumptions and the sign of the average treatment effect.

[26]Huber (2015) uses the same testing procedure on the validity of the *same sex* instrument for a sample of mothers in the 1980 US Census and Mourifié and Wan (2017) apply a different testing procedure in the 1990 US Census. Both studies find no evidence that the assumptions for the validity of the *same sex* strategy are violated.

of two boys and two girls, but the response to two boys is significantly smaller than to two girls (2.6 compared to 4.1 percentage points, respectively). In column (4), the association between having first a male child and reduced childbearing at higher parities turns insignificant when the *twoboys* and *twogirls* indicators are entered separately.[27] This means that there is no relationship between *boy1* and fertility when the effect of sex composition is allowed to differ by sex.

Table 3: First stage relationship: Fertility and children's sex composition

	Dependent variable: More than 2 children			
	(1)	(2)	(3)	(4)
samesex	0.033***	0.033***		
	(0.001)	(0.001)		
boy1		-0.009***		-0.002
		(0.001)		(0.002)
boy2		-0.007***		
		(0.001)		
twoboys			0.026***	0.026***
			(0.002)	(0.002)
twogirls			0.041***	0.040***
			(0.002)	(0.002)
Control variables		✓		✓
Dependent mean		0.529		
Observations		505,569		
Partial R^2 (excluded Z_i)	0.0011	0.0016	0.0012	0.0016
F(excluded Z_i)	558.55	802.75	293.65	418.09

Note: This table reports the first stage coefficients based on equation 1. Columns (1) and (2) use as instrument an indicator whether the two firstborn children are of the same sex. Columns (3) and (4) use as instruments indicators whether the two firstborn children are boys or girls. F(.) reports the robust Kleibergen-Paap Wald rk F-statistic on the excluded instruments. Control variables include indicators for the sex of the first and second child, dummies for the age of the mother in years, dummies for her age at first birth in years, dummies for the age of the second child, municipality size in categories, dummies for schooling in years, indicator for literacy, indicator for indigeneity, and state fixed effects. In column (4), *boy2* is excluded from the set of control variables because of multicollinearity. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To sum up, the first stage results suggest that women in the sample have a preference for a balanced sex mix combined with a bias for boys. The latter should not be of concern *per se*. There is no evidence for strong discrimination against girls and the sex mix of children is unlikely to affect the outcome through a channel other than family size. Hence, it seems likely that the bias for boys does only reflect cultural preferences that do not challenge the identification assumptions. Furthermore, I show results that control for the sex of the first two children. The F-Statistics on the excluded instruments are well above the corresponding critical values, and I can reject the null that the instruments are weak in all specifications. The instrument is subsequently used in order to estimate the effect of an exogenous increase in family size from two to three and more children on mothers' labor market outcomes.

---

[27]The second model allows the effect of having first two boys and having first two girls to differ. In order to avoid linear dependence of the included indicators in this model, either $boy1$ or $boy2$ must be dropped.

5.1 Fertility and mothers' labor supply at the extensive margin

Table 4: Relationship between fertility and mother's employment

	Mean	OLS		2SLS		2SLS+	NP-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:							
Employed	0.248	-0.082***	-0.076***	0.043	0.048	0.063**	0.042
		(0.001)	(0.001)	(0.037)	(0.031)	(0.029)	(0.030)
Formally employed	0.143	-0.080***	-0.058***	0.004	0.005	0.011	0.001
		(0.001)	(0.001)	(0.029)	(0.027)	(0.026)	(0.028)
Informally employed	0.105	-0.002***	-0.018***	0.038	0.043*	0.052**	0.041*
		(0.001)	(0.001)	(0.025)	(0.025)	(0.024)	(0.023)
Control variables			✓			✓	✓	✓
Observations				505,569			

Note: This table illustrates the relationship between mothers' employment status and the treatment. Each row shows the results for the stated outcome variable. Column (1) shows the mean of the dependent variable. Columns (2) and (3) show the estimated coefficients on the treatment in an OLS regression without and with control variables. Columns (4) and (5) show the 2SLS regression results using *samesex* as instrument without and with control variables and (6) using *twoboys* and *twogirls* as instruments plus control variables. (7) shows the results from the nonparametric IV estimation with *samesex* as instrument. *Employed* is an indicator equal to 1 if a woman is employed in the week prior to the survey and 0 otherwise. *Formally (informally) employed* is an indicator equal to 1 if a woman is employed in a formal (informal) job in the week prior to the survey and 0 otherwise. Control variables are as in Table 3. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table 4, I present the results from OLS and IV estimations of the relationship between maternal employment and fertility. The OLS estimates in column (2) illustrate that women with more than two children have a weaker labor market attachment than those with two children: They are an average of 8 percentage points less likely to be employed, which is mainly driven by lower employment in the formal sector. Including a set of control variables in column (3) indicates that only a small amount of this negative correlation is driven by differences in observable characteristics across these two fertility groups. In contrast, effects based on the IV models in columns (4) to (7) reveal that the effect of having more than two children on maternal employment is non-negative at the 10% significance level. This definition includes both employment in formal and informal jobs. In the following, I look at them separately. Employment in the informal sector increases by 4.3 percentage points when fertility increases beyond two children. This effect is statistically significant at the 10% level in all specifications and corresponds to a 17% increase relative to the average employment level in the sample. In contrast, there is no effect on employment in the formal sector. In Table 3, I show that mothers in my sample tend to have more children when their two firstborn are of the same sex independent of whether they are boys or girls. In column (6), I split up the samesex-instrument into the two separate instruments, *twoboys* and *twogirls*. The coefficients in these alternative 2SLS estimations increase slightly, and standard errors become a little bit smaller, but they are very similar to before. The coefficients in the nonparametric IV estimations in column (7) are slightly smaller than in the

linear models. These results are robust to the inclusion of different sets of control variables and to different sample restrictions regarding maternal age (see Appendix Tables D4 and D5).

The comparison of the OLS and IV estimates suggests that the preference for further childbearing are inversely related to working in the labor market, particularly in the formal labor market. An exogenous increase in childbearing beyond two does not have an impact on the probability to be formally employed, but it does increase the propensity of informal employment. As discussed earlier and in line with previous results, this might happen because informal jobs offer more flexibility to combine work and family responsibilities. In the following, I look at the distributional impact of higher parity childbirth on hours to allow for different responses at the extensive and the intensive margin of labor supply.

5.2 Fertility and maternal working hours responses

I estimate the quantiles of the potential outcome distributions for compliers based on Frölich and Melly (2013) using the samesex-instrument.[28] Figure 2 illustrates the estimated quantiles of maternal working hours separately for formal (a) and informal employment (b).[29] The blue, solid (red, dashed) line shows the estimated quantiles of working hours for compliers with more than two children (with only two children). The difference between the two lines corresponds to the local quantile treatment effect at the τ quantile. Appendix Table D7 quantifies the results and provides the corresponding standard errors. First, if they had only two children, a large fraction of women, around 85% and 90%, would have zero working hours in formal and informal employment, respectively. Second, increasing fertility beyond two children does not have an impact at the extensive margin of formal work, but it seems to have a positive effect at the intensive margin by shifting a few quantiles from around 20 to 40 hours. The second panel in Appendix Table D7 shows that this effect is statistically significantly different from zero. Third, having more than two children increases some quantiles that would be equal to zero or to a small number of hours in informal employment if women had two children only. The quantiles are augmented by statistically significant 13 to 15 hours at this part of the outcome distribution (see the third panel of Appendix Table D7). Summing up, fertility increases maternal employment in the formal sector at the intensive margin by shifting a few quantiles from around 20 to 40 hours, while having mainly an impact at the extensive margin in the informal sector.

---

[28]Appendix Table D6 shows the mean effects of having more than two children on hours worked per week. The coefficients are positive when instrumenting for the potentially endogenous fertility decision, but they are not statistically distinguishable from 0 in almost all specifications.

[29]The effect on hours in both formal and informal employment is presented in Appendix Figure D1.

Figure 2: Distributional impact of fertility on mother's hours worked per week
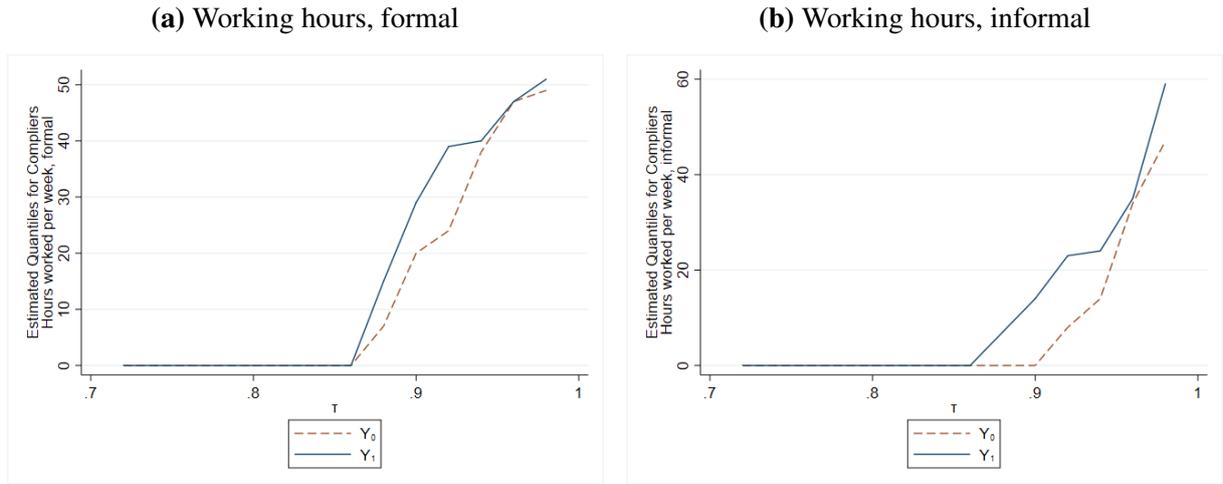
(a) Working hours, formal	**(b)** Working hours, informal

Notes: This graph illustrates the estimated quantiles for the potential outcome distributions of maternal weekly working hours in formal (top) and informal jobs (bottom). The blue, solid and the red, dashed line show the estimated quantiles of working hours for compliers with more than two children, $\hat{Q}^{\tau}_{Y_1|c}$, and with only two children, $\hat{Q}^{\tau}_{Y_0|c}$, respectively. The model includes the control variables that are listed in Table 3.

6   Mechanisms

The estimated employment effects contrast with findings from several other studies. I already show that having more than two children increases maternal informal employment. Employment opportunities in the informal sector tend to be more prevalent in low- and middle-income compared to high-income countries, potentially explaining some differences in the results. In the following, I discuss three more potential mechanisms behind my findings. First, I show that husbands in couples do not increase their employment or working hours in response to an increase in their family size. Second, I discuss a change in the provision of informal child care within extended households that favors monetary relative to time investments by the mother. Third, I look at the response of women conditional by different proxies of household wealth.

6.1   Labor supply responses of married and cohabiting couples

Most women in my sample live together with their partners.[30] Table 5 shows the estimation results for the subsample of married and cohabiting couples.[31] For employment, the coefficients on having more than two children are similar for the sample of women who live with their partners compared to those in the overall sample. The response in maternal informal employment and working hours is positive and statistically significant for partnered women at the 5% and 10% significance level, respectively.

---

[30]For 5.1% of the women who report being married or cohabiting, I do not observe their partner in the data. The cohabiting partner of a mother might not necessarily be the biological father of her children. In the following, "father" refers to the biological father, stepfather, or adoptive father of the children of the women in the sample.

[31]Descriptive characteristics for the subsample can be found in Appendix Table D2.

Table 5: Relationship between fertility and maternal and paternal employment

	Mean	OLS		2SLS		2SLS+	NP-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:							
Mother employed	0.211	-0.070***	-0.056***	0.033	0.043	0.052	0.035
		(0.001)	(0.001)	(0.037)	(0.034)	(0.034)	(0.035)
Mother employed, formal	0.114	-0.068***	-0.041***	-0.019	-0.011	-0.010	-0.019
		(0.001)	(0.001)	(0.028)	(0.026)	(0.026)	(0.027)
Mother employed, informal	0.097	-0.002***	-0.015***	0.052**	0.055**	0.062**	0.054**
		(0.001)	(0.001)	(0.026)	(0.026)	(0.026)	(0.027)
Mother hours worked	7.531	-2.793***	-2.316***	2.345	2.732*	3.017*	2.410**
		(0.054)	(0.063)	(1.565)	(1.494)	(1.472)	(1.279)
Father employed	0.903	-0.017***	0.000	-0.006	-0.004	-0.004	-0.004
		(0.001)	(0.001)	(0.026)	(0.026)	(0.026)	(0.023)
Father hours worked	43.319	-2.561***	-0.117	0.152	-0.178	-0.520	-0.036
		(0.071)	(0.084)	(2.038)	(1.965)	(1.936)	(2.109)
Control variables			✓			✓	✓	✓
Observations				412,452			

Note: This table illustrates the relationship between a set of employment outcomes for parents and the treatment similar to Table 4. The sample is restricted to households in which both parents are present in the household at the time of the survey and in which employment outcomes for both of them are observed. Control variables are as in Table 3. Robust standard errors in parentheses. $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$

Column (1) shows that fathers have, on average, a much stronger labor market attachment than mothers both in terms of employment and working hours: 90% of all fathers are employed, and they work an average of 43 hours per week. Column (2) illustrates that fathers with more than two children have worse labor market outcomes, on average, than those with two. In contrast to women, these differences in labor market outcomes are to a large extent explained by differences in observable characteristics, which I include in the regressions in column (3). In contrast, there is no effect of having more than two children on paternal employment in the IV estimations and the coefficients are close to zero.

In a model with a unitary household, we would expect that the partner with the higher wage worked unambiguously more when having additional children because he or she had a comparative advantage in providing earnings for consumption and monetary inputs. The other partner would work more inside the home because his or her opportunity costs were lower (Heath, 2017). Surprisingly, the results here indicate that mothers increase their labor supply, while fathers do not react at all. Although wages are unobserved for most women and some men, one would probably expect that potential wages are higher men.[32] One potential explanation for the contrary results are labor market frictions that were larger for men than for women. Around 2010, labor market outcomes of Mexican men were hit substantially by the Great Recession. At that time, the labor market environment was more favorable for women and, as a result, women

---

[32]Only 5% of women in the sample earn more than their partners in terms of their hourly wages among all couples in which both partners work.

were increasingly responsible for the provision of financial resources to the family (World Bank, 2012). Another explanation is that the men's labor supply is less elastic.

6.2 Informal childcare through extended families

Another potential explanation for the positive effect of fertility on employment is that the availability of informal child care provided by the extended family makes time investments less attractive relative to monetary investments. Hence, they raise the incentives to increase labor supply and focus on monetary inputs. Several studies show that the availability of grandparent-provided childcare significantly increases maternal labor supply (e.g., Gong and van Soest (2002), Compton and Pollak (2014)).

In the data, I can observe whether the parents or parents-in-law are present in households. Women who live in the same household as their parents (or parents-in-law) tend to be different from those who live only with their nuclear family. For instance, co-residence is correlated with being a single mother. Moreover, co-residence with parents might be itself determined by fertility: Mothers (and their spouses) might decide to join their parents' household, or parents might choose to live with their children in order to offer them childcare when their family size increases. In order to examine the latter, I use the same IV estimation strategy as before, putting an indicator of whether grandparents are present in the household as the dependent variable. I assume here that the sex composition of the two firstborn children does not affect co-residence with grandparents other than through its impact on fertility.

Table 6: Relationship between fertility and household composition

	Mean	OLS		2SLS		2SLS[+]	NP-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:							
Maternal grandparent present	0.100	-0.064***	-0.040***	0.053**	0.046*	0.044*	0.046
		(0.001)	(0.001)	(0.026)	(0.025)	(0.025)	(0.031)
Paternal grandparent present	0.076	-0.019***	-0.000	0.011	0.005	0.004	0.005
		(0.001)	(0.001)	(0.022)	(0.022)	(0.022)	(0.021)
Control variables			✓		✓	✓	✓
Observations				505,569			

Notes: This table illustrates the relationship between the household composition and the treatment similar to Table 4. $Maternal\ (paternal)\ grandparent\ present$ is an indicator equal to 1 if at least one maternal (paternal) grandparent lives in the same household as the mother and her children. Control variables are as in Table 3. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 illustrates the relationship between fertility and the presence of grandparents in the household. On average, mothers with two children are less likely to live with either their or their partners' parents. When I instrument for the potential endogeneity, there is a marginally significant positive relationship between having more than two children and co-residence with maternal grandparents. The 4.6 percentage points increase is considerable given that 10 percent

of the women in the sample live in the same household as their parents. Unfortunately, there is no information on child care in the data set, so it is not possible to test the hypothesis that the increased likelihood to live with grandparents allows women to work more when they have more children. Still, this pattern is consistent with the hypothesis.

6.3   Household wealth

Negative budgetary consequences from fixed costs of children are likely to be more severe in households with low wealth. As a result, an increase in the family size of a "poor" households should have a larger impact on the mother's propensity to take up employment. I propose three different measures to proxy household wealth. The first one is the number of rooms in the house, the second is a mixed wealth measure generated using a principal component analysis using several characteristics of the households' dwelling (such as ownership of a dwelling, car, TV, or radio, as well as the type of electricity used and water supply), and the third is based on spousal earnings. Importantly, I do not find an effect of the instrumented treatment on any of these three variables.[33]

Table 7: Effect of fertility on maternal employment by household wealth

	Employment		Hours		Obs.
	Formal	Informal	Formal	Informal	
	(1)	(2)	(3)	(4)	(5)
Household wealth measure:					
1. Number of rooms					
3 or less	-0.001	0.062*	1.019	2.945**	327,123
	(0.036)	(0.036)	(1.589)	(1.502)	
More than 3	0.021	0.029	0.151	0.238	174,833
	(0.044)	(0.038)	(1.968)	(1.728)	
2. Mixed wealth measure					
Low	0.009	0.032	1.436	1.803	255,249
	(0.040)	(0.047)	(1.782)	(1.864)	
High	0.001	0.050	0.100	1.265	250,320
	(0.037)	(0.030)	(1.652)	(1.399)	
3. Husband's earnings					
Low	-0.025	0.068	-1.638	3.876**	197,536
	(0.031)	(0.042)	(1.260)	(1.757)	
High	0.004	0.048	1.716	1.457	218,110
	(0.039)	(0.033)	(1.636)	(1.476)	

Notes: This table illustrates the relationship between maternal employment and hours with the treatment by different household measures. Columns (1) to (4) show the estimated coefficients on the treatment in a 2SLS regression using $samesex$ as instrument and including the control variables listed in Table 3. The mixed wealth measure is generated through a principal component analysis based on different dwelling characteristics. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The corresponding results from the second stage using the samesex-instrument are presented

---

[33]See Appendix Table D8.

in Table 7. The findings provide suggestive evidence that mothers in households with relatively low household wealth particularly increase their labor supply in response to an increase in their fertility beyond two children. There is no statistically significant effect of larger family size on formal employment or hours for any subgroup. In contrast, the positive effects on informal employment are larger and more likely to be statistically significant among women in low-wealth households.

These findings are in line with the idea that low-wealth households gain a particularly high value from additional labor income when their family size increases.

7 Beyond local treatment effects

So far, I estimated and discussed treatment effects for the group of compliers. The *internal validity* of these estimates hinges on the assumptions discussed in Section 4, namely conditional random assignment, exclusion, and monotonicity. In the following, I discuss the *external validity* of my results. To be more precise, I show whether the estimation strategy allows for drawing inferences about treatment effects for individuals in the sample other than those affected by the instrument.

The following discussion is based, in large part, on the concept of *marginal treatment effects* introduced by Heckman and Vytlacil (1999, 2005).[34] Under the conditional exogeneity of the instrument Z_i, the monotonicity condition introduced in Section 4 is equivalent to the existence of a weakly separable treatment selection equation $D_i = \mathbb{1}[p(X_i, Z_i) - U_i \geq 0]$, where $p(X_i, Z_i)$ is defined as the probability of receiving treatment based on X_i and Z_i, $P[D_i = 1|X_i, Z_i]$, and where U_i is a measure of the unobserved distaste for treatment that is uniformly distributed over $[0, 1]$ conditional on X_i (Vytlacil, 2002). As $P(.)$ increases, individuals with successively higher unobserved distaste for treatment select into treatment. Treatment effects can vary across households with the same X_i in a way that depends on the unobservable component of the treatment choice U_i. The marginal treatment effect, $MTE(u, x) = E[Y_{1i} - Y_{0i}|U_i = u, X_i = x]$, can be interpreted as the treatment effect for individuals who are indifferent to treatment if they are exogenously assigned a value of Z_i such that their propensity score $P(.)$ is equal to u. For a U_i close to zero, the MTE measures the effect of treatment on individuals with unobservables that make them most likely to participate in treatment. If U_i is large, $P(.)$ would have to be large to induce people to participate. The $MTE(u, x)$ is equal to $MTO(u, x) - MUO(u, x)$, where MTO refers to the marginal treated outcome and MUO to the marginal untreated outcome.

Subsection 7.1 is carried out maintaining the LATE assumptions from Section 4 only, while Subsection 7.2 relies on further assumptions that are subsequently introduced and discussed.

---

[34]This paragraph additionally builds on a survey by Cornelissen et al. (2016).

7.1 External validity under the LATE assumptions

In the following, I group individuals according to their unobserved distaste for treatment U_i. Since Z_i and D_i are both binary, I can partition the population into *three types*, T_i (Angrist et al., 1996).[35] *Always-takers* have low resistance for having additional children: they choose to have three or more children irrespective of their previous children's sex composition. *Compliers* have an intermediate distaste for having more than two children: They are treated when the instrument is switched on and abstain from it when not, i.e., they have more than two children if the two firstborn children are of the same sex and otherwise not. *Never-takers* have a high resistance for having additional children. This means that always-takers select into treatment before compliers who themselves select into treatment before never-takers. More formally, define $u_1 \equiv P(D_i = 1|Z_i = 1)$ and $u_0 \equiv P(D_i = 1|Z_i = 0)$. Then, always-takers are individuals with an unobserved distaste for treatment $0 < U_i < u_0$, compliers with $u_0 < U_i < u_1$, and never-takers are individuals with $u_1 < U_i < 1$. The previously discussed $LATE$ parameter is a version of the MTE and can be expressed as $LATE = E[Y_{1i} - Y_{0i}|u_0 < U_i < u_1]$.

Table 8: Compliance types: Estimated shares & fraction employed

Compliance type (T)	Fraction ($\hat{\phi}$)	$E[\widehat{Employed}_1|T]$	$E[\widehat{Employed}_0|T]$
	(1)	(2)	(3)
Always-taker	0.512	0.208	
	(0.001)	(0.001)	
Complier	0.033	0.259	0.220
	(0.001)	(0.020)	(0.030)
Never-taker	0.455		0.294
	(0.001)		(0.001)

Note: This table shows the estimated share of the compliance types in the sample ($\hat{\phi}$) and their estimated average potential employment levels with and without treatment.

The compliance type of an individual cannot, in general, be identified. However, under independence (random assignment and exclusion) and monotonicity some always-takers, individuals observed with $Z_i = 0$ and $D_i = 1$, and some never-takers, individuals observed with $Z_i = 1$ and $D_i = 0$, can be detected. The percentage of never-takers, ϕ_n, is equal to $P(D_i = 0|Z_i = 1)$, that of always-takers, ϕ_a, is equal to $P(D_i = 1|Z_i = 0)$, and the share of compliers can be deduced by $\phi_c = 1 - \phi_n - \phi_a$ (Imbens and Rubin, 1997). Column (1) in Table 8 shows the estimated share of the different compliance types among the households in the sample. The estimated share of compliers is, at 3%, very low. Never-takers are estimated to be 46% of the women; 51% are always-takers. The small number of compliers raises concerns

---

[35]In the following, I abstract from covariates X_i. This is under the more restrictive assumption of unconditional, rather than conditional, independence and exclusion. This is supported by the fact that my previous results are invariant to the inclusion of control variables.

about the external validity of the previously estimated local treatment effects since I estimate the effect over a very narrow range of the distribution of U_i.

Table 9: Mean of observable characteristics by types

	Always-taker	Complier	Never-taker
Age (yrs.)	30.02	30.21	28.63
	(0.01)	(0.16)	(0.01)
Age at first birth (yrs.)	19.03	19.75	20.59
	(0.01)	(0.11)	(0.01)
Age second child (yrs.)	8.26	7.02	4.35
	(0.01)	(0.11)	(0.01)
Indigenous	0.40	0.24	0.28
	(0.00)	(0.20)	(0.00)
Years of schooling	6.34	7.72	8.35
	(0.01)	(0.14)	(0.01)
Literate	0.87	0.94	0.95
	(0.00)	(0.01)	(0.00)
Rural	0.56	0.47	0.42
	(0.01)	(0.22)	(0.02)

Notes: This table reports average characteristics of the different compliance types in the sample. The standard errors are based on a bootstrap method and obtained with the STATA command *ivdesc*.

In order to investigate the possibility that the LATE is representative for the treatment effect of non-compliers, I exploit some more information that can be obtained using the LATE assumptions (Abadie, 2003; Kowalski, 2016). Using independence, the average characteristics of always-takers, compliers, and never-takers can be identified. If they were the same across all types, then this would suggest that the LATE is indicative for the treatment effects of non-compliers. Table 9 shows the mean observable characteristics for the different groups in the sample. The average complier in my setting seems to be different from the average never-taker and always-taker in terms of her pre-determined characteristics. Never-takers are an average of about 1.5 years younger than always-takers and compliers. The mean age of mothers at first birth and of their second-born children is increasing and decreasing, respectively, when going from always-takers to compliers to never-takers. Education in terms of years of schooling is increasing in the unobserved distaste for more children. Furthermore, always-takers are more likely to live in rural areas and to be indigenous compared to compliers and never-takers. These differences in observable characteristics that are likely related to potential employment outcomes suggest that treatment effects might be heterogeneous across U_i.

Even though it is not possible to identify treatment effects for never- and always-takers, I can identify the expectation of Y_{1i} for always-takers, of Y_{0i} for never-takers, and of both Y_{1i} and Y_{0i} for compliers (Imbens and Rubin, 1997). These outcomes can be informative: A difference in the average Y_{0i} of compliers and never-takers provides evidence of selection into a larger family size. A difference in the average Y_{1i} of compliers and always-takers provides evidence of selection, treatment effect heterogeneity, or both (Kowalski, 2016). Columns (2) and (3) in Table 8 show the estimated average potential employment with and without treatment,

respectively. The percentage of employed women with only two children is estimated to be substantially larger for never-takers (29.4%) than for compliers (22.0%). This means that a low preference for having more than two children is related to a higher employment level.[36] Furthermore, the average employment levels with more than two children is around 5 percentage points larger for compliers compared to always-takers. Given the set of assumptions, I cannot disentangle to what extent this difference is due to selection or treatment effect heterogeneity.

To sum up, the size of the group of compliers is small, and it appears that they are different from the rest of the sample in terms of their observable characteristics and their potential outcomes. Hence, the effect of fertility on maternal employment is potentially different for women who do not comply with the samesex-instrument. Still, imposing further assumptions on the potential outcomes as functions of the unobserved distaste for treatment allows me to learn more about the effects of non-compliers.

7.2 External validity imposing further assumptions on marginal outcomes

Next, I consider the external validity of the LATE assuming that maternal employment is either non-increasing or non-decreasing in the distaste for additional children. This weak monotonicity assumption is proposed in Brinch et al. (2017) and it implies that maternal employment with more than two and with two children varies monotonically from always-takers to compliers to never-takers. How covariates change between the three types can provide some indication whether the assumption is plausible (Kowalski, 2016). Table 9 shows that average characteristics are weakly monotonic in the unobserved distaste for children. Assuming weak monotonicity in the marginal maternal employment with and without treatment allows for estimating a bound on the average treatment effect for never- and always-takers, respectively (Kowalski, 2016).

A stronger assumption is that marginal maternal employment with more than two as well as with two children is a linear function of the unobserved distaste for children (Brinch et al., 2017). This means that a one percentage point change in the unobserved distaste for children leads to an average change in maternal employment that is the same across all values of the unobserved distaste for children. Linearity gives point identification of the MTO and MUO, and hence of the MTE function and allows to formally test for unobserved treatment effect heterogeneity. If the slope in the linear MTE model is non-zero so that the MTEs are non-constant, I reject the external validity of the LATE. The test can simply be implemented by regressing Y on D, Z, and the interaction between them and by subsequently performing a two-sided t-test on the interaction coefficient (Brinch et al., 2017). The test results can be found in the Appendix Table D10.

Figures 3 and 4 illustrate the estimated marginal outcomes (maternal employment and maternal informal employment, respectively) and the $\widehat{MTE}$ under the assumption of weak mono-

---

[36]The comparisons of formal and informal employment, and the working hours of never-takers and untreated compliers in Appendix Table D9 go into the same direction.

tonicity and linearity. The vertical axis measures the outcome, whereas the horizontal axis measures the unobserved distaste for having more than two children. Tracing the potential outcomes and the effect of treatment over the unit interval shows how they vary with the unobserved component, where higher values of U_i correspond to lower propensities to have additional children. The black lines in both graphs illustrate estimates of moments that are point-identified under the respective assumption, while gray lines are the bounds. The estimated marginal outcome with and without treatment (or their bounds) are indicated with a solid and dotted line, respectively. The marginal treatment effect is drawn with a dashed line and is equal to the difference between the former two lines.

Figure 3: Estimated marginal outcomes and MTE for employment
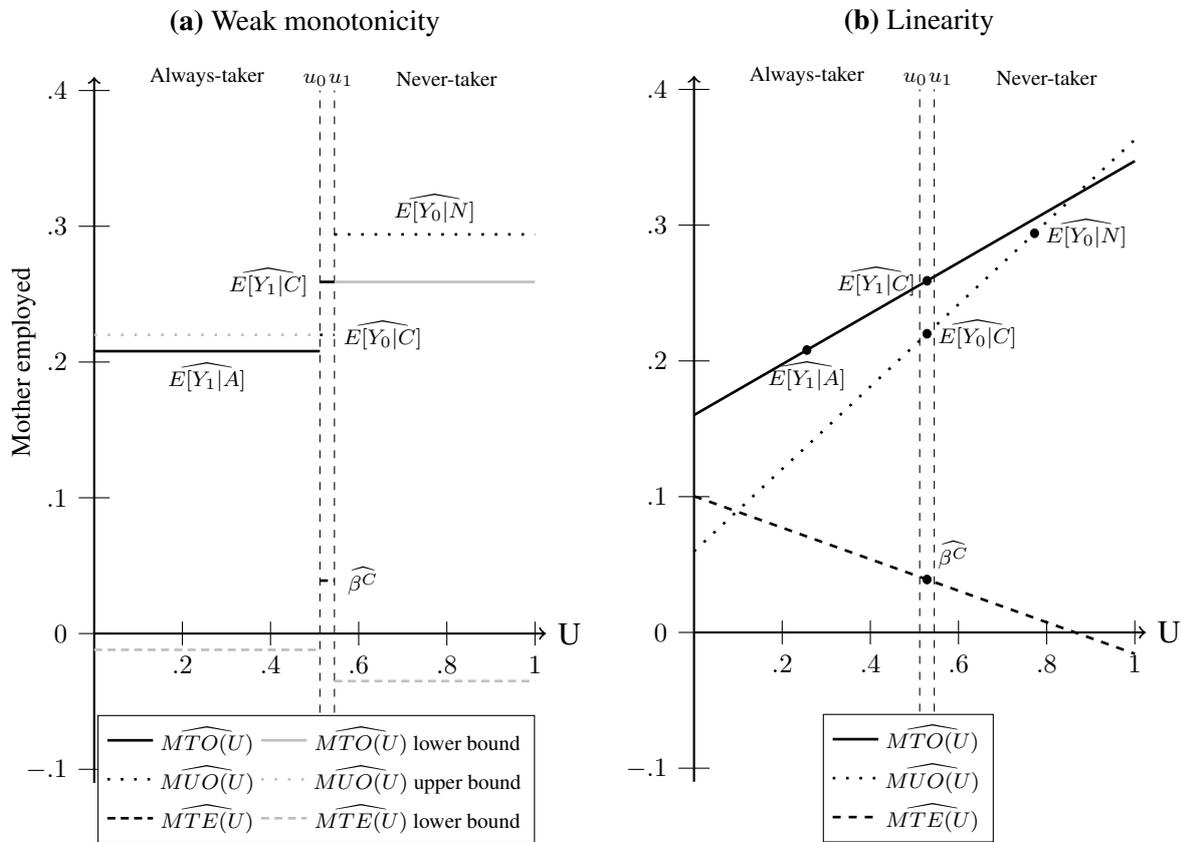
Notes: This figure illustrates point estimates and estimated bounds for the (marginal) treated outcomes ($\widehat{MTO}$), untreated outcomes ($\widehat{MUO}$), and treatment effects ($\widehat{MTE}$) as functions of the unobserved resistance toward having more than two children, U. Panel (a) assumes weak monotonicity, whereas (b) assumes linearity in the marginal treated and untreated outcome. The outcome variable is an indicator equal to 1 if a mother is employed.

Figures 3 focuses on maternal employment in both formal and informal jobs. In Panel (a), I impose the weak monotonicity assumption on the potential untreated and treated outcome. Given that the estimated average outcome without treatment is increasing from compliers to never-takers, i.e., $\widehat{E[Y_0|N]} > \widehat{E[Y_0|C]}$, the weak monotonicity assumption implies that the $\widehat{MUO}$ is weakly upward-sloping in U_i. This suggests that the average outcome without treatment for compliers is an upper bound for always-takers. Similarly, the average outcome with

treatment is increasing from always-takers to compliers implying that the $\widehat{MTO}$ is weakly upward-sloping as well. This means that the mean outcome with treatment for compliers is a lower bound for never-takers. From these two bounds, I can infer a lower bound of the effect of having more than two children on maternal employment of -0.01 for always-takers and of -0.03 for never-takers. So treatment effects are bounded to be close or slightly below zero also for non-compliers given the weak monotonicity assumption suggesting that there is no strong negative effect. Panel (b) shows that under linearity, the estimated marginal treated outcome, $\widehat{MTO}$, is larger than the estimated marginal untreated outcome, $\widehat{MUO}$, up to a value of U_i close to 0.8. This means that an increase in family size from two to more children increases employment for all mothers with an unobserved distaste for treatment below the 80th percentile of the distribution of U. The falling $\widehat{MTE}$ implies that low-resistance households have a higher treatment effect compared to those with a higher resistance toward having more than two children. Testing for the external validity of LATE under linearity, I cannot reject the null hypothesis of a constant MTE (p-value 0.427).

Figure 4: Estimated marginal outcomes and MTE for informal employment
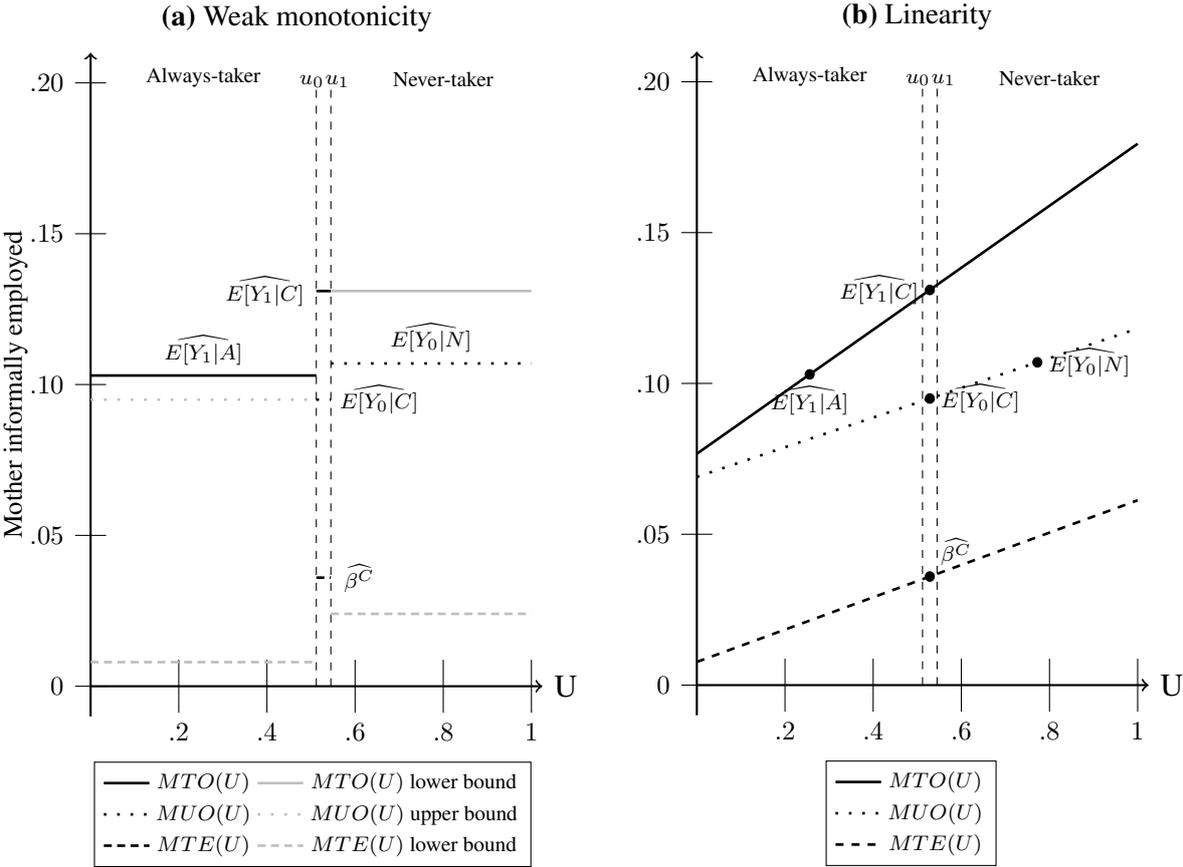
(a) Weak monotonicity   **(b)** Linearity

Notes: This figure illustrates point estimates and estimated bounds for the (marginal) treated outcomes (MTO), untreated outcomes (MUO), and treatment effects (MTE) as functions of the unobserved resistance toward having more than two children, U. Panel (a) assumes weak monotonicity, whereas (b) assumes linearity in the marginal treated and untreated outcome. The outcome variable is an indicator equal to 1 if a mother is informally employed.

In parallel, Figure 4 illustrates the estimated marginal outcomes and the treatment effect on

informal maternal employment. The effect of an increase in family size on the probability that a mother is informally employed is bounded to be non-negative for all women in the sample under the weak monotonicity assumption in panel (a). These lower bounds are informative: They rule out negative effects of having more than two children on informal employment for always-takers and never-takers in my sample. Under linearity, I find that the treatment effect is positive and increasing in the unobserved distaste for having more than two children, as illustrated in panel (b). The latter implies that mothers with a higher distaste for additional children are on average more likely to increase informal employment compared to those with a lower distaste when their family size increases. However, I cannot reject the null hypothesis of constant MTE (p-value 0.614).[37]

To sum up, I provide evidence that the small group of compliers in my setting is different from the group of always- and never-takers in terms of observable characteristics and potential outcomes. This challenges the one-to-one transferability of the effect for compliers to non-compliers. However, in further analysis, I can bound the effect of having more than two children on the propensity to be informally employed to be non-negative for all mothers in the sample under further assumptions.

8   Conclusions

Declining fertility is often assumed to be an important shifter for female labor supply that is likely to account for part of the increase in women's labor force participation over the last decades. In contrast, the results of this study suggest that increased childbearing beyond the second child does not decrease female labor supply in a setting with low household wealth and high availability of informal child care and employment opportunities. Using Mexican census data from 2010 and exploiting exogenous variation in fertility that derives from parents' preferences for mixed-sex siblings, I find that mothers significantly increase their employment in the informal sector when their number of children increases beyond two. Further analyses show that an increase in the fertility beyond two children increases the propensity of living together with the maternal grandparents. Moreover, there is some indication that women in households with low wealth are the ones who especially increase their labor supply.

There might be ample scope for policies, if policy-makers are interested in increasing the welfare of working mothers. It seems to be the case that women substantially decrease their leisure time to provide for an additional child by taking up employment in the informal sector. Informal jobs often provide low pay and inadequate social protection but may offer the only possibility to work due to inflexible working hours in the formal market. Public and corporate policies should encourage more flexible working contracts in the formal labor market, for ex-

---

[37]Similarly, Appendix Figure D2 focuses on maternal formal employment. Under the linearity assumption, the $\widehat{MTE}$ of an increase in fertility beyond two children on formal employment is decreasing in U_i. The effect is positive for always-takers, close to zero for compliers, and negative for never-takers.

ample by extending parental leave and by providing access to workplace support and part-time work to young mothers and fathers. Additionally, more flexible preschool and school schedules can help to combine employment in the formal sector and family responsibilities.

Talking about the interpretation and implications of the findings of this study, one has to keep in mind that the estimated effects are specific to the parity of going from two to more than two children, which does not necessarily generalize. The compliant subpopulation of women for whom the effects are estimated is, moreover, very small and is likely not representative for the rest of the population. Still, under further assumptions, I can reject that there is a negative effect of an increase in fertility on overall employment for most women and on informal employment for all women in the sample.

References

Aaronson, D., Dehejia, R., Jordan, A., Pop-Eleches, C., Samii, C., and Schulze, K. (2017). The Effect of Fertility on Mothers' Labor Supply over the Last Two Centuries. NBER Working Paper No. 23717.

Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics*, 113(2):231–263.

Agüero, J. M. and Marks, M. S. (2008). Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks. *American Economic Review Papers & Proceedings*, 98(2):500–504.

Agüero, J. M. and Marks, M. S. (2011). Motherhood and Female Labor Supply in the Developing World. Evidence from Infertility Shocks. *Journal of Human Resources*, 46(4):800–826.

Almond, D. and Mazumder, B. (2011). Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy. *American Economic Journal: Applied Economics*, 3(4):56–85.

Ángeles, G., Gadsden, P., Galiani, S., Gertler, P., Herrera, A., Kariger, P., and Seira, E. (2014). The Impact of Daycare on Maternal Labor Supply and Child Development in Mexico. International Initiative for Impact Evaluation Report No. 6.

Angrist, J. D. (2001). Estimation of Limited Dependent Variable models with Dummy Endogenous Regressors. *Journal of Business & Economic Statistics*, 19(1):2–28.

Angrist, J. D. and Evans, W. N. (1996). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. NBER Working Paper No. 5778.

Angrist, J. D. and Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, 88(3):450–477.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434):444–455.

Attanasio, O. and Lechene, V. (2002). Tests of Income Pooling in Household Decisions. *Review of Economic Dynamics*, 5(4):720 – 748.

Bisbee, J., Dehejia, R., Pop-Eleches, C., and Samii, C. (2017). Local Instruments, Global Extrapolation: External Validity of the Labor Supply - Fertility Local Average Treatment Effect. *Journal of Labor Economics*, 35(1):99–147.

Bosch, M. and Campos-Vazquez, R. M. (2014). The Trade-Offs of Welfare Policies in Labor Markets with Informal Jobs: The Case of the "Seguro Popular" Program in Mexico. *American Economic Journal: Economic Policy*, 6(4):71–99.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond LATE with a Discrete Instrument. *Journal of Political Economy*, 125(4):985–1039.

Bronars, S. G. and Grogger, J. (1994). The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment. *American Economic Review*, 84(5):1141–1156.

Bütikofer, A. (2011). Sibling Sex Composition and Cost of Children. Unpublished manuscript, nhh bergen.

Cáceres-Delpiano, J. (2012). Can We Still Learn Something From the Relationship Between Fertility and Mother's Employment? Evidence From Developing Countries. *Demography*, 49(1):151–174.

Chen, L.-Y. and Szroeter, J. (2014). Testing Multiple Inequality Hypotheses: A Smoothed Indicator Approach. *Journal of Econometrics*, 178:678–693.

Compton, J. and Pollak, R. (2014). Family Proximity, Childcare, and Women's Labor Force Attachment. *Journal of Urban Economics*, 79:72–90.

Conti, G. and Ginja, R. (2017). Who Benefits from Free Health Insurance: Evidence from Mexico. IFS Working Paper No. 26.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions. *Labour Economics*, 41:47–60.

Cristia, J. P. (2008). The Effect of a First Child on Female Labor Supply: Evidence from Women Seeking Fertility Services. *Journal of Human Resources*, 43(3):487–510.

Cruces, G. and Galiani, S. (2007). Fertility and Female Labor Supply in Latin America: New Causal Evidence. *Labour Economics*, 14(3):565 – 573.

Dahl, G. B. and Moretti, E. (2008). The Demand for Sons. *Review of Economic Studies*, 75(4):1085–1120.

De Chaisemartin, C. (2017). Tolerating Defiance? Local Average Treatment Effects without Monotonicity. *Quantitative Economics*, 8(2):367–396.

Frölich, M. (2007). Nonparametric IV Estimation of Local Average Treatment Effects with Covariates. *Journal of Econometrics*, 139(1):35–75.

Frölich, M. and Melly, B. (2013). Unconditional Quantile Treatment Effects Under Endogeneity. *Journal of Business & Economic Statistics*, 31(3):346–357.

Gong, X. and van Soest, A. (2002). Family Structure and Female Labor Supply in Mexico City. *Journal of Human Resources*, 37(1):163–191.

Heath, R. (2017). Fertility at Work: Children and Women's Labor Market Outcomes in Urban Ghana. *Journal of Development Economics*, 126:190–214.

Heckman, J. and Vytlacil, E. (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences*, 96(8):4730–4734.

Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738.

Hesketh, T. and Xing, Z. W. (2006). Abnormal Sex ratios in Human Populations: Causes and Consequences. *Proceedings of the National Academy of Sciences of the United States of America*, 103(36):13271–13275.

Huber, M. (2015). Testing the Validity of the Sibling Sex Ratio Instrument. *LABOUR*, 29(1):1–14.

Huber, M. and Mellace, G. (2015). Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints. *Review of Economics and Statistics*, 97(2):398–411.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Imbens, G. W. and Rubin, D. B. (1997). Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *Review of Economic Studies*, 64(4):555–74.

INEGI (2014). La Informalidad Laboral. Encuesta Nacional de Ocupación y Empleo. Marco conceptual y Metodológico. National Institute of Statistics and Geography (INEGI), Aguascalientes, Mexico.

International Labor Organization (2013). *Measuring Informality: A Statistical Manual on the Informal Sector and Informal Employment*. International Labor Office, Geneva, Switzerland.

International Labour Organization (2009). 2008 Labour Overview - Latin America and the Caribbean. Technical report.

Jacobsen, J. P., Pearce, J. W., and Rosenbloom, J. L. (1999). The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment. *Journal of Human Resources*, 34(3):449–474.

Kleven, H., Landais, C., and Søgaard, J. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Kowalski, A. E. (2016). Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments. NBER Working Paper No. 22363.

Lee, J. (2008). Sibling Size and Investment in Children's Education: An Asian Instrument. *Journal of Population Economics*, 21(4):855–875.

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

Machado, C., Shaikh, A. M., and Vytlacil, E. J. (2019). Instrumental Variables and the Sign of the Average Treatment Effect. *Journal of Econometrics*, 212(2):522–555.

Melly, B. and Wüthrich, K. (2017). Local Quantile Treatment Effects. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*. CRC Press, London.

Minnesota Population Center (2014). Integrated Public Use Microdata Series - International. Original data by National Institute of Statistics and Geography Census Mexico, microdata sample, 2010.

Mourifié, I. and Wan, Y. (2017). Testing Local Average Treatment Effect Assumptions. *Review of Economics and Statistics*, 99(2):305–313.

OECD (2017). *Building an Inclusive Mexico: Policies and Good Governance for Gender Equality*. OECD Publishing, Paris.

Parker, S. W. and Todd, P. E. (2017). Conditional Cash Transfers: The Case of Progresa/Oportunidades. *Journal of Economic Literature*, 55(3):866–915.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., and Saavedra, J. (2007). *Informality: Exit and Exclusion*. World Bank Latin America and Caribbean Studies, World Bank, Washington D.C.

Rodin, D. L., McNeill, K., Vite-Leon, N., and Heymann, J. (2012). Determinants of Informal Employment among Working Mothers in Mexico. *Community, Work & Family*, 15(1):85–99.

Rosenzweig, M. R. and Wolpin, K. I. (1980). Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models. *Journal of Political Economy*, 88(2):328–48.

Ruiz, A. M. and Campos-Vazquez, R. (2013). ¿Niña o Niño? Un Estudio sobre las Preferencias de los Padres Mexicanos por el Sexo de sus Hijos. *Estudios Económicos*, 28(2):217–248.

Schultz, T. P. (2008). Population Policies, Fertility, Women's Human Capital, and Child Quality. In Schultz, T. P. and Strauss, J. A., editors, *Handbook of Development Economics*, pages 3249–3303. Elsevier.

Staab, S. and Gerhard, R. (2010). Childcare Service Expansion in Chile and Mexico: For Women or Children or Both? UNRISD Gender and Development Programme Paper No. 10.

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, 70(1):331–341.

Wolpin, K. I. and Rosenzweig, M. R. (2000). Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, 38(4):827–874.

World Bank (2012). The Effect of Women's Economic Power in Latin America and the Caribbean. Latin America and Caribbean Poverty and Labor Brief No. 74103.

Appendix
Fertility as a Driver of Maternal Employment

Julia Schmieder

July 2020

This Appendix provides additional material discussed in 'Fertility as a Driver of Maternal Employment' by Julia Schmieder. Section A provides an overview of the literature on the relationship between fertility and maternal employment. Section B compares and discusses related results on Mexico. Section C describes the definition of informal employment used in the paper and compares it to an alternative definition. Section D analyzes fertility preferences. Section E contains additional figures and tables of the empirical analysis.

A   Relationship to the literature

Table A1: Selected studies on fertility and maternal employment (extensive margin): High-income countries

	Country	Data	Sample	Effect of interest and empirical strategy	Mean	Coefficient (SE)	Subgroup
Rosenzweig and Wolpin (1980)[a]	US	survey 1965–73	women aged 15–44 with at least one child	reduced form effect of twin first births on current employment (effect on number of children is 1.055, 0.654, 0.631, 0.150, and 0.312, respectively)	0.290	-0.371 (0.212)	age 15-24, first birth under 25
					0.362	-0.102 (0.105)	age 25-34, first birth under 25
					0.362	-0.096 (0.096)	age 25-34, first birth under 35
					0.448	0.142 (0.102)	age 35-44, first birth under 25
					0.448	0.060 (0.078)	age 35-44, first birth under 35
Bronars and Grogger (1994)[a]	US	Census 1970, 1980	women with all children born in household, oldest child age 18 or younger	reduced form effect of twin first births on labor force participation (effect on number of children is 0.808, 0.750, 0.638, and 0.676, respectively)	0.430	-0.036 (0.036)	unmarried at first birth, 1970
					0.557	-0.035 (0.017)	unmarried at first birth, 1980
					-	-0.023 (0.011)	married at first birth, 1970
					-	-0.010 (0.008)	married at first birth, 1980
Angrist and Evans (1998)	US	Census 1980, 1990	women aged 21–35 with at least two children, oldest child at most 18 years old	effect of having more than two children on work for pay in year before census using sex composition of two firstborn children as instrument	0.565	-0.120 (0.025)	all women, 1980
					0.662	-0.092 (0.024)	all women, 1990
Jacobsen et al. (1999)	US	Census 1970, 1980	women with all children born in household, oldest child age 18 or younger, married at first birth	effect of number of children on work for pay in year before census using twin first births as instrument	0.48	-0.021 (0.0135)	1970
					0.62	-0.025 (0.008)	1980
Cristia (2008)	US	survey 1995	women aged 19-38 who sought help to achieve pregnancy	effect of having a first child younger than one year on employment using childless women as controls	0.877	-0.277 (0.046)	
Lundborg et al. (2017)	DK	admin 1994-2005	women without children who obtain first IVF treatment	effect of having children on indicator of having positive earnings using success at first IVF as instrument	0.90	-0.072 (0.006)	0-1 years after birth
					0.90	-0.041 (0.012)	2-5 years after birth
					0.90	-0.015 (0.022)	6-10 years after birth
Kleven et al. (2019)	DK	admin 1985–2003	women with first child, observed in each year between 5 years before and 10 years after childbirth	event study of labor force participation around birth of first child exploiting variation in event time conditional on time and age	-	-0.110 (-)[b]	0-1 years after birth
					-	-0.120 (-)[b]	2-5 years after birth
					-	-0.140 (-)[b]	6-10 years after birth

Notes: Cells with - indicate that estimate not reported in study. Aaronson et al. (2017) and Bisbee et al. (2017) study fertility and maternal employment across high-, middle-, and low-income countries. I include their results on Mexico in Appendix Table A3.

[a] Studies report reduced form estimates of twin first births on number of children and employment, but no IV estimates

[b] Coefficients are approximated from figures, since there are no tables available, all coefficients are statistically significantly different from 0 at 95% level.

Table A2: Selected studies on fertility and maternal employment (extensive margin): Low- and middle-income countries

	Country	Data	Sample	Effect of interest and empirical strategy	Mean	Coefficient (SE)	Subgroup
Cruces and Galiani (2007)	AR, MX	Census	women aged 21-35 with at least two children, oldest child aged at most 18 years	effect of having more than two children on work for pay using sex composition of two firstborn children as instrument	0.315	-0.082 (0.032)	Argentina 1991
					0.239	-0.063 (0.037)	Mexico 2000
Agüero and Marks (2008)	several Latin American countries[a]	DHS	non-sterilized, sexual active women who are not using contraceptives, aged 20-44	effect of different fertility measures on work for pay using infertility as instrument	0.52	0.003 (0.018)	treatment: no. of children at home
					0.52	0.007 (0.038)	treatment: no. of children under 6
					0.52	0.017 (0.090)	treatment: at least one child
Agüero and Marks (2011)	26 low- and middle-income countries	DHS	non-sterilized, sexual active women who are not using contraceptives, aged 20-44	effect of number of children at home on employment measures using infertility as instrument	0.64	0.006 (0.008)	outcome: worked in the last 12 months
					0.49	-0.009 (0.008)	outcome: worked for a wage
					-	-0.019 (0.014)	middle-income countries, outcome: worked in the last 12 month
					-	0.000 (0.015)	middle-income countries, outcome: worked for a wage
					-	-0.010 (0.008)	low-income countries, outcome: worked in the last 12 month
					-	-0.021 (0.008)	low-income countries, outcome: worked for a wage
Cáceres-Delpiano (2012)	40 developing countries	DHS	women aged 18-40 with first birth aged 15-35, with at least one (two) child (children)[b]	effect of number of children on current employment measures using twin first (second) birth as instrument[b]	0.516	-0.030 (0.012)	1+ parity, outcome: working
					0.346	-0.022 (0.019)	1+ parity, outcome: informal job
					0.524	-0.016 (0.015)	2+ parity, outcome: working
					0.358	-0.028 (0.019)	2+ parity, outcome: informal job
Heath (2017)	GH	panel survey 2004-13	women in four cities	effect of number of children on work in last week using individual fixed-effect estimation flexibly controlling for age	-	-0.018 (0.009)	outcome: working
					-	0.020 (0.007)	outcome: self-employment cond. on working

Notes: Cells with - indicate that estimate not reported in study. Aaronson et al. (2017) and Bisbee et al. (2017) study fertility and maternal employment across high-, middle-, and low-income countries. I include their results on Mexico in Appendix Table A3.
[a] PE (1996), GT (1998), CO (1995), BO (1994 and 1998), NI (1998), and DO (1996)
[b] the study also presents results on subsamples of women with more than 3 and 4 children using twin third and fourth birth as instrument, respectively.

B Comparison and discussion of related results on Mexico

There are other studies that are close to mine, also using the same methodology with Mexican census data. Cruces and Galiani (2007) provide results for the year 2000, while Bisbee et al. (2017) and Aaronson et al. (2017) report estimates for several years as part of an analysis that includes a large number of countries across multiple time periods. In Appendix Table A3, I list these different results alongside mine for the year 2010. Additionally, I report the results from applying my empirical procedure to the census data for other years. First, I compare the estimated effects across the studies for a given year. My estimates are close to those reported in the other studies. This indicates that my findings are not specific to the empirical procedure that I apply. The estimates are almost identical to Cruces and Galiani (2007) in 2000 who use the same labor supply measure, namely employment. In contrast, Bisbee et al. (2017) and Aaronson et al. (2017) look at labor force participation, finding slightly different results. My point estimate for 2010 is smaller and more precisely estimated compared to Aaronson et al. (2017), but they have the same sign.[1] Second, comparing the estimated effects over time there is a significant negative effect of having more than two children on maternal labor supply in 1990 and 2000. The effect turns positive in 2010 and is close to zero in 2015. There are different potential explanations why the effect changes over time. Around 2010, labor market outcomes of Mexican men were hit substantially by the Great Recession, with median household income (excluding maternal income) decreasing.[2] At that time, the labor market environment was more favorable for women and, as a result, women were increasingly responsible for the provision of financial resources to the family (World Bank, 2012). Furthermore, changes in the institutional setting might have impacted the relationship between children and employment. For example, there were two large reforms in the early 2000s that reduced the price of childcare substantially, especially for informally employed mothers. These changes can potentially explain why the effect is negative before 2010 and non-negative thereafter. A final explanation relates to the heterogeneity of treatment effects: the estimated effects are specific to women who are induced to change their fertility because of their children's sex composition. The share of women who is treated, i.e., those who have more than two children, changes from 63% in 1990 to 48% in 2015. Hence, women who comply with the instrument over the years might have very different unobserved costs of having more children that are related to different marginal treatment effects.

---

[1]The difference might be related to a different empirical model: Aaronson et al. (2017) use a linear 2SLS specification controlling for mother's age, her age at first birth, and an indicator for the gender of the first child. Moreover, they weight their estimated effects with household weights, which I do not.

[2]Appendix Table A4 shows that household income (excluding maternal income) dropped by around 10% between 2000 and 2010 and increased again thereafter. This change comes mainly from a decreasing share of spouses with a positive income and a decreasing spousal income conditional on employment.

Table A3: Findings for Mexico: Effect of having more than two children on maternal labor supply using the samesex instrument

year	Aaronson et al. (2017)[a]			Bisbee et al. (2017)[b]		Cruces and Galiani (2007)[c]			My results[d]		
	more than two children	outcome: in labor force	coeff (se)	outcome: in labor force	coeff (se)	more than two children	outcome: work	coeff (se)	more than two children	outcome: work	coeff (se)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1990	0.605	0.159	-0.059 (0.038)	0.27	-0.079 (0.039)				0.630	0.152	-0.079 (0.036)
2000	0.499	0.285	-0.083 (0.048)	0.35	-0.031 (0.039)	0.592	0.239	-0.063 (0.037)	0.558	0.249	-0.065 (0.037)
2010	0.434	0.337	0.082 (0.089)						0.529	0.248	0.048 (0.031)
2015	0.407	0.328	-0.030 (0.069)						0.481	0.234	-0.016 (0.037)

Notes: The four studies use sample restrictions as in Angrist and Evans (1998). I do not report the results for 1970 and 1995, because the fraction of individuals sampled is much smaller compared to the other years (0.4% and 1% compared to 10%) and the coefficients are very imprecisely estimated.

[a] Results can be found in Appendix Table A1 in their paper, outcome variable is dummy whether economically active (employed or unemployed).

[b] Results can be found in Appendix Table A1 in their paper, outcome variable is dummy whether economically active (employed or unemployed).

[c] Results can be found in Table 2 in their paper, outcome variable is dummy whether worked for pay.

[d] Results are based on the main specification in Table 4 in column (5). The results for the year 2010 are from the main paper. For the other years, I applied the same data cleaning and construction, sample restrictions and estimations as in the main specification in the main paper. The data was obtained via the IPUMS International database.

Table A4: Household income (excluding maternal income) over time

Year	median non-mother income	Spouse			
		present	unemployed	positive earnings	median income
	(1)	(2)	(3)	(4)	(5)
1990	4571.50	0.913	0.016	0.861	4586.73
2000	3394.88	0.883	0.009	0.798	3638.18
2010	3070.05	0.861	0.035	0.700	3582.32
2015	3429.00	0.846	0.034	0.733	3643.00

Notes: Income measures are in real Mexican Pesos, 2015 prices. Notice that in 1993 there was a currency change, which I take into account by translating 1000 "old" Mexican Pesos (MXP) into 1 "new" Peso (MXN). Non-mother income in column (1) includes all household income minus maternal income and the median refers to all households. Spousal labor market outcomes in columns (3) to (5) are reported for all households in which the spouse is present.

C   Informal employment definition

In this section, I describe my definition of informal employment, explain the ILO definition (International Labor Organization, 2013) and the National Institute of Statistics and Geography in Mexico definition (INEGI, 2014), and explain how informal employment in my sample compares to informal employment according to the definition by the INEGI using data from the National Survey of Occupation and Employment (ENOE) 2010, first trimester from Mexico.[3]

My definition of informal employment comprises own account workers (self-employment with no employees hired), day laborers, and unpaid family workers who work outside home, whereas formal employment includes business owners with employees as well as blue- and white-collar employees who work for a company or a government for payment.

There are international standards on informal employment proposed by the International Labor Organization (ILO). According to the ILO, informal employment is defined as comprising employers and own-account workers employed in their own informal sector enterprises or engaged in the production of goods exclusively for own final use, contributing family workers, members of informal producers' cooperatives, and employees holding informal jobs (jobs without social protection and other job-related benefits) both in the formal and informal sector (International Labor Organization, 2013). The National Institute of Statistics and Geography in Mexico defines informal employment based on these international standards (INEGI, 2014). Unfortunately, I can neither observe if employees are in jobs with social protection and job-related benefits nor whether enterprises are formally registered.

To compare my informal employment definition to the INEGI definition, I restrict the ENOE data to women aged 21–35 with at least 2 children. The data already includes a variable identifying informal and formal workers based on the INEGI definition. In my sample, 42 percent of all employed women have an informal job. In the ENOE sample, this number is 63 percent. I do not worry much about assigning individuals to informal employment, even if they are actually

---

[3]The data is freely available and can be downloaded from `www.inegi.org.mx/programas/enoe/15ymas/`, last access 10/2019.

formal: according to the ENOE sample, 93 percent of own-account workers are informal; day laborers and unpaid family worker outside home are both informal by definition. This implies that I identify around two thirds of all informal workers. I mainly fail to identify informal workers among the salary workers in private or government companies who do not have access to social security.

In Figure B1, I show the ten occupations (out of about 300 occupations) with the largest representation of informal employees for (a) my sample and (b) for the ENOE sample. Seven of these occupations are the same in my sample and in the ENOE sample. Moreover, Figure B2 shows that the working hours distributions of formal and informal workers are very similar for the two definitions.

Figure B1: Occupations with highest shares among informal employees

(a) Sample

(b) ENOE

Notes: (a) 2010 census data from Mexico with restrictions as outlined in the text. The units of observation are women aged 21–35 with at least 2 children. Informal employment is defined as in the text. (b) National Survey of Occupation and Employment 2010 (first trimester) from Mexico; the data can be downloaded from `https://www.inegi.org.mx/programas/enoe/15ymas/`, last access 11/2019. Informal employment is defined by the National Institute of Statistics and Geography in Mexico (see INEGI. La informalidad laboral. Encuesta Nacional de Ocupación y Empleo. Marco conceptual y metodológico. 2014) and is based on the ILO definition. In both graphs, the units of observation are women aged 21–35 with at least 2 children.

Figure B2: Distribution of working hours

(a) Sample   **(b)** ENOE
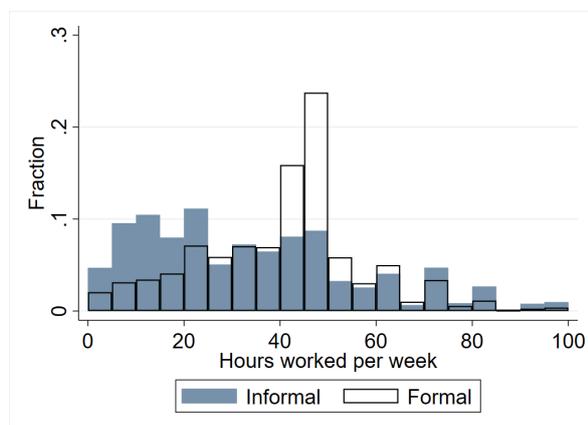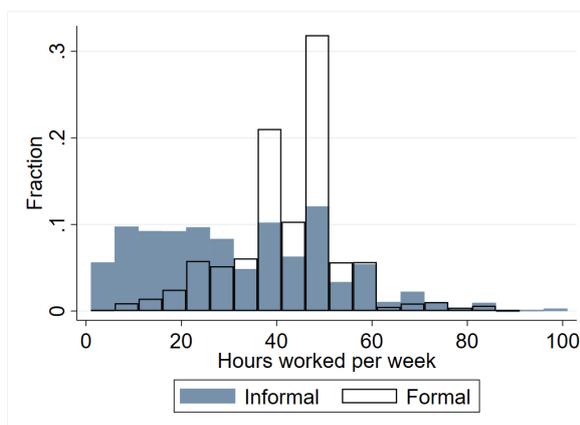
Formal: Mean: 39.7, Median: 40, SD: 17.8   Formal: Mean: 42.8, Median: 45, SD: 11.8
Informal: Mean: 33.7, Median: 30, SD: 23.8   Informal: Mean: 31.9, Median: 30, SD: 19.1

Notes: (a) 2010 census data from Mexico with restrictions as outlined in the text. Informal employment is defined as in the text. (b) National Survey of Occupation and Employment 2010 (first trimester) from Mexico; the data can be downloaded from `https://www.inegi.org.mx/programas/enoe/15ymas/`, last access 11/2019. Informal employment is defined by the National Institute of Statistics and Geography in Mexico (see INEGI. La informalidad laboral. Encuesta Nacional de Ocupación y Empleo. Marco conceptual y metodológico. 2014) and is based on the ILO definition. In both graphs, the units of observation are women aged 21–35 with at least 2 children.

D   Fertility Preferences

This section investigates fertility preferences in Mexico in order to support the validity of the monotonicity assumption regarding the instrument in the empirical setting of this paper. Monotonicity requires that while having mixed-sex siblings, compared to same-sex siblings, may not affect the fertility of some women, all of those affected are assumed to be affected in the same direction. It fails, for instance, if there are mothers who have a preference for at least two children of the same sex and, thereby, choose to have a third child if the first children have different sexes. This discussion is based on three different data sets.

National Survey of Demographic Dynamics 2014

The National Survey of Demographic Dynamics 2014 (ENADID, "Encuesta Nacional de la Dinámica Demográfica")[4] is a national survey of Mexican households. It contains questions on the fertility history of the mother, including a section on her fertility preferences. There is a question on the desired number of children ("If you could go back to the time when you did not have any children and could choose how many children you would have in your life. How many would you have?"), but the survey does not include questions regarding the desired sex of the children. For women for whom the number of desired is smaller than the actual number of children, there is a question "What is the main reason that you have had more children than desired?" The respondents can choose among five predefined answers[5] or, alternatively, they can provide their own answer. The data additionally includes information on the number, age, and sex of the children in the household.

In the group of women aged between 21 and 35 with at least two children (N=17,682), 6,154 (35.9%) women state that their ideal number of children is two. Among them, 1,381 (22.4%) had more than two children. When asked about the reason they had more children than desired, 35.8% respond that they did not use contraceptives, 32.3% say that their contraceptive method failed, and 15.7% respond that their partner wanted to have more children. A total of 37 (2.8%) gave the answer that "they were wishing for a child of the opposite sex than the children they already had." Among women with two firstborn children of the same sex this proportion is 5.2%. This share is slightly above the estimated complier share in my sample (3.3%). Moreover, the data does not indicate that any women stated that they had more children than desired because they wanted to have two children of the same sex.

Mexican Family Life Survey 2002-2012

---

[4]The data and its documentation can be downloaded for free from `http://www.inegi.org.mx/programas/enadid/2014/`, last access 10/2019. It is conducted by the National Institute of Statistics and Geography.

[5]The answers are 1 - they did not use contraceptives, 2 - they did not know contraceptive methods, 3 - their contraceptive method failed, 4 - their partner wanted to have more children and 5 - for religious reasons.

The Mexican Family Life Survey is a longitudinal survey representative of the Mexican population. It contains information collected in three rounds in 2002, 2005-2006 and 2009-2012.[6] The survey asks female household members aged between 14 and 49 years the question, "How many *more* children would you like to have?" Those who want more children are asked "Among the children you still wish to have, how many of each, boys or girls, would you like to have?". Hence, the survey only provides information about the desired sex of children wanted in the future, but not about the ideal sex composition of all children. In order to observe preferences about the desired sex composition of the two firstborn children, I focus on women who desire two children but who have not yet reached this number. The data additionally includes information on the number, age, and sex of the children in the household. Taking into account the sex of their first child, I can infer the desired sex composition of the two firstborn children for women who already have one child.[7]

First, there are 6,075 individual-survey year observations for whom I can infer the desired sex composition of the two firstborn children. Among these women, 79.2% would like to have one boy and one girl, 4.0% two boys, and 4.7% two girls. The remaining are indifferent. Hence, the share of women who either have a preference for two children of different sexes or who are indifferent is one order of magnitude larger than the share of women who desire two boys or two girls only.

Second, I follow the women who had either no or one child and stated they would like to have a total of two children in 2002 (N=1,927). For these women, I can infer the preferences regarding the sex composition of their two firstborn children. In 2005, I observe 1,553 of these women (32 with a preference for two boys, 38 for two girls, and 940 for a mix in 2002). Among these, only 153 had at least two children and 13 at least three in 2005. In 2009, the respective numbers are 901 (25 with a preference for two boys, 31 for two girls, and 740 for a mix). Among them, 227 had at least two children and 59 at least three in 2009. Hence, given the short time period between the surveys, relatively few of the women are at the point of deciding whether they had more than two children in 2005 and 2009. Within the different preference subgroups, in particular for those with preferences for two girls or two boys, I have too few observations to draw any conclusions about whether the realized fertility matches the stated preferences.

Demographic and Health Survey Colombia 2010

As there is no data available on Mexico that includes information on preferences regarding

---

[6]The data is public and can be downloaded without any charge from http://www.ennvih-mxfls.org, last access 10/2019. It is developed and managed by the Universidad Iberoamericana (UIA) and the Centro de Investigación y Docencia Económicas (CIDE) in collaboration with researchers from Duke University.

[7]In contrast to the main analysis, I keep women aged between 14 and 35. I choose the lower age limit to observe more women that do not have children yet (hence, I can observe the ideal sex composition for all their desired children).

the number and the sex of children for all women irrespective of the number they already have, I examine Demographic and Health Survey (DHS) data from 2010 on a sample of mothers from Colombia, which is geographically and culturally close.[8] In this survey, there is a question on "the ideal number of children that the respondent would have liked to have in her whole life, irrespective of the number she already has," as well as questions on the ideal number of boys, the ideal number of girls, and the ideal number of either sex.

I first compare some fertility-related statistics between the Colombian DHS and the Mexican data used in the main analysis in order to illustrate that they are not too different. The first stage coefficient on the samesex instrument in the DHS data for Colombia is with 3.25% very close to the coefficient for Mexico. Mothers are, on average, 29 years old and their age at first birth is, on average, 19.5 years. These numbers are very close to the Mexican ones. In contrast, only 39.2% have more than two children (in Mexico: 53.3%). In the Colombian DHS data, among women whose ideal number of children is two, 81.0% have a preference for a mixed sex composition, 4.4% have a preference for two boys, and 7.6% for two girls. The rest is indifferent. These numbers are close to those in the Mexican Life Survey reported above. They support the argument that a preference for two children of the same sex is rather rare relative to a preference for a mixed sex composition.

How do these stated preferences relate to realized fertility? Among women whose first two children are a boy and a girl, 0.77% have three children or more and retrospectively declare that their ideal sex composition would have been two boys and no girl, or no boy and two girls. In comparison, among women whose first two children are of the same sex, 12.7% have at least three children and retrospectively declare that their ideal sex composition would have been one boy and one girl. Relating these numbers to the share of women who have a preference for two boys and two girls and for a mixed sex composition, respectively, it does not seem to be the case that the realized fertility is more responsive to unmet preferences of the former group.

To sum up, the share of women who have fertility preferences that might induce them to act as defiers with respect to my instrument is very small, especially in comparison to the share of women with preferences that might induce them to act as compliers. Moreover, it does not seem to be the case that women with preferences for two boys and two girls react stronger in terms of their realized fertility if their preferences are not realized.

---

[8]The data was retrieved from https://dhsprogram.com/Data/; last access 05/2018. I use the Standard DHS from Colombia in 2010, Individual Recode, and apply the same sample restrictions as for the main sample.

E   Further results

E.1   Figures

Figure D1: Distributional impact of fertility on mother's hours worked per week
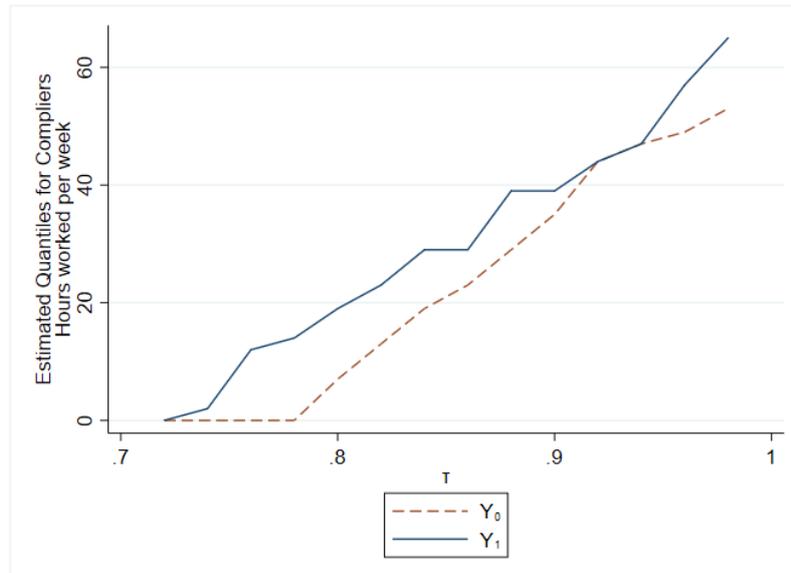
Notes: This graph illustrates the estimated quantiles for the potential outcome distributions of mother's total weekly working hours. The solid blue and the dashed red lines show the estimated quantiles of working hours for compliers with more than two children, $\hat{Q}^\tau_{Y_1|c}$, and with only two children, $\hat{Q}^\tau_{Y_0|c}$, respectively. The model includes the control variables that are listed in Table 3.

Figure D2: Estimated marginal outcomes and MTE for formal employment

(a) Weak monotonicity

(b) Linearity

Notes: This figure illustrates point estimates and estimated bounds for the (marginal) treated outcomes (MTO), untreated outcomes (MUO), and treatment effects (MTE) as functions of the unobserved resistance toward having more than two children, U. The left panel assumes weak monotonicity, whereas the right panel assumes linearity in the marginal treated and untreated outcome.

A14

E.2 Tables

Table D1: Sample compared to a broader group of Mexican women

	Women aged [21–50] yrs.		Women aged [21–35] yrs.		Women in sample	
Age (yrs)	34.11	(8.50)	27.74	(4.33)	29.39	(3.83)
Aged 21-25	0.21	(0.40)	0.36	(0.48)	0.19	(0.39)
Aged 26-30	0.19	(0.39)	0.33	(0.47)	0.38	(0.49)
Aged 31-35	0.17	(0.38)	0.30	(0.46)	0.43	(0.49)
Age at first birth (yrs)	22.34	(5.10)	20.47	(4.02)	19.76	(3.08)
Married or cohabiting	0.72	(0.45)	0.68	(0.47)	0.91	(0.29)
Primary completed	0.25	(0.43)	0.18	(0.39)	0.62	(0.48)
Secondary completed	0.51	(0.50)	0.53	(0.50)	0.12	(0.32)
University completed	0.17	(0.37)	0.20	(0.40)	0.03	(0.17)
Children in household	1.94	(1.64)	1.62	(1.50)	2.84	(1.05)
At least 2 children	0.58	(0.49)	0.49	(0.50)	1.00	(0.00)
More than 2 children	0.32	(0.47)	0.25	(0.44)	0.53	(0.50)
Employed	0.36	(0.48)	0.34	(0.47)	0.25	(0.43)
Labor market outcomes for employed						
Informal employment	0.37	(0.48)	0.32	(0.47)	0.44	(0.50)
Hours worked per week	39.53	(19.69)	40.04	(19.30)	37.19	(20.78)
Monthly income (Pesos)	3778.94	(3434.82)	3655.94	(3174.08)	2958.68	(2507.40)
Observations	2,411,685		1,372,126		505,569	

Notes: Statistics depicted are means with standard deviations in parentheses. All variables are measured at the time of the survey. The labor market outcomes refer to the week before the survey.

Table D2: Sample characteristics

	All (1)		Married | cohabiting (2)		Single (3)	
Demographic characteristics						
Age (yrs)	29.39	(3.83)	29.39	(3.82)	29.44	(3.89)
Age at first birth (yrs)	19.76	(3.08)	19.78	(3.09)	19.57	(3.00)
Literate	0.91	(0.28)	0.91	(0.29)	0.93	(0.26)
Primary completed	0.62	(0.48)	0.62	(0.48)	0.62	(0.48)
Secondary completed	0.12	(0.32)	0.12	(0.32)	0.14	(0.35)
University completed	0.03	(0.17)	0.03	(0.17)	0.03	(0.17)
Household characteristics						
Rural	0.49	(0.50)	0.50	(0.50)	0.41	(0.49)
Number of household members	5.41	(1.94)	5.38	(1.85)	5.72	(2.66)
Maternal parent present	0.10	(0.30)	0.06	(0.23)	0.49	(0.50)
Paternal parent present	0.08	(0.27)	0.08	(0.28)	0.00	(0.03)
Monthly non-mother income	3,913	(5,634)	4,039	(5,650)	2,712	(5,333)
Ownership dwelling	0.80	(0.40)	0.81	(0.39)	0.75	(0.43)
Fertility characteristics						
Children in household	2.84	(1.05)	2.87	(1.07)	2.60	(0.87)
More than 2 children	0.53	(0.50)	0.54	(0.50)	0.42	(0.49)
Same sex	0.50	(0.50)	0.50	(0.50)	0.51	(0.50)
Two boys	0.26	(0.44)	0.26	(0.44)	0.26	(0.44)
Two girls	0.24	(0.43)	0.24	(0.43)	0.25	(0.43)
Age second child (yrs)	6.44	(3.83)	6.43	(3.83)	6.55	(3.87)
Age youngest child (yrs)	3.74	(3.07)	3.65	(3.03)	4.59	(3.35)
Labor market outcomes						
Employed	0.25	(0.43)	0.21	(0.41)	0.60	(0.49)
Labor market outcomes for employed						
Informal employment	0.42	(0.49)	0.46	(0.50)	0.30	(0.46)
Hours worked per week	37.19	(20.78)	35.76	(20.88)	42.00	(19.68)
Monthly (Pesos)	2,959	(2,507)	2,946	(2,622)	3,001	(2,076)
Observations	505,569		457,546		48,023	

Note: 2010 census data from Mexico with restrictions as outlined in the text. The units of observation are women aged 21–35 with at least 2 children. Statistics depicted are means with standard deviations in parentheses. All variables are measured at the time of the survey. The labor market outcomes refer to the week before the survey.

Table D3: IV validity test based on Huber and Mellace (2015)

outcome variable:	maternal employment			maternal employment, informal		
	$st.dist0$	$st.dist1$	p-value	$st.dist0$	$st.dist1$	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
full sample	-0.037	-0.111	0.999	-0.023	-0.184	1.000
first child boy	-0.033	-0.024	1.000	-0.165	-0.0165	1.000
second child boy	-0.085	-0.013	0.999	-0.153	-0.010	1.000
no primary	-0.039	-0.007	0.990	-0.032	-0.011	1.000
primary	-0.043	-0.125	1.000	-0.025	-0.203	1.000
secondary	-0.026	-0.138	1.000	-0.017	-0.297	0.998
university	-0.029	-0.067	1.000	0.022	-0.306	0.521
no primary & rural	-0.029	-0.067	0.992	-0.031	-0.007	0.993
primary & rural	-0.171	-0.139	1.000	-0.025	-0.198	1.000
secondary & rural	0.009	-0.093	0.937	-0.124	-0.186	1.000
university & rural	0.079	0.001	0.470	0.064	-0.247	0.689
no primary, rural & 1st boy	-0.171	0.017	0.484	-0.196	0.012	0.683
no primary, rural & 2nd boy	-0.132	0.015	0.615	-0.154	0.0160	0.578
primary, rural & 1st boy	-0.122	-0.102	1.000	-0.017	-0.145	1.000
primary, rural & 2nd boy	-0.088	-0.005	0.979	-0.136	0.002	0.858
secondary, rural & 1st boy	-0.002	-0.055	0.992	-0.109	-0.159	1.000
secondary, rural & 2nd boy	-0.060	-0.081	1.000	-0.090	-0.181	1.000
university, rural & 1st boy	0.120	0.126	0.199	-0.011	-0.235	0.996
university, rural & 2nd boy	0.055	0.0158	0.808	0.095	-0.237	0.714

Note: This table shows the results from the Huber and Mellace (2015) test based on mean constraints for various sample definitions, i.e. for the full sample as well as for subsamples defined upon the values of the covariates. $st.dist0$ ($st.dist1$) is the standardized maximum distance between the point estimate of the never-takers' (always-takers') mean potential outcome and the corresponding bounds. A positive value implies that the point estimate falls outside the bounds. The p-value is based on the procedure of Chen and Szroeter (2014). Tests are based on 999 bootstrap draws. The code is provided by Martin Huber.

Table D4: 2SLS results of the effect of having more than two children on maternal employment with varying set of control variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:								
Employed	0.039	0.043	0.060	0.057	0.054	0.048	0.048	0.046
	(0.037)	(0.037)	(0.037)	(0.037)	(0.036)	(0.035)	(0.035)	(0.035)
Formally employed	0.003	0.004	0.016	0.011	0.009	0.005	0.006	0.005
	(0.030)	(0.029)	(0.030)	(0.030)	(0.028)	(0.028)	(0.028)	(0.028)
Informally employed	0.036	0.038	0.044*	0.045*	0.044*	0.043*	0.042*	0.042
	(0.026)	(0.026)	(0.027)	(0.026)	(0.026)	(0.026)	(0.025)	(0.026)
Working hours	1.818	1.975	2.681	2.506	2.389	2.134	2.161	2.068
	(1.629)	(1.618)	(1.660)	(1.647)	(1.603)	(1.570)	(1.555)	(1.564)
Working hours, formal	0.755	0.835	1.308	1.098	1.018	0.808	0.861	0.782
	(1.314)	(1.305)	(1.339)	(1.323)	(1.279)	(1.249)	(1.249)	(1.244)
Working hours, informal	1.063	1.141	1.374	1.408	1.371	1.325	1.299	1.286
	(1.096)	(1.088)	(1.119)	(1.114)	(1.108)	(1.095)	(1.052)	(1.091)
Set of controls								
$boy1$, $boy2$		✓	✓	✓	✓	✓		
age			✓	✓	✓	✓		
indigeneity				✓	✓	✓		
education					✓	✓		✓
location						✓		✓
"saturated" model							✓	✓
Observations				505,569				

Note: This table shows the effect of having more than two children on maternal employment based on 2SLS regression results using $samesex$ as instrument with a varying set of additional control variables. $boy1$, $boy2$: indicators for the sex of the first- and second-born child, age: dummies for the age of the mother in years, dummies for her age at first birth in years, dummies for the age of the second child, indigeneity: indicator for indigeneity, education: dummies for schooling in years, indicator for literacy, location: municipality size in four categories and state fixed effects. The saturated model includes six categories for the age of the mother, six categories for her age at first birth, an indicator for the sex of the first child, and the indigeneity indicator and all potential interactions between them. It additionally includes an indicator for the sex of the second child. Robust standard errors in parentheses.

A18

Table D5: 2SLS results of the effect of having more than two children on maternal employment with different sample definition

	Age 21–40	Age 18–35
	(1)	(2)
Dependent variable:		
Employed	0.043	0.048
	(0.029)	(0.035)
Formally employed	0.012	0.005
	(0.024)	(0.028)
Informally employed	0.031	0.042
	(0.022)	(0.026)
Working hours	1.151	2.071
	(1.318)	(1.588)
Working hours, formal	0.677	0.874
	(1.058)	(1.264)
Working hours, informal	0.474	1.197
	(0.930)	(1.108)
Observations	652,256	516,571

Note: This table shows the effect of having more than two children on maternal employment based on 2SLS regression results using *samesex* as instrument and including the control variables listed in Table 3 with a varying sample definition. In column (1), I restrict the sample to mothers aged 21–40 and in column (2) to mothers aged 18–35. In the main sample, mothers are aged 21–35. Robust standard errors in parentheses.

Table D6: Relationship between fertility and maternal working hours

	Mean	OLS		2SLS		2SLS⁺	NP-IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:							
Working hours	9.228	-3.492	-3.391	1.977	2.134	2.674	1.922
		(0.054)	(0.064)	(1.618)	(1.570)	(1.547)	(1.756)
Working hours, formal	5.519	-3.252	-2.542	0.691	0.650	0.813	0.503
		(0.043)	(0.051)	(1.283)	(1.230)	(1.211)	(1.073)
Working hours, informal	3.708	-0.240	-0.849	1.286	1.483	1.861	1.420
		(0.037)	(0.046)	(1.116)	(1.123)	(1.105)	(1.314)
Control variables			✓		✓	✓	✓
Observations				505,569			

Note: This table illustrates the relationship between maternal working hours and the treatment similar to Table 4. *Working hours* measure the number of hours worked in the week prior to the survey and are set to 0 for those not employed. *Working hours formal (informal)* are the working hours in a formal (informal) job. Control variables are as in Table 3. Robust standard errors in parentheses.

Table D7: Local quantile treatment effects of maternal working hours

τ	.02	...	0.72	0.74	0.76	0.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98
Hours worked per week																
$\hat{\triangle}_{IV}^{\tau}$	0	...	0	1	12	14***	12*	10*	10	6	10	4	0	0	8	12
	(0.30)	...	(0.13)	(15.2)	(37.8)	(1.26)	(6.40)	(5.72)	(5.26)	(5.84)	(13.12)	(4.91)	(6.80)	(2.03)	(9.93)	(6.82)
$\hat{Q}_{Y_0\|c}^{\tau}$	0	... 0	0	0	0	0	7	13	19	23	29	35	44	47	49	53
Hours worked per week, formal																
$\hat{\triangle}_{IV}^{\tau}$	0	...	0	0	0	0	0	0	0	0	6	10	15***	0	0	-2
	(0.37)	...	(0.12)	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(74.44)	(7.81)	(3.91)	(11.85)	(3.20)	(2.52)
$\hat{Q}_{Y_0\|c}^{\tau}$	0	...	0	0	0		0	0	0	0	3	19	24	39	47	49
Hours worked per week, informal																
τ	0	...	0	0	0	0	0	0	0	0	13***	15***	14**	13	3	12
	(0.34)	...	(0.11)	(0.11)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(3.12)	(3.32)	(6.03)	(15.04)	(12.66)	(7.15)
$\hat{\triangle}_{IV}^{\tau}$	0	...	0	0	0	0	0	0	0	0	0	0	9	14	35	47

Notes: The table reports the results of the local quantile treatment effects estimation as proposed in Frölich and Melly (2013). $\hat{\triangle}_{IV}^{\tau}$ gives the quantile treatment effect estimate for compliers at quantile τ with estimated asymptotic standard errors reported in parenthesis below the corresponding estimates. $\hat{Q}_{Y_0|c}^{\tau}$ denotes the estimated τ quantile of $Y_0|c$. In the columns (quantiles 0.04 to 0.70) that are left out, $\hat{\triangle}_{IV}^{\tau}$ and $\hat{Q}_{Y_0|c}^{\tau}$ are (as in the neighboring quantiles) equal to 0 (see also Appendix Figure D1 and Figure 2). Control variables are as in Table 3.

Table D8: Effect of fertility on maternal employment on household wealth indicators

	More than 3 rooms	Above median	
		mixed wealth measure	husband earnings
	(1)	(2)	(3)
	-0.017	0.020	0.006
	(0.038)	(0.034)	(0.037)
Observations	501,955	505,568	415,645

Notes: This table illustrates the relationship between the wealth indicators from Table 7 and the treatment. The three columns show the estimated coefficients on the treatment in a 2SLS regression using $samesex$ as instrument and including the control variables listed in Table 3. The mixed wealth measure is generated through a principal component analysis based on different dwelling characteristics. I include all observations for which the respective measures are observed. Standard errors in parentheses.

Table D9: Outcomes by compliance type

D_i	Formal Empl. 1	Formal Empl. 0	Informal Empl. 1	Informal Empl. 0	Weekly Hours 1	Weekly Hours 0
Type						
Always-taker	0.105	.	0.103	.	7.52	.
	(0.001)		(0.001)		(0.05)	.
Complier	0.128	0.125	0.131	0.095	9.36	7.54
	(0.023)	(0.021)	(0.024)	(0.018)	(1.09)	(1.13)
Never-taker	.	0.187	.	0.107	.	11.20
		(0.001)		(0.001)		(0.06)

Note: This table shows the estimated average potential outcomes of the compliance types with and without treatment. The outcome variable in the left (central) panel is an indicator equal to 1 if a mother is formally (informally) employed. The outcome variable in the right panel is equal to the maternal working hours per week.

Table D10: Test of the null hypothesis of constant MTE under linearity

Outcome variable	Coefficient	p-value
Mother employed	-0.002	0.427
Mother employed, formal	-0.003	0.154
Formally employed, informal	0.001	0.614

Note: The test is implemented by regressing Y on D, Z, and the interaction between them and by subsequently performing a two-sided t-test on the interaction coefficient (Brinch et al., 2017). The table reports the interaction coefficient and the corresponding p-value from a two-sided t-test.