Fertility as a Driver of Maternal Employment*

Julia Schmieder†

June 28, 2018

Abstract

Based on findings from high-income countries, typically economists hypothesize that having more children unambiguously decreases the time mothers spend in the labor market. Research on lower-income countries, in which low household wealth, informal child care, and informal employment opportunities prevail, is scarce. 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.

Keywords: Fertility, Female Labor Supply, Middle-Income Countries, Informality.

JEL classification: J13, J16, J22, J46.

*I gratefully acknowledge financial support from the German Research Foundation within its priority program 1764: The German Labor Market in a Globalized World. For helpful discussions, I thank Maximilian Bach, Sascha Drahs, Peter Haan, Anna Hammerschmid, Adam Lederer, Magne Mogstad, Pedro Sant'Anna, Andrea Weber, Katharina Wrohlich, and participants at the 16th ZEW Summer Workshop and the DIW Cluster Seminar.

†DIW Berlin, WU Vienna, IZA; DIW Berlin, 10108 Berlin, Germany; +49 30 89789 334; 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, stemming mainly from high-income countries, fuels this by almost exclusively finding a negative causal relationship between family size and maternal labor supply (see Lundborg et al. (2017) for a review). 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 (Heath, 2017). Beside the monetary inputs in child quality, there are some fixed costs of children that decrease household income, which make increases in labor supply more valuable by increasing the marginal utility of wealth.

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 than in high-income countries; all playing a potentially interesting role in the relationship between fertility and maternal employment. First, little wealth 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 inelastically 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/or fixed costs of work might be lower in informal compared to formal employment (Gong and van Soest, 2002; Heath, 2017).

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. The empirical analysis uses data on a sample of about 500,000 Mexican mothers aged 21 to 35. The setting provides two attractive features. First, in contrast to other low- and middle-income countries (Schultz, 2008; Wolpin and Rosenzweig, 2000), Mexican parents show a preference for a mixed-sex sibling composition of their children while at the same time there are no discriminatory practices toward one sex that might invalidate the identification strategy. Second, Mexican census data from 2010 allows for observing detailed labor market outcomes in both the informal and formal sector as well as to identify the complex household constellations and characteristics for a large sample of young mothers.

The first set of results confirms the presence of mixed-sex sibling preferences in my sample: Families with two children of the same sex are 3.3pp more likely to have another child compared to those with children of different sexes. A considerable and significantly 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 that an exogenous increase in family size beyond two children leads to a significant 4–6pp increase in women's probability to be employed. This extensive margin response is driven by mothers who start working in the informal sector. I further estimate the effect of an increasing family size at different quantiles of the maternal working hours distribution. My findings suggest that fertility increases female labor supply in the formal sector at the intensive margin for a small fraction of women, while mainly having an impact at the extensive margin in the informal sector.

What are the potential mechanisms behind this positive employment effect? A further analysis on households in which mothers live with their partners shows that they increase their labor supply, while paternal employment and hours do not react at all. This might indicate that men had a harder time increasing their labor supply compared to women in 2010 in order to provide for additional resources for their growing family. This is in line with previous findings that labor market opportunities were relatively better for women in Mexico during the Great

Recession (World Bank, 2012). I further show that having more than two children increases the probability to live with the extended family. In particular, the probability to live with the maternal grandparents rises by approximately 4pp. Moreover, I find suggestive evidence that the positive employment effect is mainly driven by women who live in households with relatively low household wealth. This is in line with the idea that these households have a particularly high value from labor income as their family size increases.

The crucial obstacle for my research design is the internal and external validity of the instrument. I provide credible evidence for the internal validity of my results and I test the identification assumptions formally by a procedure recently proposed by Huber and Mellace (2015). The impact of fertility on female labor supply is likely heterogeneous and hence, the obtained IV estimates are interpreted as local average treatment effects that pertain to the group of females who had more children than they otherwise would have due to the sex composition of their first two children. The proportion of these so-called compliers in my sample is estimated to be around 3–4% 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. Hence, the external validity of this local effect appears to be limited. I subsequently draw on recent approaches proposed by Kowalski (2016) and Brinch et al. (2017) in order to extrapolate to the treatment effect for non-compliers. My findings indicate that, for informal employment, the treatment effect is non-negative for all women in the sample.

My paper is primarily related to the large literature on the relationship between fertility and female labor supply. Browning (1992) discusses methodological issues that arise in modeling the effects of children on household behavior. The emerging empirical literature, predominantly based on quasi-experimental approaches, finds a substantial negative effect of an exogenous increase in family size on maternal labor supply in high-income countries (see e.g. Rosenzweig and Wolpin (1980), Angrist and Evans (1998), Lundborg et al. (2017) and Kleven et al. (2018)). Relatively few published papers focus on middle- and low-income countries, finding a zero to negative effect of family size on maternal employment at the extensive margin (Agüero and

Marks, 2008, 2011; Cáceres-Delpiano, 2012; Cruces and Galiani, 2007).[1] The study by Heath (2017) is closely related to this paper, but uses data on Ghanaian households and a different empirical strategy. She finds a negative effect of fertility at the extensive margin of maternal employment and a positive hours response conditional on employment. In contrast to her, I can identify intensive margin responses with a local quantile treatment effect model that is more reasonable than conditional-on-positive effects, especially in a setting in which a positive hours choice is very selective. Hence, I contribute to the literature by providing empirical evidence on a positive effect of fertility both at the intensive and extensive margin of labor supply in a middle-income country.

I further contribute to the literature by identifying potential channels behind the positive effect of a larger family size on employment. It appears that the effect is mainly driven by mothers moving into the informal sector. The results line up with Rodin et al. (2012) who find in a interview-based study in Mexico that the need for greater job flexibility to care for children and other family member forces some women to take up informal employment. A few other studies look at 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 average, but it reduces the probability of paid employment. Cáceres-Delpiano (2012) finds for a sample of several low-income countries 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. In contrast, Heath (2017) finds a positive effect of having children on self-employment in Ghana. I add to the literature on child care and labor supply by showing a positive relationship between family size and co-residence with maternal grandparents.[2] This might positively affect the availability of child care which makes time less attractive relative to monetary investments and, hence, increases the incentives to increase labor supply. This relates to Gong and van Soest

[1] My results contrast with Cruces and Galiani (2007), who find an inverse relationship between fertility and mother's employment using the same identification strategy and Mexican data from the year 2000. There are different potential explanations for this. Around 2010, the labor market environment was more favorable for women compared to men, as such, women were increasingly responsible for the provision of financial resources to the household. Furthermore, two reforms reduced the price of child care substantially between 2000 and 2010: Preschool education was declared mandatory for all children aged 3–5 years and a federal daycare program for working mothers was launched (Mateo Díaz and Rodríguez Chamussy, 2013; Staab and Gerhard, 2010). Finally, it might be that the complier subpopulation affected by the instrument changed over time.

[2] Posadas and Vidal-Fernandez (2013) provides an excellent literature review on the link between grandparental childcare and maternal labor force participation.

(2002) who show that maternal labor force participation in Mexico is higher when other adults are present in the household. Moreover, my finding that the labor supply response differs by household wealth is interesting in the light of results of Aaronson et al. (2017). They estimate the effect of fertility on maternal labor supply by per-capita income using variation across time and countries and find that the effect is not different from zero at low and larger and negative at higher income levels.

Last but not least, I contribute to the discussion about the validity of the *same sex*-instrument in the context of middle- and low-income countries. The internal validity is repeatedly criticized in the economic literature (see for example Wolpin and Rosenzweig (2000) and Schultz (2008)). In my paper I directly address the concerns that were posed and substantiate my discussion with a formal test procedure. The external validity of the results from IV estimations is rarely discussed in applied work, despite its importance for interpretation and for the derived policy implications. I show how proposed econometric procedures can be applied in order to discuss the representativeness of individuals who are affected by the instrument and to extrapolate from them.

The remainder of the paper proceeds as follows: Section 2 provides background information on households and on several aspects of the labor market in Mexico. This information is helpful to put 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 attributes 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 mechanism behind my findings. The final section offers some concluding remarks.

2 Background

As Mexico experienced substantial labor market changes and strong family dynamics in the sample period, it provides an interesting context for my analysis: Female labor force participation in Mexico was among the lowest in Latin America in the 1990s and 2000s. At the same time, Mexico revealed a strong positive trend in women's participation with an increase from

33.5% in 1990 to 43.2% in 2010, while male participation decreased from 84.4% to 80.1% over the same period. During the Great Recession, female earnings were especially crucial in compensating for the fall in men's labor income (World Bank, 2012). This positive trend in female labor force attachment was accompanied by a decrease in fertility from 3.5 to 2.3 children per woman between 1990 and 2010.[3]

The compatibility of children and maternal employment appears to be low in Mexico. 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.[4] 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 and only a small share of workers are 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). In addition, there were some policies implemented that have the potential to reduce the child care burden: Preschool education was declared mandatory for all children aged 3–5 years and a federal daycare program for working mothers was launched. The program targets mothers with children aged 1–4 from low-income households that are not covered by social security-based childcare services (Mateo Díaz and Rodríguez Chamussy, 2013; Staab and Gerhard, 2010).[5]

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 (Cunningham, 2001; International Labour Organization, 2009). Moreover, there are large gender gaps: women are

[3] Data retrieved from `http://data.worldbank.org/country/mexico`, `http://stats.oecd.org/`; last accessed 02/2018. Total fertility rate refers to number of children who would be born if a woman were to live to the end of her childbearing years and bear children according to age-specific fertility rates of the specified year.

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

[5] By 2011, more than 250,00 children were enrolled and their mothers exhibited very strong positive employment responses to the program (Ángeles et al., 2014).

more likely to work informally and to have informal jobs of lower quality than men (OECD, 2017). In the literature, informal work is both interpreted as a result from segmentation in the labor market that prevents individuals from taking jobs in the formal sector as well as 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.[6] They do not pay social security contributions, which implies that they cannot claim social benefits, such as health or parental leave benefits, or retirement payments.[7] Still, 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 (Cunningham, 2001; Heath, 2017; Perry et al., 2007; Rodin et al., 2012). Other hypotheses are that informal employment provides the possibility to take care of the children while working and that the fixed costs of work are lower in informal compared to formal employment (Heath, 2017). I discuss the relevance of these different hypotheses using my sample in Section 3.

The economic situation of many Mexican households with children is affected by the *Oportunidades* program: 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 (Parker and Todd, 2017). 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. The monetary transfer for school children increases in the grade and is slightly larger for girls than for boys starting from grade 7. In previous studies, it is shown that the time adult beneficiaries allocate to work or leisure is not affected by the program.

[6] 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`.

[7] However, there were several universal and noncontributory programs in place in Mexico in 2010, for example universal health insurance (Conti and Ginja, 2017).

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.[8] 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.[9] 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 women 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%)[10]

[8]The sample is clustered by municipality, with enumeration areas selected by simple random sampling within strata (`https://international.ipums.org/international/sample_designs/sample_designs_mx.shtml`; last accessed 08/2017).

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

[10]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 still births.

or moved out of the household (8.3%).[11] 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 21–50 (21–35), Table A1 in the Appendix shows that the females in my sample are much more likely to be married or cohabiting and they tend to have a lower education. 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 21–50 (21–35) 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

Descriptive statistics are presented in Table 1. The first part of the table shows some demographic characteristics. The women in my sample are, on average, 29 years old and gave birth to their first child at 20. The majority of women (62%) completed at most primary school. Only 3% have a university degree. These characteristics are very similar across single and married or cohabiting women. About half of the women live in very small municipalities with less than 2,500 inhabitants. This proportion is lower for single women. 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. Single women have, at 49%, a very high likelihood to live with one of their parents. The mean monthly labor income of the household, excluding mother's income, averages 3,913 Mexican Pesos (307 US$)[12]. For married and cohabiting women, the former is mainly earned by the spouse, while for single women by the maternal grandfather. Only 2% of the eldest

[11] I also exclude females for whom at least one of the following variables is missing: Age and/or sex of the two first-born 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.

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

children in my sample work.

The fertility characteristics in the third part of Table 1 demonstrate that 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 first born males being with a proportion of 26% a bit more likely than two females. This makes sense given that the natural sex ratio at birth is slightly biased toward boys (Hesketh and Xing, 2006).

Further differences between women with and without a partner arise when it comes to labor market variables presented in the last part of Table 1. At the time of the census, 60% of single women were working compared to only 21% of women in a partnership. Of these employed women, 42% work in an informal job, averaging 37 working hours per week with an income of 2,959 Pesos per month. Single women are more likely to be in formal employment, working more hours than married or cohabiting women and with slightly higher monthly earnings.[13]

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.

I define informal employment as working on one's own account (self-employment with no employees hired), day labor, or unpaid work, whereas formal employment refers to self-employment with employees or to blue- and white-collar work.[14] In the Census data, I observe Mexican women in informal employment mainly working as store merchants (20.9%), workers

[13]Compared to the sample in Cruces and Galiani (2007), who use 2000 Mexican census data and similar sample restrictions, the proportion of women working for pay increased by around 7% in the overall sample and by around 5% in the sample of married women from 2000 to 2010. Moreover, the percentage of women in the sample who have more than two children is around 18% lower and they have about 0.5 children less in 2010 than in 2000.

[14]According to the ILO, *Informal employment* is defined as comprising employers working in their own informal sector enterprise, employees holding informal jobs (jobs without social protection and other job-related benefits) both in the formal and informal sector, and own-account workers (International Labour Organization, 2009). Unfortunately, I can neither observe whether employees are in jobs with social protection and job-related benefits nor whether employers are formally registered. Still, the job types I define as informal are those typically associated with informality (Cáceres-Delpiano, 2012). Moreover, I observe the type of health insurance households have access to. Among households with informally employed mothers, 80% are not covered by any social security institution in Mexico, i.e. they are neither insured nor are they part of the public insurance. This number is substantially lower (50%) for household with formally employed mothers.

in cultivation of maize/beans (6.0%), domestic workers (5.8%), servers of food and drinks (5.0%), and sales workers (4.5%). This observation is in line with findings in OECD (2017).

In Figure 1, I compare working hours and hourly wages between women in formal and informal employment. Of course, differences in the distribution of hours and wages over the type of employment must be interpreted with caution. It is likely that women observed in the formal versus the informal sector are very different from each other. Graph 1a shows that, on average, women work fewer hours in informal relative to formal employment. Furthermore, the working hours are 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 be observed with working hours between zero and twenty, while around 40% of those in a formal job work full-time (40–50 hours). This supports the hypothesis that the fixed costs of working are higher in formal than in informal employment. These fixed costs drive women who want to work relatively few hours out of the formal labor force. Graph 1b shows that formal pays a higher average hourly net wage than informal employment. Interestingly, more than a third of all informal workers earn only 0–5 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 household labor market outcomes. I define the treatment indicator D_i equal to 1 if a women 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 absence of treatment, i.e. with two children. For any individual, Y_{1i} and Y_{0i} cannot be observed simultaneously. 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 individuals would identify the average treatment effect, $\tau = E[Y_{1i} - Y_{0i}]$, since individuals of 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 their number of children, given earning potentials and career plans. Moreover, there probably exist other unobserved factors that influence both fertility and labor market outcomes, i.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 first-born 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:

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

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

Y_i, D_i and Z_i are defined as above. Vector X_i contains relevant control variables, including indicators of the sex of the 1^{st} and 2^{nd} child plus additional variables such as education and age.[15] 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 so-called $compliers$. The target of the estimation is the average of β_i among compliers. To estimate this Local Average Treatment Effect (LATE), the baseline specification uses 2SLS with first and second stage equations given by (1) and (2). 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 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

[15]The set of additional control variables includes 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 nonlinear in these age variables and there may be discontinuities at particular ages; for instance at the school starting age or at the minimum age to enter early childhood education.

and outcomes with mass points (Melly and Wüthrich, 2017).[16] In contrast, alternative models that deal with limited outcome variables, such as the Tobit model, often rely on very restrictive underlying structural frameworks for identification (Angrist, 2001). Similar to the NP-IV estimator (Frölich, 2007), I am able to 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 LATE assumptions.

I must assume that the sex composition of the two first-born children is as good as randomly assigned. This assumption might be violated if parents can influence the sex mix of the two first born children, for example by sex-selective abortions or discrimination in care practices for one sex. 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 the 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 certain 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).

Even in absence of manipulation, the sex ratio at birth is around 105-107 male births for every 100 female births (Hesketh and Xing, 2006). This implies that having a male child among the two-first born children makes it more likely that a couple has two first-born children of the same sex, i.e. children's sex is systematically related to the *same sex* instrument. This is a

[16]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).

problem if children's sex is related to potential labor market outcomes or potential fertility.[17] Since I cannot rule out that this is the case, I impose the weaker assumption that the instrument is randomly assigned conditional on the sex of the two-first born children and account for the latter in the analysis.

Table 2 illustrates the correlation between predetermined characteristics of women and their instrument status. Column (1) and (2) give the mean characteristics of the corresponding variable in households with two children of different sexes and the same sex, respectively. Column (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. As explained earlier, having a male child among the two firstborn children is positively associated with the instrument: parents with children of the same sex have on average a significant 1.3pp (2.2pp) higher likelihood of a male first (second) child. Given these difference between mixed- and same-sex children mothers, the main model will condition on the sex of the first- and second-born child. Additional controls include the variables listed in the upper part of Table 2, among them maternal age, in a very flexible way. However, the results are not sensitive to the inclusion of these additional controls.

To interpret the IV estimates as identifying the causal effect of fertility on parents' labor market outcomes, I further have to rule out that the sex composition of the two first-born children has a systematic effect on labor supply other than through its effect on having an additional birth. This exclusion assumption might be violated if the sex composition affects child costs (Wolpin and Rosenzweig, 2000). Lower levels of expenditures for children can result from hand-me-downs (i.e. for clothing and footwear) when children have the same sex. Wolpin and Rosenzweig (2000) demonstrate that these savings are existent for Indian households and that they account for an important fraction of the household income. Nevertheless, these results might vary significantly across countries. Bütikofer (2011), for instance, examines the Mexican

[17]There is some evidence that this might be the case. Dahl and Moretti (2008) show that having a first-born boy increases the likelihood to marry and decreases the probability to divorce. Moreover, there might be differences in monetary or time cost of raising girls versus boys (Angrist and Evans, 1998; Dahl and Moretti, 2008). At the same time, marriage, divorce, and child cost are likely to be related to maternal labor market outcomes. Moreover, Dahl and Moretti (2008) show that having a male first-born reduces the likelihood to have additional children.

Family Life Survey in 2000 and 2005. She 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 sex composition related savings were existent and affect female labor supply directly, their size would have to be meaningful enough to severely violate the exclusion restriction. In contrast to India, 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.[18] The exclusion assumption might also fail to hold if the sex composition of the two first-born 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 third 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 have no effect on 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, fertility preferences are likely to vary between settings and both of the aforementioned countries are well known for their sex bias due to extreme son preferences. I show in Section 5 that, in my sample, households with two boys are more likely to increase their family size compared to those with one boy and girl. This implies that the preference for mixed-sex siblings is more prevalent than the preference for two boys. Nonetheless, this does not rule out that households with a preference for the latter exist in my sample. Similar to De Chaisemartin (2017), I estimate the proportion of potential defiers in the Demographic and Health Survey data, which has information on the desired sex composition of children. As there is no data available on Mexico, I examine a sample of mothers from Colombia, which is geographically

[18] Data is from OECD StatExtracs/Final consumption expenditure of households, `http://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE5`; last accessed 09/2017.

and culturally close.[19] Among women whose two first-born children are a boy and a girl, 0.77% had 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. Thus, only a small fraction of women were induced to have a third child because their two children were of different sexes.[20]

Beyond the argumentation in favor of the validity of the instrument based on empirical and institutional facts, I also formally test it.[21] I apply an approach proposed by Huber and Mellace (2015) for a just identified (heterogeneous) treatment effect model with endogeneity. A rejection points to the violation of independence, exclusion, monotonicity, or any combination of them. With this procedure, one can test for necessary, but not for sufficient conditions of instrumental validity. However, given the large data set, I have a high finite sample power, thus decreasing the likelihood of incorrectly retaining a false null hypothesis. In addition, more violations of the IV validity can be detected as the share of compliers becomes smaller.[22] Applying the test to the sample, I obtain a p-value close to 1. This means that I cannot reject the hypothesis that the instrument satisfies the assumptions for validity.[23]

To sum up, both the discussion based on empirical observations for the Mexican context and the formal test procedure provide little 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 between fertility and the sex mix of the two first-born children in Table 3. Column (1) suggests that

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

[20]Angrist and Evans (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.

[21]The validity of the *same sex* instrument is formally tested for a sample of mothers in the 1980 US Census by Huber (2015). He uses the same testing procedure I outline here. Mourifié and Wan (2017) apply a different testing procedure to the 1990 US Census. Both studies find no evidence that the assumptions for the validity of the *same sex* strategy are violated.

[22]I show in Section 5 that the share of compliers in the sample is actually very small.

[23]I additionally must assume common support, which states that the support of the covariates is the same for instrumented and non-instrumented individuals. This means that for any value of the confounding variables X_i, an individual can be potentially observed with treatment and without treatment.

there is a statistically significant relationship between the treatment and the *samesex* indicator: Women with two children of the same sex are 3.3pp 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 including indicators whether the first and second born child is male (*boy1* and *boy2*, respectively) 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 born children are boys or girls, labeled *twoboys* and *twogirls*, respectively. The estimated coefficients suggest that females increase childbearing both in response to the birth of two boys and two girls, but the response to two boys is significantly smaller than to two girls (2.6pp compared to 4.1pp, 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.[24] This means that there is no relationship between *boy1* and fertility when the effect of sex composition is allowed to differ by sex. 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 $per\ se$. As outlined earlier, 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 condition on the sex of the first two children in the estimations.

The small partial R^2 of the first stage results indicates that the instruments can only explain a small amount of the variation in the treatment and, hence, the IV estimates are going to have relatively large standard errors. Still, 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.

[24] The second model allows the effect of two first-born boys and 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

In Table 4, I present the results from OLS and IV estimations of the relationship between maternal employment and fertility for all women in the sample. 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 8pp less likely to be employed, which is mainly driven by lower employment rates 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, instrumented *samesex*-effects in columns (4) and (5) reveal a positive relationship between family size and the probability of employment: having more than two compared to two children induces an increase in employment by an average of about 4.8pp. Given the relatively low employment rate in the sample, this corresponds to a 19% increase. This effect is statistically significant at the 10% level. The employment definition includes both employment in formal and informal jobs (own-account work, day labor, and unpaid work outside home). In the following, I look at formal and informal employment separately. The rise in employment seems to be driven by an increase in the likelihood to be in informal employment when fertility increases beyond two children. However, the effect of fertility on informal is not statistically different from its effect on formal employment. In Table 3, I show that mothers in my sample tend to have more children when their two first-born 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.

The comparison of the OLS and IV estimates suggests that the preferences for childbearing are inversely related to work in the labor market, particularly to work 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 hold because informal jobs offer more flexibility to combine work and family responsibilities. In the following, I look at employment responses beyond the extensive margin.

5.2 Fertility and maternal working hours responses

Table 5 shows the estimated effects of $more\ than\ 2\ children$ on hours worked per week in total and both in formal and informal jobs. Similar to the probability of employment, women with more than two children are observed to work fewer hours compared to those with two in columns (2) and (3). The coefficients turn positive when instrumenting for the potentially endogenous fertility decision in columns (4)–(7), but they are not statistically distinguishable from 0 in almost all specifications. The coefficients in the regression with informal hours as outcome variable tend to be larger than those with formal hours. Subsequently, I go beyond the mean effects and 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.

I estimate the quantiles of the potential outcome distributions for compliers based on Frölich and Melly (2013) using the *samesex* instrument.

Figure 2 illustrates the estimated quantiles of maternal working hours separately for formal (2a) and informal employment (2b).[25] 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 A2 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 A2 shows that this effect is statistically significantly different from zero. Third, having $more\ than\ 2\ children$ tends to increase some quantiles that would be equal to zero or to a small number of hours

[25] The effect of the treatment on total hours in both formal and informal employment is presented in Appendix Figure A1.

in informal employment if women had two children only. The quantiles are augmented by statistically significant 13–15 hours at this part of the outcome distribution (see third panel of Appendix Table A2). Summing up, fertility increases female labor supply in the formal sector at the intensive margin for a small fraction of women, while having mainly an impact at the extensive margin of labor supply in the informal sector.

6 Mechanisms

In Section 5, I find that an increase in fertility beyond two children leads to a positive employment response of mothers. As outlined in Section 1, this finding contrasts with the negative effect found in most previous studies. There are different potential explanations for this deviation. I already show that the increase in employment is due to a rise in informal labor. Employment opportunities in the informal sector tend to be more prevalent in low- and middle-income compared to high-income countries. Most studies, focusing on the latter, find a negative relationship between fertility and female labor supply. In the following section, I discuss three more potential reasons for my empirical results. First, I show that husbands in couples do not increase their employment propensity or hours in response to an increase in their family size, which might indicate that they face difficult labor market conditions. If work opportunities were better for women around the time of the census, then this might provide a reason why they increase their labor supply. Second, I discuss 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 on different proxies of household wealth.

6.1 Labor supply responses of married and cohabiting couples

In the following, I have a closer look at households in which mothers live together with their partners.[26] Table 6 shows the estimation results for the subsample of married and cohabiting couples in a similar way as for the overall sample before. The total employment response of

[26]For 5.1% of the women who report that they are 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.

women who live with their partners is similar to that of all women in the sample. Standard errors tend to be a bit larger in the subsample and the estimated effects are not statistically significantly different from zero. The response of the probability to be informally employed and of total working hours is positive and significant for partnered women and tends to be slightly larger (but not statistically different) than in the overall sample.

Column (1) shows that average paternal labor market attachment is much stronger than maternal in terms of participation 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). The coefficients in the IV models of paternal employment on $more\ than\ 2\ children$ are very close to zero and not statistically significant.

In a model with a unitary household, one 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 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 for some men, one would probably expect that potential wages are higher for male partners.[27] One potential explanation for the contrary results are labor market frictions that impacted men differently than women in 2010, the year of the analysis. In Mexico, male labor market outcomes were the most vulnerable to the Great Recession with decreasing participation rates and income due to the economic contraction. At the same time, females labor market success was largely unaffected by the crisis (World Bank, 2012). Therefore it might have been easier for women to earn additional income that provided for the enlarged household in 2010 than it was fr men. Another explanation is that men's labor supply is less elastic.

[27] Although this is very selective, I observe that 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.

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 (see Posadas and Vidal-Fernandez (2013) for an excellent literature review). In particular, Compton and Pollak (2014) show that geographical proximity to mothers or mothers-in-law has a substantial positive effect on the labor supply of married women with young children.

In the data, I can observe whether the parents or parents-in-law are present in the households. It is likely that women who live in the same household as their parents (or parents-in-law) are not comparable to women who live with their nuclear family. For instance, the descriptive analysis already revealed that 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, incorporating an indicator whether grandparents are present in the household as a dependent variable. I make the assumption that the sex composition of the two first-born children does not affect co-residence with grandparents other than through its impact on fertility.

Table 7 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. The negative correlation between fertility and the presence of a mother's parent is sustained even if I control for a large set of variables. When I instrument the variable of interest, $more\ than\ 2\ children$, there is a marginally significant positive relationship between family size and co-residence with maternal grandparents. The 4.4–4.6pp increase is large 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

As discussed earlier, negative budgetary consequences from fixed costs of additional children are likely to be more severe in households with low wealth. As a result, an increase in the family size of these "poor" households should have a larger impact on the mother's propensity to take up employment in order to guarantee a minimum income level for the household. 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. So it is reasonable to split the sample by these measures.

The corresponding results from the second stage using the *samesex*-instrument are displayed in Table 8. The findings provide suggestive evidence that mothers in households with relatively low household wealth particularly increase their labor supply in response to an increase of their fertility beyond two children. There is no statistically significant effect of a larger family size on formal employment or hours for any subgroup in column (2) and (4), respectively. 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 labor income when their family size increases.

7 Beyond local treatment effects

So far, I estimate and discuss 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.[28] This is important, since a researcher might be interested in a policy that affects the fertility of women who are different from those that have another child due to a variation in their previous children's sex composition.

The following discussion is based, in large part, on the concept of *marginal treatment effects* introduced by Heckman and Vytlacil (1999, 2005).[29] 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.

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

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.,

[28] I am interested in the extrapolation of the effects for compliers to non-compliers *within the sample*. A paper by Bisbee et al. (2017) looks at the extrapolation of the relationship between fertility and maternal employment *across samples* from different countries and/or time periods.

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

1996).[30] *Always-takers* have a 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 first-born 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 $LATE = E[Y_{1i} - Y_{0i}|u_0 < U_i < u_1]$.

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 in the sample, 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 9 shows the estimated proportion 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 individuals; 51% are always-takers. The small number of compliers raises concerns 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.

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 (Kowalski, 2016). Using independence, I can identify the average characteristics of always-takers, compliers and never-takers. If they were the same across all types, then this

[30] 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 the control variables.

would suggest that the LATE might be indicative for the treatment effects of non-compliers. Table 10 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 am able to 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. 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). Column (2) and (3) in Table 9 show the estimated average potential employment with and without treatment, respectively. The proportion 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 stronger labor force attachment. The comparisons of formal and informal employment, and of the working hours of never-takers and untreated compliers in Appendix Table A3 go into the same direction. Furthermore, the average employment levels with more than two children is around 5pp 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 for non-compliers.

7.2 External validity imposing further assumptions on marginal outcomes

Assuming *weak monotonicity* in the marginal untreated and treated outcome from always-takers to compliers to never-takers allows for estimating a bound on the average treatment effect for always- and never-takers, respectively (Kowalski, 2016). This assumption means that the labor force attachment is either non-increasing or non-decreasing in the distaste for additional children regardless of the actual number of children. A stronger assumption is that both $MTO = E[Y_{1i}|U]$ and $MUO = E[Y_{0i}|U]$ are linear in U_i as in Brinch et al. (2017). This means that a one percentage point change in the unobserved distaste for children U_i leads to an average increase in Y_d of λ_d. Linearity gives point identification of the MTO and MUO, and hence of the MTE function.

Figure 3 illustrates the estimated marginal outcomes and the $\widehat{MTE}$ under the assumption of weak monotonicity and linearity. The vertical axis measures the outcome of interest (here the estimated probability of employment), 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.

In Figure 3a, 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 bounds, I can infer a lower bound for the marginal treatment effect over the distribution of U_i: The former is -0.01 for always-takers and -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.

Figure 3b 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 80^{th} percentile of the distribution of U. At the same time, the $\widehat{MTE}$ is falling in U_i: low-resistance households have a higher treatment effect compared to those with a higher resistance toward having more than two children. Given the linearity assumption, I can 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 (Brinch et al., 2017). In this setting, I estimate the slope to be within the 95% confidence interval of $[-0.39, 0.18]$. Therefore, I cannot reject the null hypothesis of constant MTE.

The estimated marginal treatment effect on informal employment, which is illustrated in Appendix Figure A3, is bounded to be non-negative for all values of U_i under the weak monotonicity assumption. This suggests that the positive effect of fertility on informal employment is not specific to the group of compliers, but is found for all women in our sample. Under linearity, I find that the treatment effect is increasing in the unobserved distaste for more than two children. Appendix Figure A2 shows the estimated marginal propensity to be formally employed and the corresponding $\widehat{MTE}$ similar to Figure 3. Under the linearity assumption, the $\widehat{MTE}$ of an increase in fertility beyond two children on formal employment is decreasing in absolute terms in U_i, which is the same as for the $\widehat{MTE}$ on overall employment. It is positive for always-takers, close to zero for compliers, and negative for never-takers.

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 a 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 relatively weak assumptions.

8 Conclusion

Many low- and middle-income countries demonstrate a very dynamic evolution in both their fertility rates and their female employment levels over the past decades. However, the research on the relationship between these two variables is still surprisingly scarce. This research paper uses Mexican census data from 2010 in order to investigate the complex relationship between fertility and female labor supply. Disentangling causal effects of higher parity childbirth on labor market outcomes is complicated by the endogeneity of fertility. The exogenous variation in fertility in my estimation strategy derives from parents' preference for mixed-sex siblings, as proposed by Angrist and Evans (1998). 2SLS estimations indicate that the compliant subpopulation of women significantly increases the probability of being employed due to an increase in the propensity to be in informal employment. I also find positive responses of working hours in formal employment at the intensive margin for a very small share of women. Further analyses provide suggestive evidence 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.

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 a high availability of informal child care and employment opportunities. Thus, policies targeting small family sizes are not a reasonable measure to increase the employment of women.

On the other hand, 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 so as 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 example by providing the possibility of maternity leave,[31] 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 in this research paper 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.

[31]Mothers are currently obliged to take 12 weeks of maternity leave (six weeks before and six weeks after the birth). During this period, they receive benefits equal to their full wage if they work in the formal sector and meet some work requirements. Information retrieved from `http://www.leavenetwork.org/introducing_the_network/`; last accessed 05/2018.

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, National Bureau of Economic Research, Cambridge, MA.

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.

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).

Angrist, J. D. and Evans, W. N. (1996). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. NBER Working Papers 5778, National Bureau of Economic Research.

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.

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.

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

Browning, M. (1992). Children and Household Economic Behavior. *Journal of Economic Literature*, 30(3):1434–75.

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

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.

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

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

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.

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

Cunningham, W. (2001). Breadwinner versus Caregiver: Labor Force Participation and Sectoral Choice over the Mexican Business Cycle. *The Economics of Gender in Mexico: Work, Family, State, and Market*, pages 85–132.

Dahl, G. B. and Moretti, E. (2008). The Demand for Sons. *The 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. Endogeneity, instruments and identification.

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. *The 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.

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

Kleven, H., Landais, C., and Søgaard, J. E. (2018). Children and Gender Inequality: Evidence from Denmark. NBER Working Paper No. 24219, National Bureau of Economic Research, Cambridge, MA.

Kowalski, A. E. (2016). Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments. Working Paper 22363, National Bureau of Economic Research.

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):1611–37.

Mateo Díaz, M. and Rodríguez Chamussy, L. (2013). Childcare and Women's Labor Participation: Evidence for Latin America and the Caribbean. Inter-American Development Bank Technical Note No. 586.

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*. Chapman and Hall/CRC, London.

Minnesota Population Center (2014). Integrated Public Use Microdata Series, International. *University of Minnesota*.

Mourifié, I. and Wan, Y. (2017). Testing Local Average Treatment Effect Assumptions. *The 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.

Posadas, J. and Vidal-Fernandez, M. (2013). Grandparents' Childcare and Female Labor Force Participation. *IZA Journal of Labor Policy*, 2(1):14.

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.

Schultz, T. P. (2008). *Population Policies, Fertility, Women's Human Capital, and Child Quality*, volume 4 of *Handbook of Development Economics*, chapter 52, pages 3249–3303. Elsevier.

Staab, S. and Gerhard, R. (2010). Childcare Service Expansion in Chile and Mexico: For Women or Children or Both? Gender and Development Programme Paper 10.

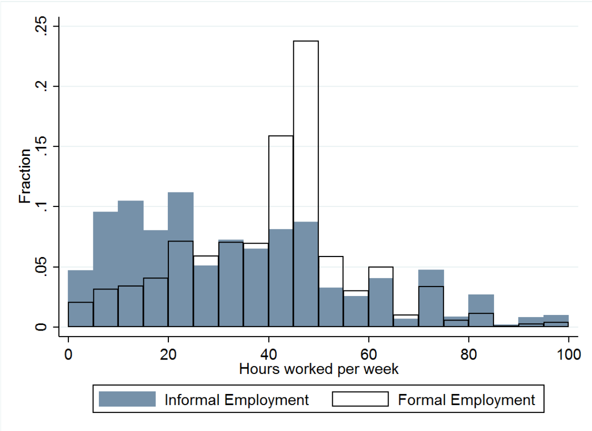
Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, 70(1):331–341.

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.

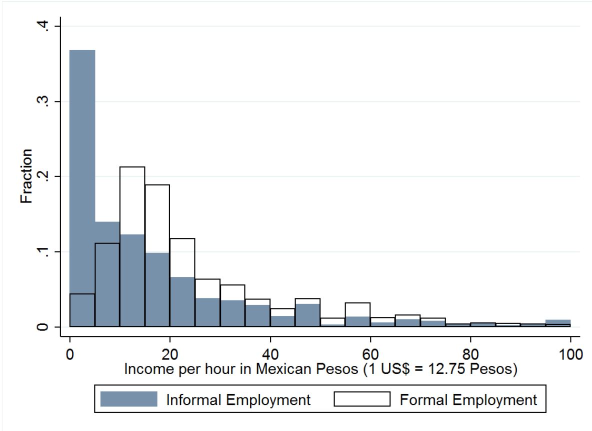
Figure 1: Distribution of working hours and hourly wages in formal vs. informal employment

(a) Working hours



Formal: Mean: 39.7, Median: 40, SD: 17.8
Informal: Mean: 33.7, Median: 30, SD: 23.8

(b) Net hourly wages

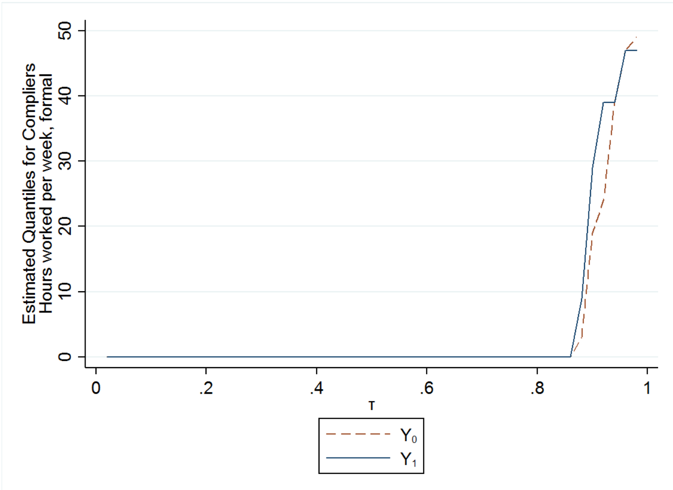


Formal: Mean: 26.5, Median: 18.7, SD: 28.8
Informal: Mean: 20.3, Median: 10.4, SD: 37.8

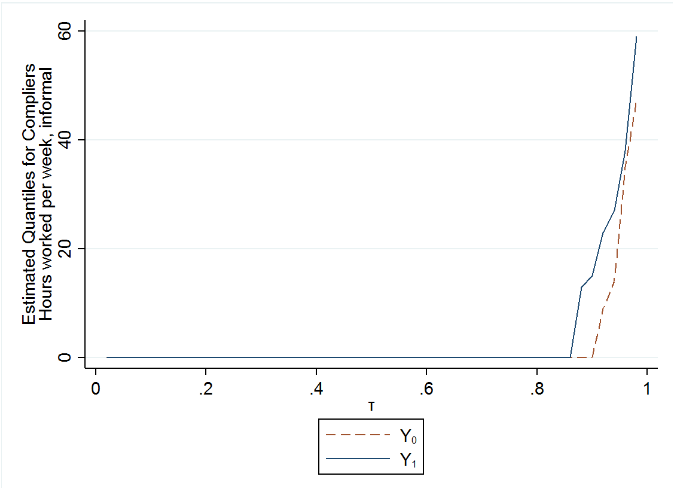
Notes: 2010 census data from Mexico with restrictions as outlined in the text. The units of observation are women aged 21–35 with at least 2 children. Informal employment is defined as in the text. The graph to the left shows the distribution of working hours and the one to the right the distribution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

Figure 2: Distributional impact of fertility on mother's hours worked per week by job type

(a) Working hours, formal

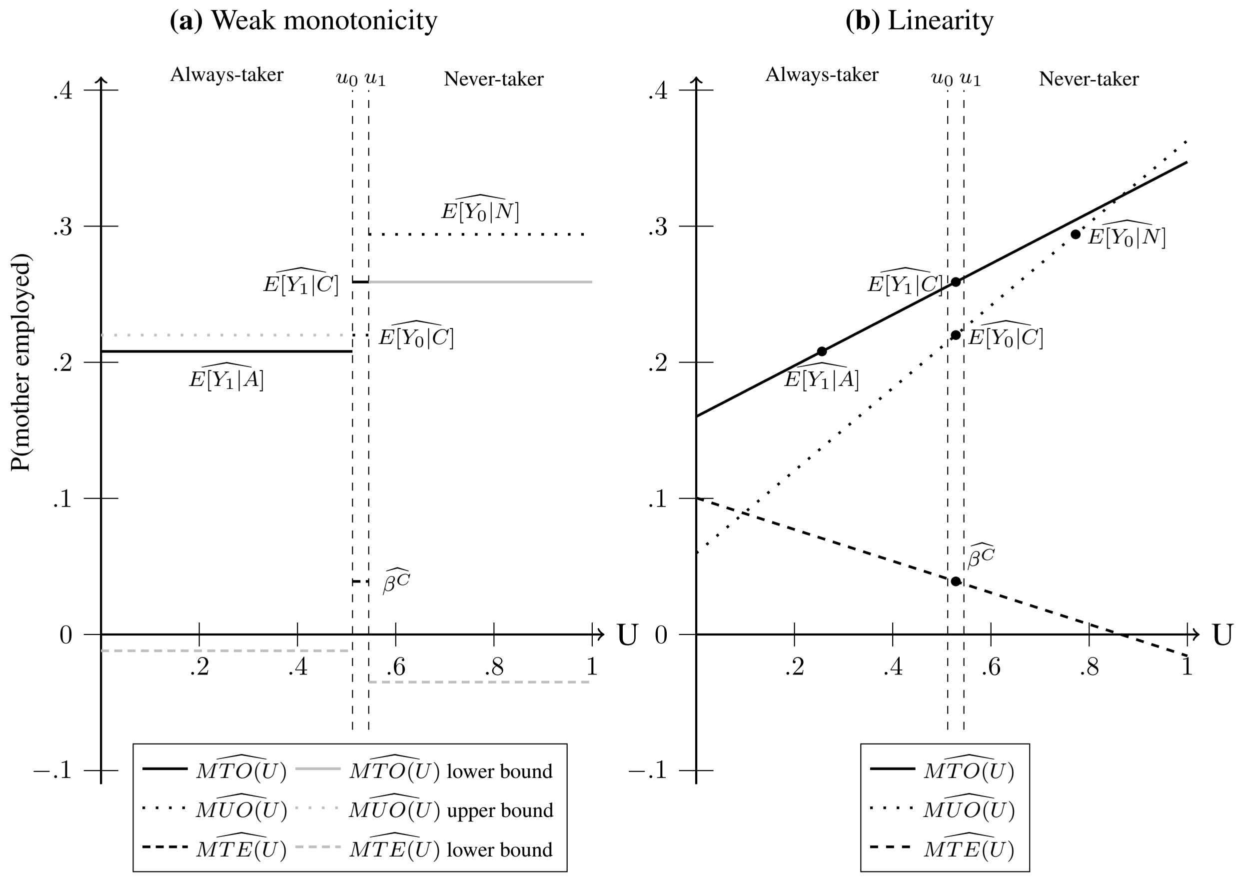


(b) Working hours, informal



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.

Figure 3: Estimated Marginal Outcomes and MTE



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 3a assumes weak monotonicity, whereas 3b assumes linearity in the marginal treated and untreated outcome.

Table 1: Descriptive Characteristics for Women in the Sample

	All		Married | cohabiting		Single	
	(1)		(2)		(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 2: Testing for 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)
Variable:				
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)
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 predetermined characteristics by the value of the instrument. Column (1) and (2) show the mean of the corresponding variable among households with a mixed and same sex composition of the two first born children, respectively. The coefficient and SE in column (3) and (4), respectively, are obtained by regressing the corresponding variable on the instrument dummy.

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)
Additional controls		✓		✓
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. Column (1) and (2) use as instrument an indicator whether the two firstborn children are of the same sex. Column (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. Additional control variables include 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Relationship between fertility and mother's employment probability

	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)
Additional controls			✓		✓	✓	✓
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. (2) and (3) show the estimated coefficients on the treatment in an OLS regression without and with additional control variables. Column (4) and (5) show the 2SLS regression results using $samesex$ as instrument without and with additional 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. Columns (2)–(7) all include indicators for the sex of the first- and second-born child. $Employed$ 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: 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)
Additional controls			✓		✓	✓	✓
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Relationship between fertility and mother's and father's 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)
Additional controls			✓		✓	✓	✓
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: 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)
Additional controls			✓		✓	✓	✓
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 8: 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					
2 or less	-0.001	0.062^{*}	1.019	2.945^{**}	327,123
	(0.036)	(0.036)	(1.589)	(1.502)	
More than 2	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. Husbands' 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. Column (1)-(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. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Compliance Types: Estimated Proportions & Fraction Employed

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

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

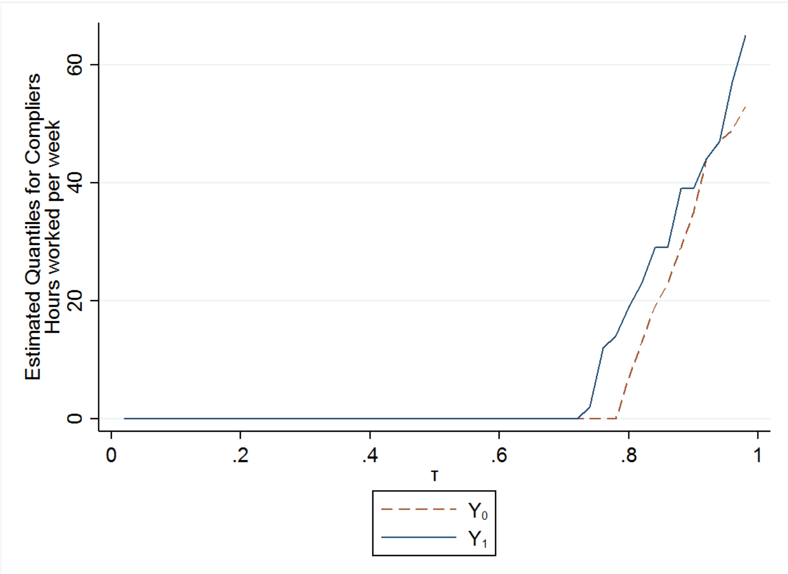
Table 10: Mean of Observable Characteristics by Types

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

Notes: This table reports average characteristics of the different compliance types in the sample.

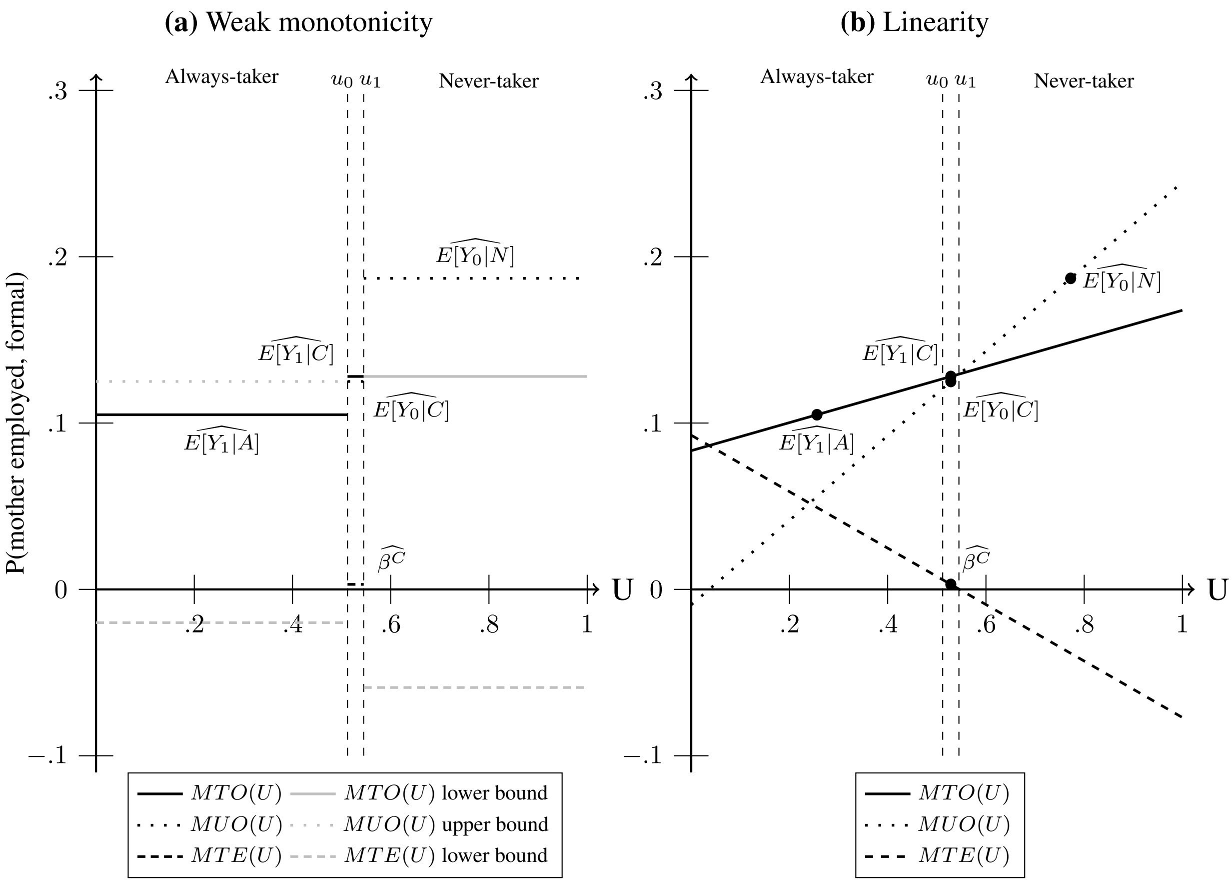
A Supplementary material

Figure A1: Distributional impact of fertility on mother's hours worked per week



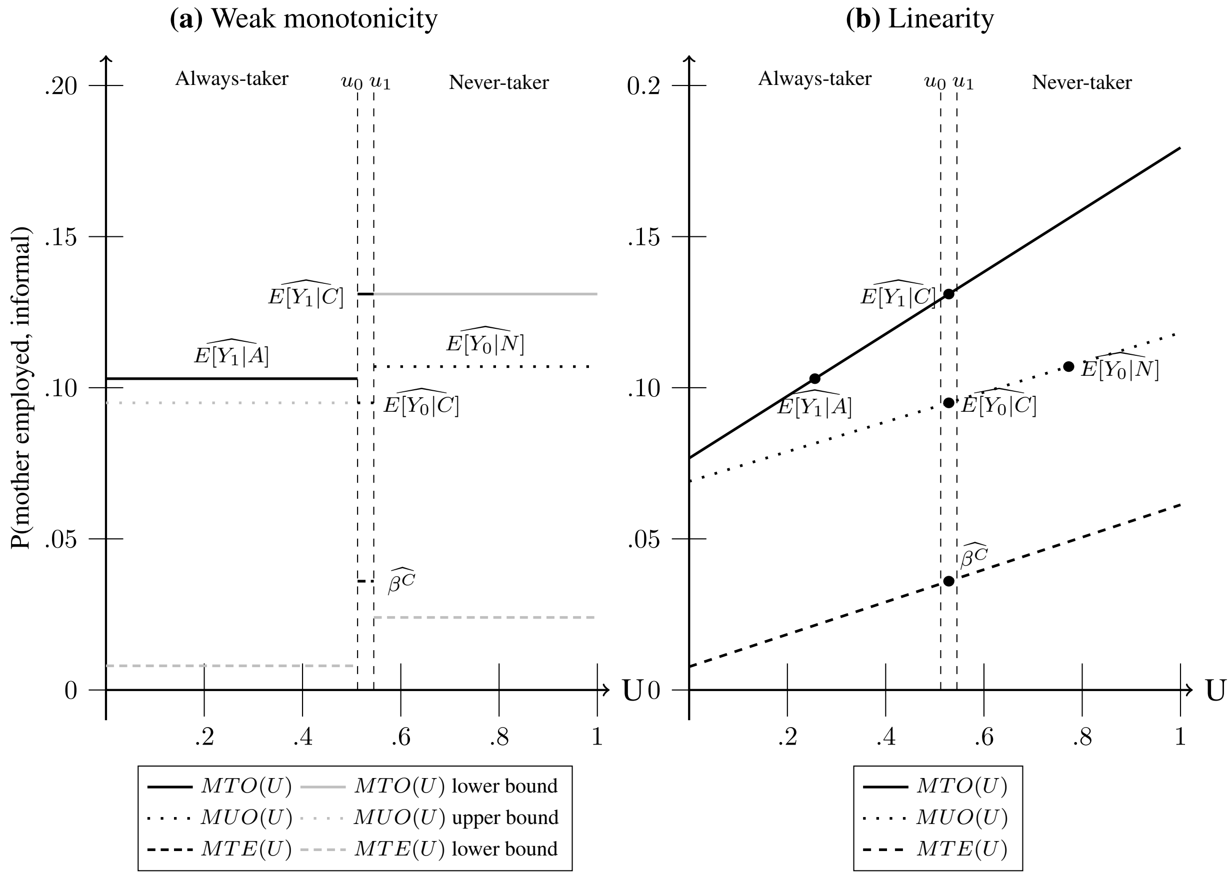
Notes: This graph illustrates the estimated quantiles for the potential outcome distributions of mother's total weekly working hours. The solid blue and the dashed red lines show the estimated quantiles of working hours for compliers with more than two children, $\hat{Q}^{\tau}_{Y_1|c}$, and with only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

Figure A2: Estimated Marginal Outcomes and MTE - Formal Employment



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.

Figure A3: Estimated Marginal Outcomes and MTE - Informal Employment



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.

Table A1: Sample compared 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)
Age at first birth (yrs)	22.341	(5.100)	20.471	(4.015)	19.762	(3.080)
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 A2: 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																
$\hat{\Delta}^{\tau}_{IV}$	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)
$\hat{Q}^{\tau}_{Y_0\|c}$	0	... 0	0	0	0	0	7	13	19	23	29	35	44	47	49	53
Hours worked per week, formal																
$\hat{\Delta}^{\tau}_{IV}$	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)
$\hat{Q}^{\tau}_{Y_0\|c}$	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)
$\hat{\Delta}^{\tau}_{IV}$	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). $\hat{\Delta}^{\tau}_{IV}$ gives the quantile treatment effect estimate for compliers at quantile τ with estimated asymptotic standard errors reported in parenthesis below the corresponding estimates. $\hat{Q}^{\tau}_{Y_0|c}$ denotes the estimated τ quantile of $Y_0|c$. In the columns (quantiles 0.04 to 0.70) that are left out, $\hat{\Delta}^{\tau}_{IV}$ and $\hat{Q}^{\tau}_{Y_0|c}$ are (as in the neighboring quantiles) equal to 0 (see also Appendix Figure A1 and Figure 2). Control variables are as in Table 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Outcomes by Compliance Type

	Formal Empl.		Informal Empl.		Weekly Hours	
$\mathbf{D_i}$	1	0	1	0	1	0
Type						
Always-taker	0.105	.	0.103	.	7.52	.
Complier	0.128	0.125	0.131	0.095	9.36	7.541
Never-taker	.	0.187	.	0.107	.	11.20