Generalizing the Causal Effect of Fertility on Female Labor Supply

Guillermo Cruces
STICERD-London School of Economics

Sebastian Galiani[*]
Universidad de San Andres

October 2003

Abstract

We study the effect of fertility on labor supply in Argentina and Mexico exploiting a source of exogenous variability in family size first introduced by Angrist and Evans (1998) for the United States. Our results constitute the first external validation of the estimates obtained for the US. External validation of empirical results is central to the making of rigorous science, but there are very few attempts to establish it. We find that the estimates for the US can be generalized both qualitatively and quantitatively to the populations of two developing countries where, compared to the US, fertility is known to be higher, female education levels are much lower and there are fewer facilities for childcare.

JEL: J13 and J22.
Keywords: Causality, Internal and External Validity, Childbearing and Female Labor Supply.

Too many social scientists expect single experiments to settle issues once and for all. This may be a mistaken generalization from the history of great crucial experiments in physics and chemistry. In actually the significant experiments in the physical sciences are replicated thousands of times…. Because we social scientists have less ability to achieve "experimental isolation," because we have good reasons to expect our treatment effects to interact significantly with a wide variety of social factors many of which we have not yet mapped, we have much greater needs for replication experiments than do physical sciences. **D. T. Campbell (1969)**

1. Introduction

One of the most significant changes in human behavior during the past century was the massive incorporation of women into the labor force. Not surprisingly, there is a wide theoretical and empirical literature attempting to explain female labor supply and its evolution (see, among others, Killingsworth and Heckman, 1986).

The relationship between fertility and female labor supply is of longstanding interest in the social sciences in general. Much of the effort has been devoted to disentangling the causal mechanisms linking childbearing and female labor supply (see, among others, Willis, 1987). Recently, Angrist and Evans (1998) (AE from now on) have made substantial progress in this area. Their identification strategy exploits the parental preferences for a mixed sibling-sex composition. In particular, as AE also confirm in their study, parents of same-sex siblings are significantly more likely to have an additional child. Since sex mix is virtually randomly assigned, an indicator variable for whether the sex of the second child matches the sex of the first child provides a plausible instrument for further childbearing among women with at least two children. AE's Instrumental Variables (IV) estimates convincingly show that additional children lead to a reduction in female labor supply and that the previous Ordinary Least Squares (OLS) estimates appear to overstate the identified causal effect of childbearing on female labor supply.

Can the causal link identified by AE be generalized? In general, we may not presume that an estimated causal relationship is universally true in the sense that it holds under all conditions with all types of people and in any circumstance. All causal statements are inevitably contingent. Thus

it is clearly useful to learn as much as possible about these contingencies and, where possible, identify the relationships that hold more consistently than others.

Any causal study faces two sources of threats to its validity: internal and external (see Campbell, 1957, Cook and Campbell, 1979 and Meyer, 1995). Most of the research effort is normally devoted to dealing with threats to internal validity. This refers to whether one can validly draw the inference that, within the context of the study, the differences in the dependent variables were caused by the differences in the relevant explanatory variables. For instance, AE assess whether there is any secular effect of child sex on female labor supply that could contaminate their IV estimates. External validity, instead, is concerned with the extent to which a causal relationship holds over variations in persons, settings and time. Thus, whenever it is possible, once an identification strategy for a causal relationship is validated internally, it is worth inquiring about the external validity of the results obtained.

Related to external validity is the idea of causal generalization, which is concerned with specifying the range of application of a causal mechanism that has been identified with at least one instance of a treatment and outcome and at least one sample of persons and settings (Cook, 2000). In practice, there is a sense in which all causal generalization is about interpolation and extrapolation. Rubin (1992) suggests that causal generalization is about estimating a response surface, i.e., mapping a third variable to an estimated causal relationship. While this is clearly an advisable procedure in conceptual terms, it is difficult to attain in empirical work.

In this perspective, we investigate the extent to which the causal link identified in AE can be generalized to the context of developing countries where, compared to the US, fertility is known to be higher, female education levels are much lower and there are fewer facilities for childcare. Thus, we question whether in such different socioeconomic environments childbearing also leads to a reduction in female labor supply and the effects are of the same order of magnitude.

More generally, we believe there is a potentially high reward in replicating valid empirical strategies. Ultimately, the external validity of causal estimates is established by replication in new

3

data sets (Angrist, 2003). In addition, external replication of valid identification strategies would lead, to the extent that it is possible, to causal generalization. Thus, replication studies would broaden our knowledge about the causal relations we are interested in.

The rest of the paper is organized as follows. First, we describe the data used and provide a set of summary statistics. Second, we present the methodology employed and our main empirical results. Third, we conduct a set of robustness checks in order to validate our estimates from existing threats to their validity. We finally provide the initial steps towards the causal generalization of AE's results. Conclusions follow.

2. Data and summary statistics

Data description

We use two large datasets gathered from the Mexico 2000 and the Argentina 1991 censuses. The Argentine data is from the 1991 National Census of Population and Housing (*Censo Nacional de Población y Viviendas*) conducted by the National Institute of Statistics and Censuses (*Instituto Nacional de Estadísticas y Censos*, INDEC). The Mexican data is from Mexico's 2000 XII General Census of Population and Housing (*XII Censo General de Población y Vivienda*), conducted by the National Institute of Statistics, Geography and Computing (*Instituto Nacional de Estadística, Geografía e Informática*, INEGI).

Both censuses were conducted by means of two questionnaires, a simple version for the whole population and an extended version covering nationally representative samples. We exploit the full extended questionnaire datasets for both censuses, resulting in unusually high sample sizes. For Argentina, we have data on 16,023,180 individuals and 4,287,580 households, covering around 50 percent of the whole population. For Mexico the sample consists of 10,099,182 individuals and 2,312,034 households (10.6 percent of the total population of 97,483,412 persons and 22,268,916 households). We provide a short data appendix describing the construction of our final samples.

4

We concentrate most of our analysis on the *Worked for pay* indicator variable (see Table I), which is also the main output variable considered by AE. Even though AE cover a wider set of labor market outcomes, in the case of Mexico, we are also able to study the impact of childbearing in women and household income since this information also was collected in the census.

We study the labor supply of women with children. We limit our sample to women between 21 and 35 years old, with at least two children, and whose oldest child was at most 18 years old at the time of the census. AE also exclude women whose second child is younger than a year old. In order to establish the generality of our results, we work with two different datasets: the "complete sample" as defined above and the "AE sample", which is the "complete sample" but excluding women whose second child was less than a year old. Finally, as in AE, all the results are reported for total and married women only, where we defined married as being legally married and living with the spouse (i.e., excluding de facto cohabitation). The final sample sizes are between 456,000 and 653,000 observations for Argentina and 355,000 and 493,764 for Mexico.[1]

Descriptive statistics

Descriptive statistics and variable definitions for covariates, instruments, and our main dependent variable (*Worked for pay*) are given in Table I. The first points that stand out refer to labor force participation and fertility. The female labor force participation rates for our married AE samples are much higher (30.5 percent) in Argentina than in Mexico (22 percent), but they are both significantly lower than the US figures for equivalent samples (52.8 percent for 1980 and 66.7 in 1990). Both in Argentina and Mexico, female labor supply is lower for married women than for unmarried women. With respect to fertility, the average number of children is higher for married women in Mexico (3.035) than in Argentina (2.985), and substantially higher than the respective US figure (around 2.5 in both the 1980 and 1990 censuses).

---

[1] See the data appendix for details on the construction of our datasets.

The causing variable in the labor supply reduced-form regressions we report here is the indicator *More than two children*. The first instrumental variable for *More than two children* is the indicator *Same sex*. The Table also shows averages for the two components of *Same sex*, the indicators *Two boys* and *Two girls*. In both Argentina and Mexico, more than 50 percent of the women in any of our samples have a third child while in the US the same figure is only about 36 to 40 percent. We also report indicators for whether the first and second children were boys. Finally, we report some demographic variables including measures of mother's age and age at first birth. In the regressions, we also control for education of the mother and the spouse, using categorical information, which is not included in Table I for the sake of brevity.

[Insert Table I about here]

Our strategy to identify the causal effect of fertility on female labor supply relies on the virtually random assignment of *Same sex*. One simple check on this claim is to compare the demographic characteristics of people who have same-sex and mixed-sex sibling compositions. There is no statistically significant difference at the standard levels for the age at first birth in Argentina and Mexico, and for age in Mexico. The difference in age between the two groups is significant at the 5% level in three of our samples for Argentina, but we attribute this to the high precision of the estimates – the difference is only around 0.02 years.[2]

3. Estimation strategy and main results

Estimation strategy

Following AE we report estimates of the effect of childbearing on labor supply by OLS and Two-Stages Least Squares (2SLS). Angrist (2001) argues that the problem of causal inference with Limited Dependent Variables is not fundamentally different from causal inference with continuous outcomes. If there are no covariates or the covariates are sparse and discrete, linear

---

[2] As an additional check, there was no difference in a series of education level indicators between women with same sex and mixed sex siblings. These results are available from the authors upon request.

models and associated estimation techniques like 2SLS are no less appropriate for LDV than for other types of dependent variables. This is because conditional expectation functions with discrete covariates can be parameterized as linear using a saturated model, regardless of the support of the dependent variable (Angrist, 2001). Thus, we first present 2SLS estimates of the parameter of interest from models where we saturate the set of control variables. However, since the conditional expectation function (CEF) of the potential outcomes of interest is also a function of the causing variable, we denote this empirical model as the "almost saturated" model. We then present 2SLS estimates of the parameter of interest from more parsimonious models. Finally, as a robustness check, we also report estimates from an IV estimator developed by Abadie (2003), which allows a flexible nonlinear approximation of the causal response function. Abadie's IV estimates have a robust causal interpretation regardless of the shape of the actual CEF for potential outcomes since identification is attained nonparametrically.

The AE application deals with the effect of going from a family size of two children to more than two children. Let D_i be an indicator for women with more than two children in a sample of women with at least two children and let Y_{1i} be the labor supply of mother i if D_i equals one and Y_{0i} denote her labor supply otherwise.

The simplest option to estimate the causal effect of childbearing on labor supply is by means of the following linear, constant-effects model:

$$E[Y_{oi}] = X_i^{'} \beta \qquad (1a)$$

and

$$Y_{i1} = Y_{oi} + \alpha \qquad (1b)$$

where X is a vector of control variables. These assumptions lead to the following linear causal model:

$$Y_i = X_i'\beta + \alpha D_i + \varepsilon_i \qquad (2)$$

that is easily estimated by 2SLS if the IV Z_i (*Same Sex*) satisfies: a) $\{Y_{0i}, Y_{1i}, D_{0i}, D_{1i} | X_i\}$ is independent of Z_i, b) $P[D_i = 1 | X_i, Z_i] \neq P[D_i = 0 | X_i, Z_i]$, and c) without loss of generality, $D_{1i} \geq D_{0i}$.

   If the additive, constant-effect assumptions in (1a) and (1b) do not hold, a 2SLS estimate of model (2) does not identify the average causal effect without further assumptions. However, under the assumptions stated above, 2SLS identifies a Local Average Causal Effect (LATE), i.e., the average effect of treatment on those whose treatment status is changed by the instrument (see Imbens and Angrist, 1994). Angrist et al. (1996) refer to this group as the population of compliers.[3]

Basic results: Sex mix and Wald estimates

   Table II contains the essence of the estimation strategy adopted here as well as our main results. It presents the means of the variables involved in the identification of the causal effect of interest (*Number of children*, *More than two children*, and *Worked for pay*) for all the samples considered and also by treatment status with respect to the binary instrument *Same sex*.

   First, note that in both Argentina and Mexico the probability of having two children of the same sex is just above one half. As expected, in both countries and for all samples women who had two children of the same sex have a higher probability of having a third child (and, naturally, also a higher number of children) than women who had two children of different sex (*Mixed sex*). The difference in these probabilities is around 3-4 percentage points in Argentina and 3-3.6 in Mexico, depending on the samples; women with two children of the same sex have on average 0.05-0.07 more children in both countries. These differences, significant at the 1 percent level, are

---

[3] Even though, as noted by Abadie (2003), in general, the coefficients associated to the control covariates X do not have a similar interpretation.

remarkably close to those found by AE for the US (0.07-0.08 more children), and this is clear evidence of the presence of the sex mix preference phenomenon in Argentina and Mexico.

With respect to labor supply, there is also some evidence of a small but significant difference in employment between women who had two children of the same sex and women with mixed sex siblings. The latter group has a participation rate around 0.3-0.4/0.2-0.3 percentage points higher (Argentina/Mexico), and these differences are statistically significant at the standard levels of confidence. In the US, AE find a similar pattern, although the differences are larger (around 0.8 percentage points for 1980 and 0.5 for 1990).

The last four lines in the two panels report Wald and OLS estimates of the effect of the total number of children and having more than two children on the probability of working for pay. The Wald estimates are equivalent to the simplest instrumental variable estimates obtained by relying on *Same sex* as an instrument for *Number of children* or *More than two children*, when no other covariates are included in the CEF. Thus, generally, they estimate the LATE parameter. Wald estimates are obtained straightforwardly as the ratio of the reduced-form relationship between *Worked for pay* and *Same sex* and between *Number of children* or *More than two children* and *Same sex.*

[Insert Table II about here]

Even with a small variation in the dependent variable, all Wald estimates for *Worked for pay* are strongly significant, with only two out of sixteen coefficients significant at the 10 percent level and most of them being different from zero at the 1 percent level of statistical significance. The results for Argentina imply that an additional child reduces the labor supply of those women whose fertility has been affected by their children's sex mix by about 5-6 percentage points, while having more than two children reduces their labor supply by about 9-10 percentage points. For Mexico, the effects are still significant but slightly lower, around 3.8-5 and 6.8-9.2 percentage points respectively. These results are quite close to the 1990 US estimates reported by AE (6.3

and 8.4 percentage points), although they are lower than the US 1980 effects (13.3 percentage points for number of children and 10.4 for more than two children).

Finally, it should be noted that for both instrumented variables the LATE estimates are larger than the simple OLS estimates (in absolute value). The rest of the paper discusses 2SLS estimates of regression models relating labor market outcomes to fertility and a variety of exogenous covariates.

Two stages least squares results

As pointed out by AE, we can use 2SLS to control for any secular additive effects of child sex when using *Same sex* as an instrument. This is desirable because *Same sex* is potentially correlated with the sex of either child, which is of concern if the sex of either child affects labor supply for reasons other than family size. Thus, among the list of control covariates included in the regression models reported here, we include additive effects for the sex of each child.

In addition, the *Same sex* indicator is easily decomposed in two variables indicating the sex composition of the first two children, *Two boys* and *Two girls*, leading to an overidentified model. AE show that this is useful because bias from any secular effects of child sex on labor supply should be different from these two instruments, while the labor supply consequences of childbearing seem likely to be independent of whether *Same sex* equals *Two boys* or *Two girls*. Thus, an appropriate specification test is the Sargan test or test of overidentifying restrictions (see Angrist and Evans, 1998).

As also noted by AE, when *More than two children* is instrumented by both *Two boys* and *Two girls*, it is not possible to control of the effects for the sex of each child, and so we report results that control only for the sex of the first child. In addition, we also include the age of the woman and her age at first birth as demographic control variables. While AE also include indicators for the woman's race, such information is not available in the Argentine and Mexican data.

The control variables included in the almost saturated models we report in Tables III and IV are a set of forty-nine mutually exclusively indicator variables that result from the interaction of the control variables age, age at first birth, sex of first child and sex of second child, once both age and age at first birth are mapped into five categorical indicator variables each.[4] Obviously, in addition, in Table IV, we include among the regressors the causing variable *More than two children.*

[Insert Table III about here]

Table III presents the results of the OLS first stage regressions. In the bottom panel of the Table we include the spouse controls instead of the women controls in the same way. The strength of the instruments is widely confirmed. The first row indicates that in both countries women who have two children of the same sex have about a 3.5 percentage points higher probability of having a third child. These coefficients are in both countries very close to the ones estimated in Table II without conditioning on any covariates,[5] suggesting that the impact of the instrument on fertility is independent of the set of added covariates.

The results for *Same sex* are a weighted average of the effect of having two girls and two boys, which are presented in the second and third rows of the Table. Women who have two girls have on average a 4.38-5.29 (Argentina) and 4.03-4.61 (Mexico) percentage points higher probability of having a third child, but these figures are markedly lower for women who had two boys (2.36-3 and 2.26-2.84 percentage points respectively). When compared with the US, AE find a larger effect of the *Same sex* instrument in the US, which is closer to 6-7 percentage points. They find a difference between women who had two boys and women who had two girls in the 1980 US Census, but for the 1990 US data the two coefficients are virtually the same. This relatively large difference between the effect of two boys and two girls seems to be an

_____

[4] The five age category indicators were chosen to contain approximately the same number of observations, and were defined as 21-25, 26-28, 29-30, 31-32 and 33-35 for age, and 17 or less, 18-19, 20-21, 22-23 and 24 and more for age at first birth.
[5] See difference of *More than two children* between *Same sex* and *Mixed sex* in Table II.

idiosyncracy of the developing countries we study. The panel at the bottom of the Table presents the results of the same first stage regressions, but using the covariates corresponding to the women's spouse. The results are fairly similar to those for married women. Thus, the evidence in Table III confirms the presence of a strongly significant mixed sex sibling preference in our data, which is the basis of the IV estimation results presented in the rest of the paper.

The first row in Table IV presents the results of estimating equation (2) by OLS. The dependent variable is *Worked for pay*, and the covariates as described above. The OLS coefficients of *More than two children* are virtually unaffected by the inclusion of these covariates, and they are all different from zero at the 1 percent level of statistical significance. A third child reduces the probability of work in the Argentine data by 9.17-9.69 percentage points for all women and 7.95-8.28 percentage points for married women, and these effects are slightly lower in Mexico (8.37-9.02 and 7.57-8.12 percentage points respectively). Of course, these OLS coefficients may be biased estimates of the average causal effect of fertility on female labor supply because of the potential endogeneity of fertility. The second row presents the main IV estimates using *Same sex* to instrument *More than two children*. The values of the coefficients of the instrumented variables are broadly in the same range as their OLS counterparts, fluctuating between 6.31 and 8.86 percentage points for Mexico and 7.9 and 9.58 percentage points for Argentina.[6] Most of these coefficients are significantly different from zero at the 5 percent level (10 percent for total samples in Mexico). In comparison with the US, the typical AE result shows a much larger difference between OLS and IV.

[Insert Table IV about here]

---

[6] A Durbin-Wu-Hausman test of differences between the IV and OLS coefficient estimates does not reject the null hypothesis at conventional levels of statistical significance.

The following row in Table IV presents Abadie's (2003) IV estimates.[7] The results are almost identical to the 2SLS just described, showing that our almost saturated model captures extremely well the CEF of female labor supply. This result coincides with those presented in Angrist (2001) and Abadie (2003).

The next row presents the IV results when the *Same sex* instrument is decomposed into its *Two boys* and *Two girls* components, allowing us to test the validity of the over-identification of the model. The results are marginally different from the estimates with *Same sex* alone: all estimates with the two instruments are smaller (in absolute terms) than the one obtained by instrumenting with *Same sex* only. However, the differences are never statistically significant at conventional levels of significance. Finally, the level of significance of the coefficients is also affected: the results in the Mexican whole samples, which were only significant at the 10 percent level with *Same sex* alone, do not reach the standard levels of confidence any more. As noted by AE, however, the theory behind the estimation refers mainly to married women in nuclear households. Finally, we also report a standard Sargan test of over-identifying restrictions. The results are reassuring. For only one case (whole sample, Mexico), we reject the null hypothesis at the 5 percent level, but we fail to reject it at the 1 percent level. In all other cases, we do not reject the Sargan null hypothesis at the conventional levels of significance.

The bottom panel of Table IV presents the results for *Spouse worked for pay* as the dependent variable, with controls (age and age at first birth) corresponding to those of the spouse. Despite the presence of a small but strongly significant effect in the OLS regressions (3.24-3.3 percentage points for Mexico, 1.25-1.5 for Argentina), none of the IV estimates are significant at standard levels of confidence. The size and level of confidence of the estimated coefficients are similar in

---

[7] We implement a simple two-step version of Abadie's (2003) estimator based on a linear specification of the Local Average Response Function. In the first step, we estimate by OLS the model $Z_i = X_i^{'} \gamma + e_i$. The predicted values from this regression are then used to construct the estimated weights K_i (Abadie 2003, Theorem 3.1). We finally use these weights to estimate the model given by equation (2), $Y_i = X_i'\beta + \alpha D_i + \varepsilon_i$, by weighted least squares (Abadie 2003, Equation 14).

the US for the 1980 sample, although for the 1990 sample, AE find a positive and significant effect of fertility on spouse labor supply.

The upper panel of Table V presents a more parsimonious version of the model reported in Table IV. The controls include the continuous variables *Age* and *Age at first birth*, and indicators for the sex of the first and/or the second child. In the bottom panel we also include a set of regional control dummy variables. The estimates of the parameter of interest are unaltered in both cases with respect to the ones reported in Table IV. The difference between the two versions of our main model is small, indicating that our instruments are orthogonal to the additional controls and geographic indicators. This result validates our strategy of using linear models to identify the parameter of interest.

[Insert Table V about here]

So far we have only studied the effect of fertility on the *Worked for pay* indicator. However, the Mexican data contains detailed income information, allowing us to estimate models given by equation (2) with alternative dependent variables. Table VI presents the results for the regression models with the logarithm of the woman's labor income, logarithm of total household income and the logarithm of income of other members of the household ("non woman income"). The first panel presents the estimates of the effect of fertility on woman's labor income (note that the women who did not work for pay were assigned a zero value). The IV coefficients of *More than two children* are significant at the 10 percent level, with the exception of the AE sample (not significant) and the whole married sample (5 percent). The results imply an elasticity of around 0.5 with respect to a third child, which is much lower than the OLS estimate, which fluctuates around 0.7-0.8 and is highly significant. Again, the data does not reject the validity of the overidentifying restrictions imposed. The results for the log of total household income (second panel of Table VI) and for the log of non-woman income (third panel) are fairly similar: while the

OLS coefficients are strongly significant, none of the IV estimates are different from zero at the standard levels of confidence.

[Insert Table VI about here]

Summing up, we have provided new evidence in support of the internal validity of exploiting *Same sex* as an instrument to identify the causal effect of fertility on labor market outcomes for women with at least two children. Nevertheless, contrary to what AE find for the US, we fail to find significant differences between the IV and OLS estimates. It is worth remembering that, in general, the *Same sex* IV strategy identifies the average effect of having more than two children on those whose fertility decisions are changed by the instrument (compliers), while OLS is suspected to fail at identifying the same effect averaged over the whole population. Thus, with this interpretation in mind, the finding is not worrisome.[8]

4. Heterogeneous effects and robustness checks

Education level and heterogeneous effects

The results from the previous section did not consider the possibility of some heterogeneity in the effects of childbearing on women's labor supply. Table VII fills in this gap, breaking down our main regressions by the education level of the woman. We first divided our samples for each country into three roughly equally sized and homogeneous groups. For the Argentine sample, these groups are composed of those who completed primary school or less, those with some high school, and those with some higher education. For Mexico, the three categories are less than primary, primary education, and high school and beyond. The three panels in Table VII present the results of the non-saturated models for these three categories, with *Same sex* as the instrument for *More than two children* in the IV estimations.

15

[Insert Table VII about here]

In the Argentine dataset, we find significant results (at the 5 percent level) in the less educated group only for married women, with the IV coefficient estimates of *More than two children* fluctuating around 10 percentage points – clearly higher than the corresponding OLS values (7-7.5 percentage points) and slightly higher than the average effects in Tables IV and V. For women with some high school, the estimated fertility effects in the IV regressions are also different from zero at conventional levels of statistical significance and quite similar to the ones estimated for the low education group (around 11 percentage points). Finally, despite significant and negative OLS coefficients, none of IV estimates are even marginally significant for the higher education group.

In the Mexican dataset, we find systematically significant coefficient estimates only for the low education group. Moreover, the size of these coefficients is remarkably higher than previous estimates for the whole sample: the coefficient of *More than two children* is around 20 to 25 percentage points, compared with only 3 to 5 percentage points in the OLS estimation and 6-8 percentage points in Tables IV and V.

These results by education group are notably close to those reported by AE in the US. For both the 1980 and 1990 Census, they find that the fertility effect is not significant for the higher education group, but that for both high school and lower education groups the IV estimates are significantly different from zero.

Robustness checks

Table VIII presents two robustness checks of our main results. The upper panel presents the results of the estimation for an alternative definition of our dependent variable (detailed in the data appendix). The *Worked for pay* variable in the Argentine and Mexican data refers to the week before the census, while in the US it refers to the previous year. We present the results

---

[8] See Card (2000) for a similar argument in reconciling IV and OLS estimates of the effect of schooling on

using *Active labor force status* (equal to one if in paid employment or looking for a job, zero otherwise) as the dependent variable as an alternative proxy of the US indicator. Reassuringly, the results are similar to those in Tables IV and V, both in the size of the coefficients and their levels of significance.

Finally, we noted in the data appendix that we had discarded around 100,000 women from the final Argentine sample and around 50,000 from the Mexican sample because the total number of surviving children of the woman did not match the number of children matched to her in the household (see data appendix). The bottom panel in Table VIII contains the estimation using the sample without the aforementioned adjustment. The coefficients are again very close to those in Tables IV and V.

[Insert Table VIII about here]

5. Generalizing AE causal estimates

So far, we have provided new supporting evidence on the internal validity of exploiting *Same sex* as an instrument to identify the causal effect of fertility on labor market outcomes for women with at least two children. What is more, AE results seem to be qualitatively valid also in developing countries –at least in the two countries we studied here. Both in Argentina and Mexico, childbearing also leads to a reduction in female labor supply.

The next question is whether the effects are of the same order of magnitude or whether they differ substantially. A response surface of the estimated effects against real GDP per capita for the 1980 and 1990 samples considered in AE and those studied here seems to suggest that the effect of fertility on female labor supply increases with the wealth of the country. However, this assertion would entail too much extrapolation. What is more, a formal test of the following hypotheses (using self-explaining notation):

---

earnings.

$$H_0: \alpha_{US90} = \alpha_{AR91}$$

$$H_0: \alpha_{US90} = \alpha_{MX00}$$

$$H_0: \alpha_{MX00} = \alpha_{AR91}$$

does not reject the null at standard levels of statistical significance. Thus, we can assert that in both the US and Argentina and Mexico, the average effect of going from a family size of two children to more than two children for those whose treatment status is changed by the *Same sex* instrument is statistically similar. Thus, AE estimates appear to be generalized both qualitatively and quantitatively to dissimilar populations that display obvious observable differences. In addition, as in AE, we find that the effect of fertility on female labor supply disappears for educated women, and that husbands do not change much their labor-market behavior in response to a change in family size.

6. Conclusion

We study the effect of women's fertility on labor supply in Argentina and Mexico by means of an instrumental variable estimator first introduced by Angrist and Evans (1998) for the United States. Our investigation shows that the "mixed sex sibling preference", the basis of the Angrist and Evans estimation strategy, is present in Argentina and Mexico. Moreover, we provide new supporting evidence of the internal validity of exploiting *Same sex* as an instrument to identify the causal effect of fertility on labor market outcomes for families with at least two children.

Our purpose was twofold. On one hand, we established a causal link between fertility and labor supply in Argentina and Mexico. But, most importantly, while the IV strategy of AE has been subject to intensive tests of its internal validity in the context of United States Census data, our results constitute the first check – that we are aware of – of the external validity of AE estimates. The idea of external validation of empirical results is central to the making of rigorous science, but there have been very few attempts to establish it.

We believe that the results of our application to developing countries strengthen the claims in this literature. We find that AE estimates for the US can be generalized both qualitatively and

quantitatively to the populations of two developing countries with notable observable differences.

Finally, we find that the effect of fertility on female labor supply disappears for educated women

while husbands do not change much their labor supply in response to a change in family size.

Data appendix

The Argentine dataset contains information on 16,023,180 individuals and 4,287,580 households, from a total population of 32,245,467 individuals and 8,927,289 households. We constructed this dataset from the original data tapes. The Mexican dataset covers 10.6 percent of the total population of 97,483,412 persons and 22,268,916 households, yielding a sample size of 10,099,182 person records and 2,312,034 household records. The Mexican data used to construct our final samples was set up and provided to us by Sobek *et al.* (2002).

Matching women and their children
For both Argentina and Mexico, the relationship variable linking members of a household indicates kinship with respect to the head of the household only. In order to match women with their own children, we restrain the sample to females who are heads or spouses of the heads of households. The Mexican dataset, however, contains also a more complex constructed relationship, which may allow linking of children to other members of the household (see Sobek et al., 2002). However, in order to use comparable data, we do not rely on this constructed variable.

In order to avoid assigning all children of a male head to his current partner, who might not be the mother of all children in the household, we first check that the reported number of children alive (as asked for in a specific census question for both countries) coincides with the number of children in the household matched to the woman, restraining our samples to women for whom both numbers coincide. Finally, we made some extra adjustments based on the age of the woman and/or her husband. We discarded a small number of observations for which the age of the mother at first birth was less than 14, taking this as an indicator of data entry errors or misallocated children, since most of the ages were far too low. We also dropped from our final samples a very small fraction of married women for whom the husband's age at first birth was less than 14.

Worked for pay indicator
Worked for pay is the only outcome variable available in the Argentine census, while for the Mexican census income is also available. Neither of the two datasets collects information about hours or days worked. Thus, the main dependent variable in this paper is the *Worked for pay* indicator. In the Argentine census, an individual is classified as working for pay (*Worked for pay* indicator = 1) if he or she works and is not a family worker without remuneration. Thus individuals working for pay include employees (wage earners), self-employed, owner-managers and civil and domestic servants. In Mexico, we use the same definition and classify an individual as being working for pay if he or she does remunerated work.

In terms of generalizing AE results, a possible nuisance is that the *Worked for pay* variable in the US census refers to the previous year, while in both Argentina and Mexico it referrers to a week of reference. Thus, in our robustness checks section we report our main results using a labor force status indicator as the dependent variable, which is equal to one if the woman works for

pay, works without remuneration or is unemployed, and zero otherwise.[9] This is the group of women that works for pay on a regular basis and which might be closer to the AE definition for the US.

Income variables, Mexico

Income data is collected for every individual in the household aged twelve or more. We considered the household's income to be valid only when the income of every eligible member was reported, resulting in a sample of 470,524 households – 23,240 households (4.71 percent) were discarded for the income regressions. Finally, we also discarded a total of 19,289 (3.94 percent) observations since the data was inconsistent (women working for pay that have no earned income, and women not working but who reported labor incomes). Thus, the final total sample for the women's income regression consists of 469,542 observations. As AE, we imputed a value of one to those households (women) that have valid zero earnings, leading to a zero value when log-transformed.

---

[9] Only 2.43 percent of the women in our final sample for Argentina and 0.4 percent for Mexico are classified as unemployed.

References

Abadie, A. (2003): "Semiparametric instrumental variable estimation of treatment response models", *Journal of Econometrics* 113, pp. 231-63.

Angrist, J. (2003): "Treatment Effect Heterogeneity in Theory and Practice", NBER WP 9708, Cambridge, MA, US.

Angrist, J. (2001): "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice", *Journal of Business and Economic Statistics* 19, pp. 2-16.

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

Angrist, J. and W. Evans (1998): "Children and their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", *American Economic Review* 88(3), pp.450-77.

Card, D. (2000): "The causal effect of education on earnings" in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Vol. 3A. Amsterdam: North-Holland.

Campbell, D. T. (1969): "Reforms as experiments", *American Psychologist 24,* pp. 409-29.

Campbell, D. T. (1957): "Factors relevant for the validity of experiments in social settings", *Psychological Bulletin* 54, pp. 297-312.

Cook, T. and D. T. Campbell (1979): *Quasi-Experimentation: Design and Analysis issues for field settings*. Boston: Houghton Mifflin.

Cook, T. (2000): "Toward a Practical Theory of External Validity", in L. Bickman (ed.), *Validity and Social Experimentation*: *Donald Campbell's Legacy*. Sage Publications.

Imbens, G. and J. Angrist (1994): "Identification and Estimation of Local Average Causal Effects", *Econometrica* 62, pp. 467-75.

Killingsworth, M. and J. Heckman (1986): "Female Labor Supply: A Survery" in Orley Ashenfelter and Richard Layard (eds.), *Handbook of Labor Economics*, Vol. 1. Amsterdam: North-Holland.

Meyer, B. (1995): "Natural and Quasi-Experiments in Economics", *Journal of Business & Economic Statistics* 13 (2), pp. 151-62.

Rubin, D. (1992): "Meta-Analysis: Literature Synthesis or Effect-Size Surface Estimation?", *Journal of Educational Statistics* 17, pp. 363-74.

Sobek, M., S. Ruggles and R. McCaa (2002): "Integrated Public Use Microdata Series-International: Preliminary Version 1.1", Minneapolis: Minnesota Population Center, University of Minnesota.

Willis, R. (1987): "What Have We Learned from the Economics of the Family?", *American Economic Review* 77(2), pp. 68-81.

Table I - Summary statistics

	Women 21-35, two or more children*							
	Argentina 1991				Mexico 2000			
	Complete Sample	Married	AE Sample	AE Sample, married	Complete Sample	Married	AE Sample	AE Sample, married
Worked for pay	0.314	0.305	0.315	0.305	0.234	0.216	0.239	0.220
(=1 if worked for pay, 0 otherwise)	(0.464)	(0.460)	(0.465)	(0.460)	(0.423)	(0.412)	(0.426)	(0.414)
More than 2 children	0.548	0.528	0.596	0.574	0.551	0.552	0.592	0.593
(=1 if mother had more than two children, 0 otherwise)	(0.498)	(0.499)	(0.491)	(0.495)	(0.497)	(0.497)	(0.491)	(0.491)
Number of children	2.976	2.905	3.062	2.985	2.957	2.963	3.029	3.035
	(1.223)	(1.165)	(1.240)	(1.183)	(1.175)	(1.184)	(1.188)	(1.197)
Same Sex	0.507	0.506	0.506	0.505	0.503	0.503	0.503	0.503
(=1 if first two children were the same sex, 0 otherwise)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Two boys	0.260	0.261	0.260	0.261	0.261	0.261	0.261	0.261
(=1 if two children were boys, 0 otherwise)	(0.438)	(0.439)	(0.438)	(0.439)	(0.439)	(0.439)	(0.439)	(0.439)
Two Girls	0.246	0.245	0.246	0.244	0.243	0.242	0.243	0.242
(=1 if two children were girls, 0 otherwise)	(0.431)	(0.430)	(0.431)	(0.430)	(0.429)	(0.428)	(0.429)	(0.428)
Boy 1st	0.508	0.510	0.508	0.510	0.512	0.512	0.512	0.512
(=1 if first child was a boy, 0 otherwise)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Boy 2nd	0.506	0.507	0.506	0.507	0.507	0.507	0.507	0.507
(=1 if second child was a boy, 0 otherwise)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Age	29.410	29.679	29.660	29.928	29.210	29.426	29.440	29.651
	(3.868)	(3.755)	(3.770)	(3.652)	(3.843)	(3.770)	(3.758)	(3.683)
Age at first birth	20.830	21.127	20.641	20.932	20.082	20.253	19.930	20.095
	(3.434)	(3.436)	(3.337)	(3.340)	(3.166)	(3.188)	(3.083)	(3.101)
Observations	653,213	497,194	599,941	456,437	493,764	382,428	458,849	355,730

Note: Means and standard deviations (in parenthesis). The samples correspond to the extended questionnaire sample of the 1991 Census, Argentina and the 2000 Census, Mexico. Samples as described in the data appendix.

Table II - Wald estimates

Argentina

	Complete Sample				Married				AE Sample				AE Sample, married			
	Proportion of sample	Worked for pay	Number of children	More than two children	Proportion of sample	Worked for pay	Number of children	More than two children	Proportion of sample	Worked for pay	Number of children	More than two children	Proportion of sample	Worked for pay	Number of children	More than two children
Total		0.3138	2.9763	0.5480		0.3050	2.9050	0.5280		0.3155	3.0619	0.5955		0.3046	2.9848	0.5741
Same sex (1)	0.5066	0.3124	3.0047	0.5637	0.5056	0.3034	2.9365	0.5460	0.5062	0.3139	3.0935	0.6131	0.5053	0.3025	3.0196	0.5940
Mixed sex (2)	0.4934	0.3153	2.9472	0.5318	0.4944	0.3067	2.8729	0.5096	0.4938	0.3171	3.0295	0.5775	0.4947	0.3066	2.9492	0.5538
Difference (1)-(2)		-0.0029	0.0574	0.0320		-0.0034	0.0636	0.0364		-0.0032	0.0640	0.0356		-0.0041	0.0704	0.0403
Wald estimate			-0.0511	-0.0917			-0.0529	-0.0924			-0.0503	-0.0905			-0.0584	-0.1021
Standard error			[0.0199]**	[0.0358]**			[0.0205]***	[0.0357]***			[0.0187]***	[0.0336]***			[0.0193]***	[0.0337]***
OLS estimate			-0.0352	-0.0783			-0.0392	-0.0800			-0.0383	-0.0913			-0.0410	-0.0875
Standard error			[0.0005]***	[0.0011]***			[0.0006]***	[0.0013]***			[0.0005]***	[0.0012]***			[0.0006]***	[0.0014]***
Observations	653,213				497,194				599,941				456,437			

Mexico

	Complete Sample				Married				AE Sample				AE Sample, married			
	Proportion of sample	Worked for pay	Number of children	More than two children	Proportion of sample	Worked for pay	Number of children	More than two children	Proportion of sample	Worked for pay	Number of children	More than two children	Proportion of sample	Worked for pay	Number of children	More than two children
Total		0.2341	2.9570	0.5506		0.3050	2.9050	0.5280		0.3155	3.0619	0.5955		0.3046	2.9848	0.5741
Same sex (1)	0.5035	0.2330	2.9854	0.5656	0.5030	0.2145	2.9944	0.5688	0.5034	0.2378	3.0598	0.6080	0.5030	0.2180	3.0684	0.6108
Mixed sex (2)	0.4965	0.2352	2.9282	0.5354	0.4970	0.2176	2.9317	0.5351	0.4966	0.2400	2.9981	0.5754	0.4970	0.2213	3.0010	0.5747
Difference (1)-(2)		-0.0022	0.0572	0.0302		-0.0031	0.0627	0.0336		-0.0022	0.0617	0.0326		-0.0033	0.0674	0.0362
Wald estimate			-0.0384	-0.0726			-0.0494	-0.0921			-0.0357	-0.0677			-0.0490	-0.0914
Standard error			[0.0199]**	[0.0399]*			[0.0205]***	[0.0396]**			[0.0187]***	[0.0385]*			[0.0193]***	[0.0383]**
OLS estimate			-0.0228	-0.0448			-0.0241	-0.0486			-0.0274	-0.0609			-0.0278	-0.0620
Standard error			[0.0211]*	[0.0012]***			[0.0212]**	[0.0013]***			[0.0203]*	[0.0013]***			[0.0206]**	[0.0014]***
Observations	493,764				382,428				458,849				355,730			

Note: * significant at 10%; ** significant at 5%; *** significant at 1%. The samples correspond to the extended questionnaire sample of the 1991 Census, Argentina and the 2000 Census, Mexico. The "AE samples" exclude women whose second child is less than a year old. Samples as described in the data appendix.

Table III - First stages, almost saturated model

Dependent variable: More than two children	Argentina				Mexico			
	Complete Sample	Married	AE Sample	AE Sample, married	Complete Sample	Married	AE Sample	AE Sample, married
Coefficient of:								
Same Sex[1]	0.0335	0.0380	0.0366	0.0413	0.0312	0.0346	0.0336	0.0371
	[0.0011]***	[0.0013]***	[0.0012]***	[0.0014]***	[0.0013]***	[0.0014]***	[0.0013]***	[0.0015]***
Two Boys[2]	0.0236	0.0277	0.0260	0.0300	0.0226	0.0261	0.0247	0.0284
	[0.0016]***	[0.0018]***	[0.0017]***	[0.0019]***	[0.0018]***	[0.0020]***	[0.0019]***	[0.0021]***
Two Girls[2]	0.0438	0.0486	0.0475	0.0529	0.0403	0.0435	0.0429	0.0461
	[0.0016]***	[0.0018]***	[0.0017]***	[0.0019]***	[0.0018]***	[0.0020]***	[0.0019]***	[0.0021]***
With spouse controls								
Same Sex[1]		0.0330		0.0412		0.0309		0.0368
		[0.0012]***		[0.0014]***		[0.0013]***		[0.0015]***
Two Boys[2]		0.0236		0.0303		0.0215		0.0274
		[0.0016]***		[0.0020]***		[0.0018]***		[0.0022]***
Two Girls[2]		0.0427		0.0525		0.0407		0.0467
		[0.0017]***		[0.0020]***		[0.0019]***		[0.0022]***
Observations	653,213	497,194	599,941	456,437	493,764	382,428	458,849	355,730
Geographic Controls	None				None			

Note: Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. [1]Control for sex of first and second children. [2]Control for sex of first child. All regressions include main effects and interactions for five categories of age, five categories of age at first birth, and sex of the first children (49 indicator variables in total). Samples as described in the data appendix.

Table IV - Two Stages Least Squares results, almost saturated model

Dependent variable: Worked for pay	Argentina				Mexico			
Instrumented: More than two children	Complete Sample	Married	AE Sample	AE Sample, married	Complete Sample	Married	AE Sample	AE Sample, married
OLS[1]	-0.0917	-0.0795	-0.0969	-0.0828	-0.0838	-0.0757	-0.0903	-0.0812
	[0.0012]***	[0.0014]***	[0.0013]***	[0.0015]***	[0.0013]***	[0.0015]***	[0.0014]***	[0.0015]***
IV: Same Sex[1]	-0.0790	-0.0861	-0.0817	-0.0958	-0.0686	-0.0886	-0.0631	-0.0862
	[0.0336]**	[0.0338]**	[0.0323]**	[0.0325]***	[0.0381]*	[0.0380]**	[0.0370]*	[0.0370]**
DWH p-value	0.7040	0.8446	0.6361	0.6868	0.6907	0.7336	0.4628	0.8929
IV: Same Sex[1] - Abadie's estimator	-0.0784	-0.0863	-0.0814	-0.0953	-0.0695	-0.1098	-0.0631	-0.0862
	[0.0679]**	[0.0356]***	[0.0333]***	[0.0378]***	[0.0451]*	[0.0499]**	[0.03962]*	[0.0415]**
IV: Two Boys and Two Girls[2]	-0.0650	-0.0748	-0.0652	-0.0821	-0.0469	-0.0739	-0.0445	-0.0721
	[0.0322]**	[0.0326]**	[0.0310]**	[0.0313]***	[0.0367]	[0.0369]**	[0.0357]	[0.0360]**
Sargan p-value	0.1515	0.2055	0.0701	0.1121	0.0368	0.1145	0.0545	0.1015
DWH p-value	0.4069	0.8853	0.3050	0.9825	0.3142	0.9619	0.1994	0.8006
Spouse worked for pay								
OLS[1]		-0.0125		-0.0105		-0.0324		-0.0330
		[0.0006]***		[0.0006]***		[0.0010]***		[0.0011]***
IV: Same Sex[1]		0.0071		-0.0031		0.0014		-0.0036
		[0.0153]		[0.0137]		[0.0288]		[0.0280]
DWH p-value		0.1995		0.5904		0.2387		0.2920
IV: Two Boys and Two Girls[2]		0.0063		-0.0026		0.0062		-0.0016
		[0.0148]		[0.0132]		[0.0276]		[0.0271]
Sargan p-value		0.8481		0.8800		0.5681		0.7753
DWH p-value		0.2016		0.5497		0.1608		0.2452
Observations	653,213	497,194	599,941	456,437	493,764	382,428	458,849	355,730
Geographic Controls				None				

Note: Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. [1]Control for sex of first and second children. [2]Control for sex of first child. OLS results are virtually identical with either set of controls. All regressions include main effects and interactions for five categories of age, five categories of age at first birth, and sex of the first children (49 indicator variables in total). Standard errors for Abadie's estimator were obtained by 500 bootstrap replications. Samples as described in the data appendix.

Table V - Two Stages Least Squares results, geographic controls

Dependent variable: Worked for pay	Argentina				Mexico			
Instrumented: More than two children	Complete Sample	Married	AE Sample	AE Sample, married	Complete Sample	Married	AE Sample	AE Sample, married
No geographic controls								
IV: Same Sex[1]	-0.0737	-0.0828	-0.0775	-0.0931	-0.0660	-0.0847	-0.0610	-0.0826
	[0.0331]**	[0.0332]**	[0.0319]**	[0.0320]***	[0.0378]*	[0.0374]**	[0.0368]*	[0.0366]**
IV: Two Boys and Two Girls[2]	-0.0610	-0.0722	-0.0621	-0.0801	-0.0449	-0.0707	-0.0433	-0.0696
	[0.0320]*	[0.0322]**	[0.0309]**	[0.0311]***	[0.0365]	[0.0364]*	[0.0357]	[0.0357]*
Spouse worked for pay								
IV: Same Sex[1]		0.0054		-0.0031		0.0004		-0.0047
		[0.0148]		[0.0134]		[0.0284]		[0.0277]
IV: Two Boys and Two Girls[2]		0.0046		-0.0026		0.0049		-0.0027
		[0.0143]		[0.0130]		[0.0275]		[0.0270]
Geographic controls								
IV: Same Sex[1]	-0.0786	-0.0807	-0.0841	-0.0940	-0.0642	-0.0828	-0.0593	-0.0805
	[0.0328]**	[0.0329]**	[0.0316]***	[0.0317]***	[0.0377]*	[0.0374]**	[0.0368]	[0.0366]**
IV: Two Boys and Two Girls[2]	-0.0666	-0.0709	-0.0701	-0.0821	-0.0433	-0.0688	-0.0416	-0.0676
	[0.0318]**	[0.0320]**	[0.0306]**	[0.0308]***	[0.0366]	[0.0365]*	[0.0358]	[0.0358]*
Spouse worked for pay								
IV: Same Sex[1]		0.0047		-0.0039		-0.0014		-0.0080
		[0.0148]		[0.0133]		[0.0282]		[0.0274]
IV: Two Boys and Two Girls[2]		0.0050		-0.0019		0.0050		-0.0019
		[0.0142]		[0.0129]		[0.0142]		[0.0129]
Geographic Controls	522 Municipalities				32 States			
Observations	653,213	497,194	599,941	456,437	493,764	382,428	458,849	355,730

Note: Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. [1]Control for sex of first and second children. [2]Control for sex of first child. All regressions include the age and the age at first birth of the woman or her spouse, in addition to the sex of the first and/or second children. Samples as described in the data appendix.

Table VI - Two Stages Least Squares results for income, Mexico

Instrumented: More than two children	Complete Sample	Married	AE Sample	AE Sample, married
Dependent variable:	**Log of Woman's Labor Income**			
OLS	-0.7901	-0.7127	-0.8284	-0.7451
	[0.0097]***	[0.0108]***	[0.0100]***	[0.0111]***
IV: Same Sex	-0.4705	-0.5308	-0.4128	-0.4978
	[0.2703]*	[0.2674]**	[0.2628]	[0.2604]*
DWH p-value	0.2364	0.4960	0.1129	0.3415
Observations	469,542	364,497	436,076	338,823
Dependent variable:	**Log of Total Household Income**			
OLS	-0.5462	-0.5679	-0.5451	-0.5640
	[0.0078]***	[0.0088]***	[0.0079]***	[0.0089]***
IV: Same Sex	-0.0647	-0.1039	-0.0483	-0.0519
	[0.2171]	[0.2196]	[0.2091]	[0.2126]
DWH p-value	0.0258	0.0338	0.0169	0.0154
Observations	470,524	364,614	437,111	339,049
Dependent variable:	**Log of Non Woman Income**			
OLS	-0.5501	-0.7157	-0.5311	-0.6991
	[0.0099]***	[0.0107]***	[0.0101]***	[0.0109]***
IV: Same Sex	-0.1237	-0.0777	-0.1597	-0.0715
	[0.2762]	[0.2664]	[0.2672]	[0.2586]
DWH p-value	0.1216	0.0160	0.1636	0.0146
Observations	470,498	364,596	437,088	339,033
Geographic Controls	None			

Note: Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions include the age and the age at first birth of the woman or her spouse, in addition to the sex of the first and second children. Samples as described in the data appendix.

Table VII - Heterogeneity - Results by education level

Dependent variable: Worked for pay	Argentina				Mexico			
Instrumented: More than two children	Complete Sample	Married	AE Sample	AE Sample, married	Complete Sample	Married	AE Sample	AE Sample, married
Education level:	**Completed primary or less**				**Less than primary**			
OLS	-0.0767	-0.0699	-0.0828	-0.0749	-0.0448	-0.0303	-0.0493	-0.0339
	[0.0015]***	[0.0018]***	[0.0016]***	[0.0018]***	[0.0025]***	[0.0029]***	[0.0026]***	[0.0030]***
IV: Same Sex	-0.0629	-0.1028	-0.0613	-0.1004	-0.2018	-0.2147	-0.2256	-0.2561
	[0.0416]	[0.0399]**	[0.0407]	[0.0393]**	[0.1059]*	[0.1094]**	[0.1058]**	[0.1116]**
DWH p-value	0.7403	0.4094	0.5967	0.5150	0.1329	0.0853	0.0902	0.0407
Observations	379,290	264,142	355,560	248,477	143,032	102,182	136,882	98,265
Education level:	**High school, complete or incomplete**				**Primary education**			
OLS	-0.0923	-0.0786	-0.0946	-0.0803	-0.0543	-0.0418	-0.0588	-0.0460
	[0.0023]***	[0.0025]***	[0.0023]***	[0.0025]***	[0.0022]***	[0.0024]***	[0.0023]***	[0.0025]***
IV: Same Sex	-0.1063	-0.1107	-0.1100	-0.1219	0.0546	-0.0382	0.0629	-0.0253
	[0.0573]*	[0.0588]*	[0.0551]**	[0.0567]**	[0.0596]	[0.0583]	[0.0581]	[0.0567]
DWH p-value	0.8062	0.5847	0.7803	0.4620	0.0655	0.9503	0.0345	0.7156
Observations	194,683	161,631	176,000	146,492	154,302	118,349	144,552	111,260
Education level:	**Some university or tertiary**				**High school, univerity or technical degree**			
OLS	-0.0698	-0.0629	-0.0707	-0.0633	-0.0882	-0.0778	-0.0937	-0.0823
	[0.0037]***	[0.0039]***	[0.0038]***	[0.0040]***	[0.0023]***	[0.0025]***	[0.0024]***	[0.0026]***
IV: Same Sex	0.0280	0.0700	-0.0386	-0.0100	-0.1048	-0.0862	-0.0888	-0.0751
	[0.0945]	[0.0928]	[0.0875]	[0.0852]	[0.0518]**	[0.0513]*	[0.0502]*	[0.0498]
DWH p-value	0.2985	0.1485	0.7141	0.5302	0.7477	0.8695	0.9216	0.8858
Observations	79,240	71,421	68,381	61,468	196,430	161,897	177,415	146,205
Geographic Controls	None							

Note: Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. All regressions include the age and the age at first birth of the woman or her spouse, in addition to the sex of the first and second children. Samples as described in the data appendix.

Table VIII - Robustness checks

Dependent variable: Worked for pay	Argentina				Mexico			
Instrumented: More than two children	Complete Sample	Married	AE Sample	AE Sample, married	Complete Sample	Married	AE Sample	AE Sample, married
Using labor force status instead of worked for pay as dep. variable								
OLS[1]	-0.0956	-0.0811	-0.1007	-0.0847	-0.0907	-0.0808	-0.0958	-0.0853
	[0.0013]***	[0.0015]***	[0.0013]***	[0.0015]***	[0.0014]***	[0.0015]***	[0.0014]***	[0.0016]***
IV: Samesex[1]	-0.0748	-0.0820	-0.0809	-0.0948	-0.0567	-0.0816	-0.0530	-0.0790
	[0.0338]**	[0.0338]**	[0.0326]**	[0.0327]***	[0.0381]	[0.0378]**	[0.0371]	[0.0369]**
IV: Two Boys and Two Girls[2]	-0.0596	-0.0693	-0.0635	-0.0802	-0.0366	-0.0682	-0.0356	-0.0659
	[0.0326]*	[0.0329]**	[0.0315]**	[0.0317]**	[0.0368]	[0.0368]*	[0.0360]	[0.0360]*
Observations	653,213	497,194	599,941	456,437	493,764	382,428	458,849	355,730
Women with mismatching children born and children in household								
OLS[1]	-0.0966	-0.0816	-0.1015	-0.0848	-0.0885	-0.0776	-0.0942	-0.0825
	[0.0012]***	[0.0014]***	[0.0012]***	[0.0014]***	[0.0013]***	[0.0015]***	[0.0013]***	[0.0015]***
IV: Samesex[1]	-0.0698	-0.0827	-0.0732	-0.0920	-0.0512	-0.0985	-0.0463	-0.0962
	[0.0341]**	[0.0335]**	[0.0328]**	[0.0323]***	[0.0383]	[0.0380]***	[0.0372]	[0.0371]***
IV: Two Boys and Two Girls[2]	-0.0554	-0.0745	-0.0576	-0.0824	-0.0225	-0.0784	-0.0219	-0.0779
	[0.0328]*	[0.0324]**	[0.0316]*	[0.0313]***	[0.0368]	[0.0368]**	[0.0359]	[0.0361]**
Observations	748,157	541,529	690,103	499,000	536,719	408,384	500,226	380,996
Geographic Controls	None							

Note: Standard errors in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. [1]Control for sex of first and second children. [2]Control for sex of first child. OLS results are virtually identical with either set of controls. All regressions include the age and the age at first birth of the woman or her spouse, in addition to the sex of the first and/or second children. Samples as described in the data appendix.