

CAF - WORKING PAPER #2026/03

First version: March 2, 2026

When Does Family Size Matter for Children's Education? Causal Evidence from Latin America

Lucila Berniell¹ | Cristian Bonavida² | Dolores de la Mata³

¹Research affiliate at CEDLAS (UNLP) and CEPE (UTDT).

luberniell@gmail.com

²Carnegie Mellon University and Research affiliate at CEDLAS (UNLP).

cristianbonavida@gmail.com

³CAF-Development Bank of Latin America and the Caribbean, Research Department. mdelamata@caf.com

We examine the relationship between family size and children's human capital—commonly known as the quantity–quality trade-off—in Latin America. Despite historically high fertility rates and a rapid convergence toward levels observed in high-income countries, evidence on this trade-off in the region remains limited. Using census data from five countries and a twin instrumental variable strategy, we estimate the causal effect of family size on educational outcomes. We document a trade-off in households with three or more children: an additional sibling increases the likelihood of falling behind in school, reduces primary school completion, and lowers total years of schooling. The magnitude of the trade-off is related to tighter financial constraints, more traditional gender norms, and insufficient local school supply. Identifying the contexts in which these trade-offs arise is essential for understanding how fertility patterns shape gaps in human capital accumulation and for informing policies aimed at narrowing intergenerational disparities.

KEYWORDS

human capital investments, fertility, Latin America, twins-instrument

Small sections of text that are less than two paragraphs may be quoted without explicit permission as long as this document is acknowledged. Findings, interpretations and conclusions expressed in this publication are the sole responsibility of its author(s) and cannot be, in any way, attributed to CAF, its Executive Directors or the countries they represent. CAF does not guarantee the accuracy of the data included in this publication and is not, in any way, responsible for any consequences resulting from its use.

CAF - DOCUMENTO DE TRABAJO #2026/03

Esta versión: 2 de marzo de 2026

¿Cuándo importa el tamaño de la familia para la educación de los niños? Evidencia causal para América Latina?

Lucila Berniell¹ | Cristian Bonavida² | Dolores de la Mata³

¹Research affiliate at CEDLAS (UNLP) and CEPE (UTDT).
luberniell@gmail.com

²Carnegie Mellon University and Research affiliate at CEDLAS (UNLP).
cristianbonavida@gmail.com

³CAF-Development Bank of Latin America and the Caribbean, Research Department. mdelamata@caf.com

Examinamos la relación entre el tamaño de la familia y el capital humano de los niños—comúnmente conocida como el *trade-off* cantidad-calidad—en América Latina. A pesar de tasas de fecundidad históricamente elevadas en la región y de una rápida convergencia hacia los niveles observados en los países de altos ingresos, la evidencia sobre este *trade-off* sigue siendo limitada. Utilizando datos censales de cinco países y una estrategia de variable instrumental basada en gemelos, estimamos el efecto causal del tamaño de la familia sobre los resultados educativos. Encontramos un *trade-off* en hogares con tres o más hijos: un hermano adicional aumenta la probabilidad de rezago escolar, reduce la probabilidad de completar la educación primaria y disminuye los años totales de escolaridad. La magnitud del *trade-off* está asociada con mayores restricciones financieras, normas de género más tradicionales y una oferta educativa local insuficiente. Identificar los contextos en los que surgen estos *trade-offs* es fundamental para comprender cómo los patrones de fecundidad configuran las brechas en la acumulación de capital humano y para informar políticas orientadas a reducir las disparidades intergeneracionales.

KEYWORDS

inversiones en capital humano, fertilidad, América Latina, instrumento de mellizos

Pequeñas secciones del texto, menores a dos párrafos, pueden ser citadas sin autorización explícita siempre que se cite el presente documento. Los resultados, interpretaciones y conclusiones expresados en esta publicación son de exclusiva responsabilidad de su(s) autor(es), y de ninguna manera pueden ser atribuidos a CAF, a los miembros de su Directorio Ejecutivo o a los países que ellos representan. CAF no garantiza la exactitud de los datos incluidos en esta publicación y no se hace responsable en ningún aspecto de las consecuencias que resulten de su utilización.

1 | INTRODUCTION

The extent to which family size shapes the human capital of each child—the so-called quantity–quality trade-off—remains a central and contested question in demographic and economic research. This relationship has important implications in settings where fertility choices and educational investments jointly determine long-run development prospects and the intergenerational transmission of inequality. Correlational evidence has consistently shown that children in larger families tend to achieve worse educational outcomes. Seminal theoretical contributions by [Becker and Lewis \(1973\)](#) and [Becker and Lewis \(1973\)](#) formalized this association within a framework in which fertility and parental investments are jointly chosen under binding resource constraints, implying a trade-off between the number of children and the quality of investments per child. Yet empirical work has questioned whether family size exerts a causal effect on children’s human capital formation.

This paper estimates the causal relationship between family size and children’s educational outcomes in five Latin American countries using population census data. We draw on harmonized 10% IPUMS public-use microdata samples from Colombia (2005), the Dominican Republic (2002, 2010), Ecuador (2010), El Salvador (1992, 2007), and Nicaragua (1995, 2005)—countries in which twin births can be identified with precision. Our identification strategy exploits twin births as an exogenous source of variation in fertility. Specifically, we estimate the effects of family size on the educational outcomes of first-born children in families with at least two children, and on first- and second-born children in families with at least three children. Twin births at the second and third parity serve as instruments for family size, respectively. Following [Bhalotra and Clarke \(2020\)](#), we control for observable characteristics associated with the probability of twinning and estimate bounds for the instrumental-variable (IV) parameters under imperfect-instrument assumptions using the approaches of [Conley et al. \(2012\)](#) and [Nevo and Rosen \(2012\)](#), thereby providing additional robustness to concerns about selection into twinning.

Beyond average effects, we explore heterogeneity along dimensions that speak directly to the mechanisms underlying the trade-off. We examine variation by parental socioeconomic status, child gender, and local educational supply. This allows us to assess whether financial constraints, gender norms in parental investments, and supply-side limitations in schooling mediate the impact of fertility shocks on children’s outcomes. We also relax the standard linearity assumption and study nonlinear responses to changes in family size.

Our analysis makes three main contributions. First, we add to the longstanding empirical debate on the quantity–quality trade-off, whose findings remain far from conclusive. In high-income countries, most studies report null to negative family size effects on educational attainment ([Black et al., 2005, 2010](#); [Cáceres-Delpiano, 2006](#); [Angrist et al., 2010](#); [Åslund and Grönqvist, 2010](#); [Bhalotra and Clarke, 2020](#)). In developing contexts, the evidence is markedly more mixed. Early work by [Rosenzweig and Wolpin \(1980\)](#) finds that larger family size reduces schooling among farm households in India, while subsequent studies report effects ranging from null or even positive—such as [Marteleto and de Souza \(2012\)](#) for Brazil, [Alidou and Verpoorten \(2019\)](#) for Sub-Saharan Africa—to clearly negative, as in [Li et al. \(2008\)](#) and [Rosenzweig and Zhang \(2009\)](#) for China, [Ponczek and Souza \(2012\)](#) for Brazil, and [Bhalotra and Clarke \(2020\)](#) for a large pool of developing countries.¹ This heterogeneity underscores the importance of institutional context, the stage of the fertility transition, and local constraints in shaping the nature of the trade-off, as well as the types of

¹All these papers use the same identification strategy as our study, relying on the twin instrument. Beyond twins, the literature has also exploited other sources of plausibly exogenous variation in fertility, such as children’s sex composition and changes in access to contraception or abortion. A comprehensive review of this literature is provided by [Guo et al. \(2022\)](#).

outcomes analyzed and the data sources used.

Second, we provide new evidence for Latin America and the Caribbean (LAC), a region for which causal evidence remains scarce and mostly confined to a few country case studies, with the exception of [Li et al. \(2017\)](#).² LAC offers a particularly informative setting: while fertility remained persistently high for much of the twentieth century, recent decades have seen a rapid convergence toward the levels observed in high-income countries, alongside enduring socioeconomic gradients in both fertility and educational attainment. If a quantity–quality trade-off operates in this context, the rapid decline in family size among households of low socioeconomic status ([Onofri et al., 2026](#)) may contribute to narrowing the large human capital gaps that characterize the region ([de la Mata et al., 2022](#)). By assembling a harmonized, multi-country census database, we provide a systematic, region-wide assessment of this channel. Moreover, we extend the range of outcomes beyond standard measures of educational attainment. Specifically, we examine whether children fall behind the grade appropriate for their age (overage)—a pervasive problem in LAC that often precedes school dropout—as well as primary school completion and total years of schooling. This approach allows us to assess whether the trade-off operates while children are still in school—through slower grade progression and reduced academic persistence—or whether it mainly reflects more discrete schooling decisions, such as school enrollment.

Third, the paper addresses methodological concerns related to the twin-based identification strategy that have not been fully considered in much of the previous literature, specially for developing countries. A growing body of work has raised questions about the validity of twins as instruments. [Bhalotra and Clarke \(2019\)](#) show that mothers who bear twins tend to be positively selected on health and education, potentially violating the exclusion restriction. Moreover, twin births may directly affect parental investments, as twins are more likely to be born with adverse health endowments, which can induce either reinforcing or compensatory parental responses ([Rosenzweig and Zhang, 2009](#); [Conley, 2008](#); [Hsin, 2012](#); [Bhalotra and Clarke, 2020](#)). We address these identification concerns by controlling for covariates associated with the likelihood of twinning and by estimating IV bounds that allow for limited violations of instrument exogeneity. Separately, we improve the measurement of the instrument by restricting the sample to cases in which twin identification is exact, based on precise dates of birth rather than year-of-birth approximations. This refinement materially affects the magnitude of the estimated trade-off.

Our IV estimates show that, in families with at least three children, larger family size reduces the probability that first- and second-born children remain on track in their educational progression, as well as their primary completion and years of schooling. We find no effects for first-born children in families with at least two children. Birth-order analyses indicate that second-born children are the most affected in families with three or more children, suggesting that this sibling bears the most persistent consequences of the fertility shock. The robustness analysis highlights the importance of accounting for potential selection into twinning, and bound estimates that treat the twin IV as an imperfect instrument support our main results. Additionally, our findings indicate that identifying twins solely on the basis of children's ages may lead to an overestimation of the quantity–quality trade-off.

We find heterogeneous effects by family and child characteristics. The negative effect of family size is concentrated among children whose parents have low levels of education, while effects are null among those with highly educated parents, aligning with the hypothe-

²Causal evidence for LAC shows mixed results. Beyond [Marteleto and de Souza \(2012\)](#) and [Ponczek and Souza \(2012\)](#) for Brazil (mixed findings), studies using a sex-composition instrument include [Baez \(2008\)](#) for Colombia (negative effects), [Fitzsimons and Malde \(2014\)](#) for Mexico (no effects on girls), and [Li et al. \(2017\)](#) for 12 LAC countries (negative on the probability of secondary education attendance but smaller effects than in Asia).

sis that the quantity–quality trade-off is more pronounced in families facing tighter financial constraints. Gender-specific estimates point to an important role for social norms. Female children appear less adversely affected—and in some cases positively affected—by fertility shocks, whereas male children experience increased probabilities of labor force participation and school disengagement in larger families. This pattern is consistent with households re-allocating male labor toward income generation under tighter financial constraints, thereby amplifying the trade-off for boys.

Exploiting the geographic richness of the census data, we examine heterogeneous effects by the level of educational supply in the district of residence, proxied by the ratio of teachers to school-age children. Negative effects of family size are concentrated in districts with intermediate levels of educational supply, consistent with an inverted U-shaped pattern: in low-supply districts, both small and large families face binding access constraints, whereas in high-supply districts educational provision mitigates the impact of additional children. Taken together, these results suggest that the quantity–quality trade-off in Latin America is most pronounced in institutional settings where access to education is neither severely constrained nor fully abundant.

Finally, nonlinear estimates indicate that the impact of an additional child is not constant across the parity distribution. Marginal effects differ by birth order and family size, highlighting the importance of accounting for heterogeneous effects of fertility shocks at different parity levels.

The remainder of the paper is organized as follows. Section 2 describes the data, sample construction, and fertility trends motivating the country selection. Section 3 presents the identification strategy and the twin-based IV framework. Section 4 reports the main estimates and the IV bounds under imperfect instrument assumptions. Section 5 examines heterogeneous effects and discusses potential mechanisms. Section 6 relaxes linearity and explores nonlinear responses. Section 7 concludes.

2 | BACKGROUND AND DATA

2.1 | Fertility rates in the selected LAC countries

Latin American and Caribbean countries are experiencing a profound demographic transition, characterized by a rapid and sustained decline in fertility. While fertility rates were close to six children per woman in the 1960s—almost twice as high as in high-income countries—they fell sharply to around two children per woman by the 2010s, according to United Nations estimates (United Nations, Department of Economic and Social Affairs, Population Division, 2024), nearly closing the gap with high-income economies (see Figure A.1a in the Appendix). This downward trend has been particularly steep in Colombia, the Dominican Republic, Ecuador, El Salvador, and Nicaragua, the countries analyzed in this study. Notably, despite these large declines, all countries except Colombia still exhibit fertility levels above the regional average in recent decades (see Figure A.1a and Table A.1 in the Appendix). Family sizes are much larger for the less educated women, although the declining fertility occurs across all education levels (see Figure A.2 in the Appendix).

2.2 | Data and sample selection

We use data from IPUMS International (IPUMS-I) datasets, which consist of 10% public use microdata samples from population censuses conducted in various countries and harmonized over years (Ruggles et al., 2025).³ Specifically, our analysis draws on census

³Integrated Public Use Microdata Series - International (<http://international.ipums.org/international/>).

data from Colombia (2005), Ecuador (2010), El Salvador (1992, 2007), Nicaragua (1995, 2005) and the Dominican Republic (2002, 2010). Unfortunately, we are unable to include all Latin American countries with available harmonized censuses in IPUMS-I. Our empirical strategy requires the availability of key variables to select the sample, identify twin births, measure outcome variables of interest in the IV specifications, and control for confounding variables.

From the selected censuses, we construct our analytical sample by selecting children aged 6 to 18 years who were living with their biological mother at the time of the census. We focus on first- and second-born children of mothers aged 18 to 50 years, with at least two children ever born. The father may or may not be present, but the presence of the biological mother is required to identify the total number of siblings of the same mother living in the household. Because censuses usually do not directly show the links between children and their biological parents, we use the IPUMS-I “Constructed Family Interrelationship Variables.” This method identifies the biological mothers of children through a well-established probabilistic method (Ruggles et al., 2015). Additionally, we restrict our sample to cases in which the total number of children ever born alive to the mother matches the number of children co-residing with her. This condition ensures that we observe the outcomes of all children, as we do not have information on non-resident children. It also avoids the potential misclassification of the birth order when some children are not present in the home. After applying these criteria, we identify 793,586 children of 618,581 mothers.⁴

We classify as twins the individuals who meet either of the following conditions: (i) same year and same month of birth, or (ii) same age and same month of birth. This results in the identification of 4,458 twins. A key advantage of using census data from these countries is the large sample size, which is crucial for our identification strategy, given the rarity of twin births. We do not allow discrepancies in the month of birth, even when age and birth year coincide. Allowing such discrepancies would risk including siblings who were born very close in time but are not twins.

As fertility rates were high in the countries and periods we analyze, our sample selection does not imply leaving out a large share of families. Among all families where the mother is between 18 and 50 years old at the census date, those with two or more children represent 75.2% of all families. Then, restricting further to the cases in which all children co-reside with the mother, we are left with 60.1% of the original sample (see Figure A.3a in the Appendix), and that sample selection is stable across the countries considered (see Figure A.3b in the Appendix).

Considering all mothers, the co-residence rate is high and stable around 75% for women between 20 and 35 years old, and then, as expected, starts to decrease (see Figure A.4a in the Appendix) as older mothers also have older children, who are more likely to leave the home.⁵ For mothers aged 18 to 32, the average number of children among those co-residing with all of them closely mirrors the average for all mothers in that age group. By contrast, among women older than 32, those who co-reside with all their children tend to have fewer children than their counterparts in the overall population.⁶ To the estimation purpose, this means that the co-residence requisite leave us with a sub-sample of smaller families, so if

⁴To arrive at the final sample, we exclude the following cases: (i) children for whom the difference between their reported year of birth and the census year minus their age is greater than 1 or less than -1; (ii) children of mothers who reported more than 12 children ever born; (iii) children of mothers who reported more than 5 child deaths; and (iv) children living in households where the mother and father lack information about their years of education.

⁵Co-residence rates are relatively lower for women below 20 years old, but especially for those who have three or more children.

⁶Figure A.4b also shows that the average number of children increases with the age of the mother. This pattern partially reflects the declining fertility rate in the region, but also the fact that younger mothers are further from reaching the end of their reproductive life (i.e., the number of children is a truncated variable).

we were to find a relevant trade-off in the estimation sample, the same restriction should operate for bigger families.

2.3 | Estimation samples

Following the literature of the twins IV approach (Black et al., 2005; Angrist et al., 2010), we split the full sample into two subsamples. A first subsample, which we call “2+ sample”, focuses on the first non-twin born child in families with 2 or more births. A second subsample, which we call “3+ sample”, focuses on the first and second non-twin born children in families with 3 or more births. Table 1 describes these samples.⁷ Restricting the sample to siblings prior to the twin’s arrival has the purpose of avoiding selection. Considering siblings in higher orders could be problematic if families who choose to have another child after a twin birth were different from those who choose to have another child after a singleton birth.

As human capital outcome variables of children, we consider age-standardized years of schooling and a dummy variable that indicates whether they are enrolled in formal education at the time of the census. We focus on these two measures since we restrict the analysis to children of school-age, that is, they are still in the process of acquiring education. The years of schooling are standardized (z-scores) within country-year and age of the children. Table A.9 in the Appendix shows the mean and standard deviation of years of schooling in the whole population by country, year and age of children. Additionally, we consider whether children have completed primary, but only for a subsample of those with age above the country-specific cutoffs at which children are expected to have finished this level.⁸ We choose not to present estimates for secondary completion since we are left with a small sample for children between 18-25 after conditioning on co-residence, which may also induce sample selection. We also construct a fourth outcome to measure whether the child is behind the expected completed years of schooling, considering the modal distribution for children in the same census, age and birth quarter. We define a child as being “Behind” when his or her completed years of schooling fall at least two years below the expected level.⁹

Parents’ and households’ characteristics measured with IPUMS data include the mother’s age at the time of the census and at the time of the first birth, the parents’ education, and a dummy variable that indicates whether a child ever born to the women has died. This variable results from the difference between the reported number of children ever born to each woman and the number of children born who were alive at the time of the census. Although the age and moment at which born children died are not available in the census, most child deaths typically occur during the first years of life. Since child mortality is negatively associated with maternal health (Bhalotra and Rawlings, 2011), it is expected that healthier mothers in our dataset are less likely to have experienced the loss of a child. Therefore, we use this child mortality variable as an indirect measure of maternal health.

⁷Table A.2 shows descriptive statistics only for children in families that have twins in the n th birth.

⁸These cutoffs are reported in Table A.8, which are precisely the theoretical entrance ages to secondary education in each country-year.

⁹The analysis is restricted to children in cohorts where the mode of schooling is two or more, as these cohorts display variation in the “Behind” indicator. In cohorts where the modal value equals zero, we use the second mode as the reference. Therefore, children with zero years of schooling are classified as behind in all cases.

TABLE 1 Sample descriptive statistics

	2+ Sample	3+ Sample	
	First-born	First-born	Second-born
	(1)	(2)	(3)
Age (years)	11.9	12.6	11.6
Female (%)	48.2	48.0	48.3
Birth month	6.6	6.6	6.6
Age difference with nth birth (years)	3.6	6.4	3.8
Number of children	2.9	3.6	3.7
3 children or more (%)	55.3	-	-
4 children or more (%)	21.6	39.0	42.9
Father or mother's spouse at home (%)	83.5	85.4	83.8
Father at home (%)	81.7	83.2	82.6
Twins at nth birth (%)	0.7	0.8	0.8
Mother's age at first child (years)	21.3	20.4	21.2
Mother's age at census (years)	33.3	33.1	35.0
Years of schooling (mother)	7.6	6.5	6.6
Primary complete (mother) (%)	57.2	66.3	65.7
Secondary complete (mother) (%)	32.7	26.5	26.6
University complete (mother) (%)	10.1	7.1	7.7
Mother's health (% have not an alive child)	6.5	7.6	8.3
Years of schooling (father)	7.3	6.4	6.4
Primary complete (father) (%)	59.1	66.3	65.6
Secondary complete (father) (%)	29.5	25.1	25.1
University complete (father) (%)	11.4	8.7	9.4
Primary complete (%) (*)	82.6	78.8	76.9
Secondary complete (%) (*)	40.0	33.7	33.8
Attending school (%)	87.1	83.4	86.0
Behind 2 years of schooling or more (%)	24.7	31.9	29.3
Behind 1 year of schooling (%)	19.9	19.9	21.1
Not behind (%)	55.4	48.3	49.6
Labor force participation (%)	18.8	21.6	20.7
Population of district of residence (logs)	12.0	11.9	11.9
Colombia 2005 (%)	42.9	40.1	40.2
Dominican Republic 2002 (%)	8.9	10.0	9.5
Dominican Republic 2010 (%)	9.8	9.7	10.0
Ecuador 2010 (%)	17.0	15.8	16.0
Nicaragua 1995 (%)	4.4	6.2	5.9
Nicaragua 2005 (%)	5.5	6.2	6.3
El Salvador 1992 (%)	4.7	5.6	5.1
El Salvador 2007 (%)	6.8	6.5	6.9
Observations	377,328	206,097	210,161

Notes: The table presents the mean values of selected characteristics for two subsamples. Column 1 shows the “2+ sample”, which includes the first non-twin child in families with two or more births. Columns 2 and 3 show the “3+ sample”, focusing on the first and second non-twin children, respectively, in families with three or more births. (*) The length of primary and secondary education differ by country. We use these country specific cut-offs to compute primary and secondary education completion on the population who is above this age cut-offs. *Source:* Data from IPUMS International samples.

3 | EMPIRICAL STRATEGY

We first examine the relation between education outcomes of the first-born or first- and second-born children and family size using ordinary least squares (OLS). Then, we use twinning at the second or third births to instrument the number of children in the 2+ sample and the 3+ sample, respectively.

Specifically, our OLS specification in the sample of families with n or more children is the following:

$$y_{ihjd} = \alpha + \beta_1 Nchild_{ih} + \beta_2 X_i + \beta_3 X_{ih}^M + \beta_4 X_d + \gamma_j + u_i, \quad (1)$$

where y_{ihjd} represents the outcome of interest of child i in family h , in district d and country-year j . $Nchild$ is equal to the total number of children in the family (also referred to as fertility). X_i is a set of characteristics for the child i (gender, dummies for age, age difference with the n th sibling, fixed effect for the birth month and birth order in the 3+sample), X_{ih}^M are mothers' characteristics (age at census, age at first birth, education, a proxy of her health and an interaction term between the latter), X_d are characteristics of the district of residence (population in log and square), and γ_j are country-year of census fixed effects.

The parameter of interest β_1 will be biased in the OLS estimation as potential non observed third factors correlated with fertility decisions may also affect families' investments in child human capital formation. The argument for using twins as an instrumental variable for fertility is that the birth of twins results in an increase in family size that is out of parents' control. However, a recent paper by [Bhalotra and Clarke \(2019\)](#) shows that twin births are not as random as they were presumed to be. The authors find that the distribution of twins in the population is skewed in favour of healthier women with healthier behaviors, both in developed and developing countries. They also found a positive association of twin births with the mother's education, which is coherent with the fact that education facilitates the access and the uptake of health-related information and practices.

We follow [Bhalotra and Clarke \(2020\)](#), and include a set of mothers' characteristics X_h^M , including her education and the measure of mothers' health described in the previous section. Table A.3 in the Appendix shows that the probability of giving birth to twins is positively and significantly associated with mothers' health as well as with mothers' education, conditional on maternal age, age of first birth, place of residence, and year-country fixed effects. Nevertheless, we are aware that these variables may only adjust partially for all dimensions characterizing healthier mothers, and, as a result our IV estimates may be upward biased (too close to zero) as described by [Bhalotra and Clarke \(2020\)](#). In this sense, we think of our results as lower bounds and we test this in the next section implementing an IV bound estimation.

The first stage for our instrumental variable approach in the sample of families with n or more children is as follows:

$$Nchild_{ih} = \pi_0 + \pi_1 \times Twins_{ih} + \pi_2 X_i + \pi_3 X_{ih}^M + \pi_4 X_d + \gamma_j + \epsilon_i, \quad (2)$$

where $Twins_h$ is a binary variable that takes value 1 if the n th birth in family h was a twin birth and 0 if it was a singleton, for $n=2$ (the "2+ sample") and for $n=3$ (the "3+ sample").

The first stage is shown in Table 2 in the first row for Panel A and B, estimating the effect of twin birth in the n th birth. The results demonstrate that the twin instrument is highly predictive of family sizes in both samples. The point estimates indicate that the incidence of

twins raises total fertility by about 0.65 to 0.79 births. In line with previous literature, this estimate is less than one.¹⁰

4 | RESULTS

Table 2 presents the OLS and IV estimates of family size on education. For all outcomes, OLS estimates show a strong negative correlation between family size and children's educational outcomes. After instrumenting for family size with the twin instrument, we find no significant negative impacts in the 2+ sample. The coefficients for years of school and primary are even positive, although neither is statistically significant at conventional levels. This result is in line with much of the previous literature (Black et al., 2005; Cáceres-Delpiano, 2006; Angrist et al., 2010; Marteleto and de Souza, 2012; Alidou and Verpoorten, 2019; Bhalotra and Clarke, 2020), which generally rejects the quantity-quality trade off for first-born children in the 2+ sample.

We find, however, that a negative quantity-quality relation holds true for behind two years, years of schooling and primary completion in our 3+ sample. As expected, the coefficients are much smaller than in the OLS specification. In particular, according to the IV estimates for the 3+ sample, having an additional child in the household increases the probability of falling two or more years behind the expected level of schooling by 2.8 percentage points. This effect is statistically significant at the 5% level. Also, one additional child in the household causes a reduction of about 0.036 units in the z-score. This magnitude corresponds to the equivalent of between 0.076 and 0.109 years of schooling for a child of 15 years old in the countries considered.¹¹ In addition, it leads to a reduction of 2.1 percentage points in the probability of completing primary education, equivalent to a 3% decrease relative to the sample mean of primary completion. Both effects are statistically significant at the 10% level. The probability of attending formal education at the moment of the census has a negative coefficient (-0.009), but is not affected by family size.

Our results suggest that in larger families (with three or more children), the resources available for each child become more limited, giving rise to a quantity-quality trade-off when a new sibling is born. The increased competition or tension in the allocation of both material resources and parental time and attention within the household may lead to disruptions in school progression, increasing the likelihood of educational delay and reducing the chances of completing primary education. The fact that these effects are observed despite continued school attendance highlights the role of academic performance and persistence as key margins through which the quantity-quality trade-off materializes.

The quantity-quality trade-off observed in the 3+ sample reflects the average marginal effect for the first- and second-born children. In Table 3, we explore heterogeneity by birth order. We find that for the z-score, the negative effect is now concentrated on the second-born child and is statistically significant at 5%. Interestingly, we now also observe a negative effect on school attendance for the second-born. Although splitting the sample reduces our ability to estimate the effects precisely, these results suggest that, for the child closest in age to the newborn, the school attendance channel may also be operating.

For the probability of falling behind in school, no significant differences are found between the first and second-born; the effect is quite similar in magnitude and both are statistically significant at 5% level. Taken together, the results suggest that family size imposes a negative effect on school delay for both siblings, but it is the second-born who

¹⁰Bhalotra and Clarke (2020) suggests that this is evidence of partial reduction of future fertility following twin births (i.e., a compensating fertility behavior).

¹¹The coefficient is multiplied by the standard deviation of completed years of education of children of 15 years old (see Table A.9).

appears to bear more lasting consequences, ultimately facing lower educational attainment, either due to reduced school attendance or difficulties in completing grades.

TABLE 2 OLS and Twins IV estimates

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Sample 2+</i>								
First-stage		0.719*** (0.016)		0.719*** (0.016)		0.699*** (0.017)		0.648*** (0.024)
Number of children	-0.136*** (0.002)	0.037 (0.023)	-0.029*** (0.001)	-0.004 (0.008)	0.059*** (0.001)	-0.011 (0.011)	-0.042*** (0.001)	0.014 (0.013)
Observations	377,328	377,328	376,555	376,555	330,009	330,009	209,453	209,453
Mean	0.176	0.176	0.871	0.871	0.247	0.247	0.826	0.826
<i>Panel B: Sample 3+</i>								
First-stage		0.789*** (0.018)		0.789*** (0.018)		0.781*** (0.019)		0.740*** (0.025)
Number of children	-0.149*** (0.002)	-0.036* (0.022)	-0.034*** (0.001)	-0.009 (0.008)	0.064*** (0.001)	0.028** (0.011)	-0.050*** (0.001)	-0.021* (0.012)
Observations	416,258	416,258	415,141	415,141	370,172	370,172	236,762	236,762
Mean	0.060	0.060	0.847	0.847	0.306	0.306	0.780	0.780

Notes: The table shows the OLS and twins IV estimates of the coefficient β_1 of Equation 1 for the “2+ sample” (Panel A) and “3+ sample” (Panel B). The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). The “2+ sample” comprises first-born non-twins aged 6-18, while the “3+ sample” includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children’s age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

TABLE 3 IV estimates by birth order

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	First	Second	First	Second	First	Second	First	Second
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Sample 3+</i>								
First-stage	0.812*** (0.018)	0.767*** (0.019)	0.811*** (0.018)	0.766*** (0.019)	0.809*** (0.019)	0.751*** (0.021)	0.774*** (0.024)	0.697*** (0.029)
Number of children	-0.018 (0.027)	-0.053** (0.027)	-0.002 (0.010)	-0.018* (0.010)	0.027** (0.013)	0.029** (0.014)	-0.018 (0.014)	-0.025 (0.017)
Observations	206,097	210,161	205,552	209,589	188,880	181,292	129,977	106,785
Mean	0.060	0.060	0.834	0.860	0.319	0.293	0.788	0.769

Notes: The table shows the twins IV estimates of the coefficient β_1 of Equation 1 by birth order of the child (first or second) in the “3+ sample”. The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). The “3+ sample” comprises first and second-born non-twins aged 6-18 (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children’s age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

4.1 | Robustness analysis

| IV estimates when selection into twinning is not controlled for

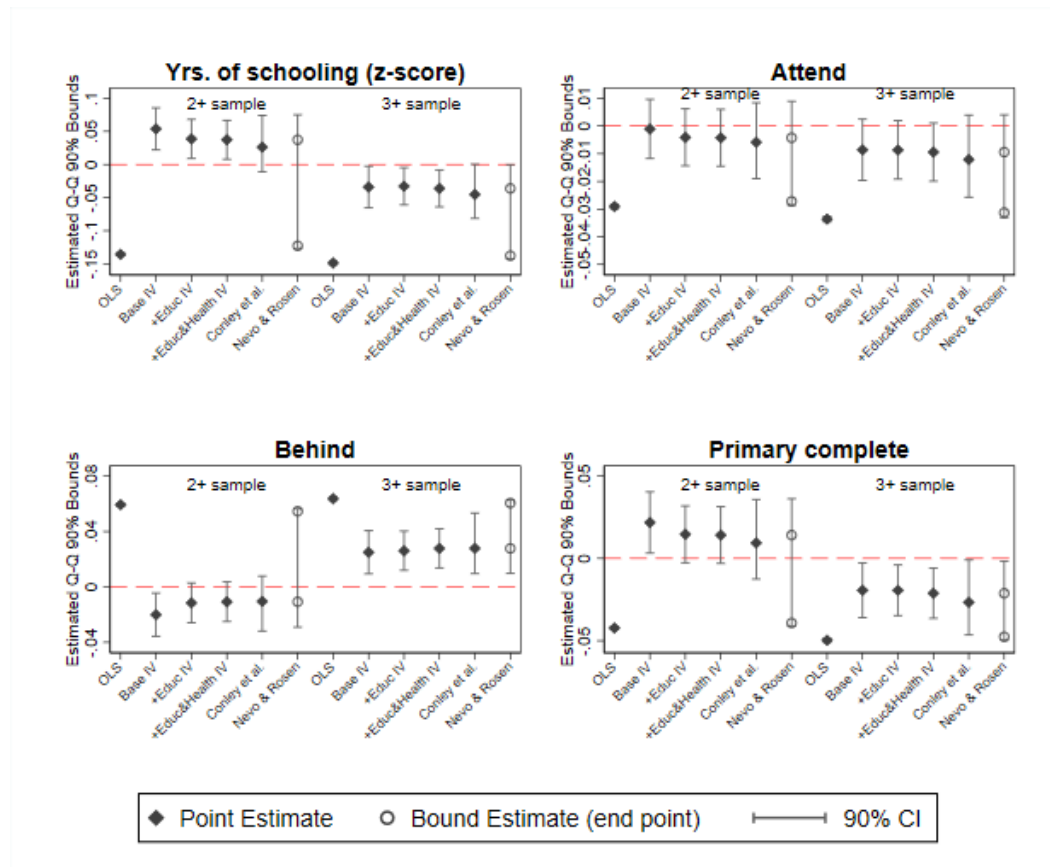
Bhalotra and Clarke (2020) shows that the omission of mother’s characteristics that are positively correlated with the probability of having twins makes the IV estimate of the β_1 parameter closer to zero or more positive than the true estimate (and/or estimates to be less precise). In this sense, the IV estimation without controlling for these characteristics can be considered an upper bound of the true estimate, and that the inclusion of these mother’s characteristics can help tightening this bound. We perform an analysis that excludes these characteristics and compare them with our OLS and IV estimates. Figure 1 shows, for each sample and outcomes, the OLS estimate, a base IV estimate that does not control for mother’s education or health characteristics (Base IV), an IV estimate that only controls for mother’s education (+Educ IV), and the IV estimate of our main specification which controls for education and health, and interactions between the two (+Educ&Health IV). In all cases, we find that $\hat{\beta}_1^{OLS}$ and $\hat{\beta}_1^{baseIV}$ bound the IV estimates that control for selection into twinning and that, as expected, $\hat{\beta}_1^{baseIV} > \hat{\beta}_1^{IV, Educ} > \hat{\beta}_1^{IV, Educ+Health}$.

| IV Bounds with an Imperfect Instrument

The variables we use to control for selection into twinning may only adjust partially for all dimensions characterizing healthier mothers leaving room for concerns regarding instrument validity. To overcome this concern, we implement two procedures to estimate IV bounds with an imperfect instrument.

First, we estimate bounds following the approach developed by Conley et al. (2012) for cases when the instrument is plausible but fails the exclusion restriction. Consider the model:

FIGURE 1 Parameter and bound estimates of the quantity-quality trade-off



Note: Each set of estimates refer to the 90% confidence intervals on parameters bounds of the impact of family size on child educational outcomes, together with the OLS point estimates reported in Table 2. “Base IV” refers to the IV estimate that does not control for mother’s education or health characteristics, “+Educ IV” the IV estimate that only controls for mother’s education, and “+Educ&Health IV” the IV estimate of our main specification which controls for education and health, and interactions between the two reported in Table 2. “Conley et al” and “Nevo Rosen” report the 90% confidence intervals estimated using the procedures of Conley et al. (2012) and Nevo and Rosen (2012), respectively, and implemented using the Stata commands `plausexog` and `imperfectiv` programmed by Clarke and Matta (2018). The hollow circles refer to the lower-bound and upper-bound point estimates of the Nevo and Rosen (2012) procedure, $\beta_{v(1)}^{iv}$ and β_z^{iv} (see footnote 1045 for further details). Source: Data from IPUMS International samples.

$$y_{ihjd} = \delta_0 + \delta_1 Nchild_{ih} + \lambda twins_{ih} + \delta_2 X_i + \delta_3 X_{ih}^M + \delta_3 X_d + \gamma_j + u_i, \quad (3)$$

The parameter λ reflects how close the exclusion restriction is to being satisfied. When it is satisfied, it is equivalent to imposing ex-ante that λ is precisely equal to zero. One can define plausible exogeneity as a situation in which λ is nearly, but not equal to zero. Since the parameters δ_1 and λ are not jointly identified, the method proposed by [Conley et al. \(2012\)](#) uses prior information or assumptions about λ to obtain estimates of the parameter of interest, δ_1 .¹² We assume λ to be, in our setting, between 0 (in which case the IV is valid) and 20% of the absolute value of the effect of number of children on the corresponding outcome that results from our baseline IV estimates. For example, in the case of schooling years z-score in the sample 3+ this value is 0.007, which implies a direct effect of the twins dummy of 0.7% of a standard deviation of this outcome. These priors are guided by the discussion in [Bhalotra and Clarke \(2020\)](#) and [Bhalotra and Clarke \(2019\)](#), which provide evidence consistent with λ being different from zero in a similar context than ours, although small in magnitude.

The second approach to construct valid IV bound estimates follows [Nevo and Rosen \(2012\)](#). The classical IV approach produces consistent estimates based on the unobservable validity assumption that the instrumental variable is not correlated with the error term in Equation 1. [Nevo and Rosen \(2012\)](#) documents that bounds inference in a linear IV model can proceed with weaker-than-classical assumptions by replacing the zero correlation assumption with an assumption on the sign of this correlation.¹³

Figure 1 plots the 90% confidence intervals estimated for the bounds of each procedure. The bound estimates provide some relevant information. First, they enable us to determine that the effect is negative in the 3+ sample for the outcomes behind two years with both methods. Second, it confirms the direction of the point estimates when we allow for some degree of selection into twinning, estimating a higher effect for family size. This validates that our estimates are an upper bound of the true effect, as expected, even when we are not able to distinguish it from being statistically different from zero for schooling and primary, due to the higher standard errors associated with these methods. Additionally, the bounds of both procedures reject a positive effect of family size on educational outcomes in the sample 2+. Lastly, the bound estimates never exclude the point estimate of our main IV estimate but they do exclude the OLS estimates.

In the case of the method proposed by [Conley et al. \(2012\)](#), the point estimate reported in the figure assumes a specific value for the parameter λ . Rather than selecting an arbitrary value, it is possible to conduct a sensitivity analysis of the estimator across a range different

¹²[Conley et al. \(2012\)](#) show that bounds for the IV parameter β_1 from equation 1 can be generated under a series of assumptions regarding λ . The assumptions can be made regarding the range of values that λ can take, regarding the entire distribution of λ , or a fully Bayesian approach, in which that researcher must provide a prior for λ , as well as priors for each model parameter and the distribution of the error term. Our estimates are based on the assumption regarding the support of λ (their “Union of Confidence Intervals”, or UCI, approach) which we implement using the Stata command `plaussexog` programmed by [Clarke and Matta \(2018\)](#). Given this assumption, the model $(y_{ihjd} - \lambda twins_{ih}) = \delta_0 + \delta_1 Nchild_{ih} + \delta_2 X_i + \delta_3 X_{ih}^M + \delta_3 X_d + \gamma_j + u_i$ can be estimated by 2SLS using the transformed dependent variable $y - \lambda twins$. In practice, bounds consists of the union of all confidence intervals in the assumed range of λ .

¹³Consider the model $Y = \alpha + \beta X + \epsilon$, where X is the endogenous variable and Z a plausible instrument for X . [Nevo and Rosen \(2012\)](#) show that as long as: i) $\rho_{x,\epsilon} \rho_{z,\epsilon} \geq 0$, ii) $|\rho_{x,\epsilon}| \geq |\rho_{z,\epsilon}|$, and iii) $\rho_{x,z} < 0$, the true parameter β is bounded by the values $\beta_{v(1)}^{iv}$ and β_z^{iv} , where $\beta_{v(1)}^{iv}$ is the 2SLS estimator using the imperfect IV and $\beta_{v(1)}$ is the 2SLS estimator using the transformed (valid) instrument $V(\lambda^*) = (\sigma_x Z - \lambda^* \sigma_z X)$, with $\lambda^* = \frac{\rho_{z,\epsilon}}{\rho_{x,\epsilon}} = 1$. If iv) $\rho_{x,\epsilon} < 0$, then the upper and lower limits can be defined as $\beta_{v(1)}^{iv} \leq \beta \leq \beta_z^{iv}$. If assumption (ii) were not to meet, then $\beta^{OLS} \leq \beta \leq \beta_z^{iv}$.

values, which reflect varying degrees of violation of the exclusion restriction. This is shown in Figure A.6 for the 2+ sample and Figure A.5 for the 3+ sample in the Appendix. The effect of the quantity–quality trade-off becomes increasingly significant as selection into twinning becomes more pronounced in the estimation. Given that our controls for maternal health are partial and less comprehensive than those used by Bhalotra and Clarke (2020), it is likely that we are not fully ruling out selection. Therefore, positive values of λ are to be expected. Even under relatively small magnitudes, the sensitivity analysis confirms a lower bound below zero for the sample 3+, particularly in the cases of falling behind.

| Identification of twins

As we mentioned before, one advantage of our data is that it allows for the correct identification of twins because we have access to their exact date of birth (month-year). Some existing studies in the literature (e.g., Ponczek and Souza (2012)) rely only on information about the actual age of children for this classification, as many databases lack information about individuals' exact dates of birth. This implies that some siblings with the same age at the date of the census are classified as twins when they are not actually twins (e.g. when the birth interval between them is between nine and 11 months). Potentially, families with closely spaced births are those of lower socioeconomic status; hence, if this were the case, measurement error would be correlated with the outcome of interest. We present evidence that the misclassification of twins biases our estimates, indicating that the correct identification of twins is crucial for obtaining accurate estimates of the parameter of interest.

Table 4 displays the results of our main OLS and IV specifications, considering the alternative identification of twins based exclusively on the current age of children. The new classification adds 3,350 new twins to our initial sample. We know from our data that all of them are siblings who have the same age at the census date but were born in different month-years. When comparing these results with Table 2, we find that the misclassification of twins leads to a larger quantity-quality trade-off, while the first stage remains almost the same. For all outcomes and samples, the quantity quality trade-off is negative and larger in magnitude than estimates in Table 2, and generally statistically significant.

TABLE 4 Replication of main results following an imperfect twin sample definition

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Sample 2+</i>								
First-stage		0.775*** (0.013)		0.775*** (0.013)		0.761*** (0.014)		0.718*** (0.020)
Number of children	-0.136*** (0.002)	-0.052*** (0.016)	-0.029*** (0.001)	-0.017*** (0.006)	0.059*** (0.001)	0.035*** (0.008)	-0.042*** (0.001)	-0.016 (0.010)
Observations	378,512	378,512	377,738	377,738	331,070	331,070	210,123	210,123
Mean	0.176	0.176	0.871	0.871	0.247	0.247	0.825	0.825
<i>Panel B: Sample 3+</i>								
First-stage		0.878*** (0.016)		0.878*** (0.016)		0.872*** (0.018)		0.853*** (0.023)
Number of children	-0.148*** (0.002)	-0.068*** (0.016)	-0.034*** (0.001)	-0.012** (0.006)	0.064*** (0.001)	0.041*** (0.008)	-0.050*** (0.001)	-0.039*** (0.009)
Observations	417,487	417,487	416,367	416,367	371,272	371,272	237,469	237,469
Mean	0.0592	0.0592	0.847	0.847	0.307	0.307	0.779	0.779

Notes: The table replicates the estimates and specifications detailed in Table 2 for the same outcomes, with the unique difference in the twin sample definition. This criteria leaves aside the birth month and the birth year of siblings and directly defines a twin pair as two siblings of the same age. The “2+ sample” comprises first-born non-twins aged 6-18, while the “3+ sample” includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children’s age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

5 | HETEROGENEOUS EFFECTS

We next look at heterogeneous effects of family size according to parents’ characteristics, children’s characteristics and key characteristics of the district of residence. This analysis allows us to shed light on the potential mechanisms affecting the relationship between family size and children’s human capital. First, we estimate the heterogeneous effects by parent’s characteristics. A first dimension we look at is parents’ educational attainment as a proxy of their socioeconomic status.¹⁴ Financial restrictions may be more binding for households of lower socioeconomic backgrounds; hence, fertility shocks could have a larger effect on children’s human capital investments of these families.

A second dimension in the heterogeneity analysis refers to children’s characteristics. We look at the heterogeneous effects by the gender of the child. Gender norms may affect parental investment decisions in their children’s human capital upon the onset of a fertility shock. Finally, we analyze heterogeneous effects by characteristics of the district of residence.¹⁵ We split the sample according to the provision of education services across

¹⁴We divide parents’ educational attainment into “Low” (parents with incomplete secondary education or less) and “High” (at least one parent with completed secondary education or more).

¹⁵In general, the district refers to the lowest level of sub-national government (municipalities or departments)

districts, which we measure as the ratio of teachers to school-age children.¹⁶ We expect that in districts with high supply restrictions both large and small families face limitations to school access, while in districts with many teachers both types of families are not constrained in this dimension. Hence, in both of these cases the effect of family size on human capital of children is expected to be weaker.¹⁷

5.1 | Parents' socioeconomic characteristics

Considering families with low-educated parents in the 3+ sample, one additional child reduces the z-score of schooling years, reduces the probability of completing primary school and increases the chances of falling behind in school progress (Table 5). The magnitude of these effects for the column of low-educated parents is higher than Table 2, whereas the effects are null for children in the same sample but with highly educated parents. Taking the probability of falling behind and primary completion as examples, we estimate effects of 0.6 and 0.4 percentage points, respectively, for children of highly educated parents, neither of which is statistically distinguishable from zero. In contrast, for children of low-educated parents, the estimated effects amount to 3.7 and 3.6 percentage points, respectively—both significant at the 5% level. For this latter group, the estimates correspond to roughly a 10% and 5% decrease relative to the mean.

This is in line with results from [Bhalotra and Clarke \(2020\)](#) who found that negative effects in developing countries are concentrated for mother of low educational level, and with findings from [Ponczek and Souza \(2012\)](#) and [Dayioğlu et al. \(2009\)](#) indicating that detrimental effects on schooling are more pervasive among low-educated mother families and that limited financial resources play an important role in bringing about the sibling composition effects.

For the 2+ sample, we do not find evidence of heterogeneous effects, and interestingly, having one more child has a slightly positive effect on the sibling z-score only for families with high socioeconomic status. This is in line with results from [Alidou and Verpoorten \(2019\)](#) where the positive result is shown to be mainly driven by children from non-poor families.

The results align with the hypothesis that the tension between quality and quantity manifests markedly in families where household resources are more limited, because of a larger family composition and a lower parental socioeconomic status.

in each country. The number of districts in the IPUMS-I data are: 433 in Colombia (2005), 105 in Dominican Republic (2002), 65 in Dominican Republic (2010), 77 in Ecuador (2010), 68 in Nicaragua (1995, 2005), and 103 in El Salvador (1992, 2007).

¹⁶We identify teachers through occupation codes available in the census data. We do not distinguish teachers according to the educational level in which they teach.

¹⁷We classify the districts in three terciles (q1, q2, and q3) measuring the ratio of education workers to school-age children constructed using the entire sample of districts.

TABLE 5 IV estimates by maximum education level of parents

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	Low	High	Low	High	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Sample 2+</i>								
First-stage	0.688*** (0.024)	0.773*** (0.017)	0.688*** (0.024)	0.772*** (0.017)	0.665*** (0.026)	0.754*** (0.019)	0.600*** (0.036)	0.724*** (0.027)
Number of children	0.027 (0.034)	0.046* (0.028)	-0.004 (0.013)	-0.007 (0.008)	-0.012 (0.017)	-0.008 (0.012)	0.016 (0.022)	0.017 (0.013)
Observations	236,221	141,107	235,567	140,988	206,673	123,336	132,946	76,507
Mean	-0.004	0.479	0.827	0.944	0.336	0.0966	0.766	0.928
<i>Panel B: Sample 3+</i>								
First-stage	0.753*** (0.023)	0.878*** (0.020)	0.752*** (0.023)	0.877*** (0.020)	0.740*** (0.026)	0.872*** (0.021)	0.689*** (0.032)	0.862*** (0.028)
Number of children	-0.054* (0.029)	-0.001 (0.029)	-0.013 (0.011)	0.004 (0.008)	0.037** (0.015)	0.006 (0.013)	-0.036** (0.017)	0.004 (0.012)
Observations	300,094	116,164	299,100	116,041	265,674	104,498	169,572	67,190
Mean	-0.091	0.450	0.812	0.936	0.379	0.121	0.726	0.915

Notes: The table shows the twins IV estimates of the coefficient β_1 of Equation 1 by the maximum education level of the parents (Low and High). The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). “Low” education refers to at most primary education complete or lower; “High” education refers to secondary education complete or more. The “2+ sample” comprises first-born non-twins aged 6-18, while the “3+ sample” includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children’s age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

5.2 | Children’s gender

Parental decisions regarding human capital investments tend to differ depending on the gender of children at the time of the fertility shock. Table 6 shows and interest pattern: now the mechanism of school attendance become relevant, with a increase in family size affecting negatively to the second-born male, decreasing their probability of being enrolled in education in 3.0 percentage point (a 3.5 % decrease relative to the sample mean), whereas for females, no significant effect is registered. For the remaining outcomes, the point estimates have the expected sign, but they are not statistically distinguishable from zero at conventional levels for either females or males.

We further explore potential heterogeneity in the probability of falling behind by splitting the analysis by the child’s age within each gender (see Table A.11 in the Appendix). The table reveals that in the 3+ sample, the negative effect on the probability of falling behind is verified for 10 to 14-year-old boys, with no effect for the same age female siblings. In contrast, female children in the 2+ sample families appear to experience positive effects.

These differential effects are consistent with the idea that gender norms and roles within the household influence intra-household allocation of resources and responsibilities. In

larger families where financial constraints are binding, boys may be expected to contribute economically by entering the labor market, which increases their likelihood of dropping out of school or falling behind academically. For these children, the fertility shock intensifies the quality-quantity trade-off.

TABLE 6 IV estimates by birth order and gender of the child (3+ Sample)

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	First	Second	First	Second	First	Second	First	Second
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Females</i>								
First-stage	0.781*** (0.025)	0.749*** (0.026)	0.781*** (0.025)	0.747*** (0.026)	0.778*** (0.027)	0.727*** (0.029)	0.737*** (0.035)	0.660*** (0.041)
Number of children	-0.004 (0.039)	-0.056 (0.040)	-0.003 (0.015)	-0.004 (0.014)	0.030 (0.019)	0.021 (0.021)	-0.026 (0.021)	-0.026 (0.026)
Observations	98,898	101,604	98,633	101,332	90,493	87,522	61,884	51,083
Mean	0.151	0.149	0.853	0.875	0.273	0.249	0.819	0.805
<i>Panel B: Males</i>								
First-stage	0.842*** (0.025)	0.782*** (0.027)	0.840*** (0.025)	0.782*** (0.027)	0.837*** (0.027)	0.771*** (0.030)	0.807*** (0.034)	0.722*** (0.042)
Number of children	-0.032 (0.037)	-0.044 (0.036)	-0.002 (0.014)	-0.030** (0.015)	0.024 (0.018)	0.031 (0.020)	-0.015 (0.019)	-0.020 (0.023)
Observations	107,199	108,557	106,919	108,257	98,387	93,770	68,093	55,702
Mean	-0.024	-0.024	0.815	0.846	0.361	0.335	0.761	0.737

Notes: The table shows the twins IV estimates of the coefficient β_1 of Equation 1 by birth order (first or second) and gender of the child in the “3+ sample”. The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). The “3+ sample” comprises first and second-born non-twins aged 6-18 (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children’s age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Source: Data from IPUMS International samples.

In contrast, in the 2+ sample -where families are typically less constrained— an additional child might alter household dynamics in ways that might benefit older girls. A plausible hypothesis is that the arrival of a younger sibling encourages formal schooling for both the newborn and the older sister. In the absence of the fertility shock, the older daughter might have assumed caregiving roles or been withdrawn from school to help with domestic work. Thus, the fertility shock may reduce her household responsibilities, enabling continued or improved educational participation. These patterns may reflect parental preferences and expectations shaped by traditional gender roles, where boys are expected to support the household financially, while girls face competing demands between school and domestic chores.

To explore these mechanisms, we estimate the same IV regressions using two alternative outcomes: children’s labor force participation and involvement in household chores. Table 7, Panel B, shows that an increase in family size significantly raises the probability that 15–18-year-old male children are working at the time of the census, whereas female children are not affected. The effect is significant statistically and economically, representing a 18% higher chance to be involved in labor activities for a 10 to 15-year-old male sibling.

Furthermore, male children appear to spend less time on household chores, likely due to their engagement in income-generating activities outside the home.

TABLE 7 IV estimates by gender of the child for Labor Force Participation and Household Chores

Dependent variable:	Labor Force				Household Chores			
	10-14		15-18		10-14		15-18	
	Female	Male	Female	Male	Female	Male	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Sample 2+</i>								
First-stage	0.748*** (0.034)	0.705*** (0.035)	0.597*** (0.050)	0.601*** (0.050)	0.766*** (0.037)	0.713*** (0.040)	0.591*** (0.052)	0.603*** (0.054)
Number of children	-0.008 (0.009)	0.004 (0.014)	0.022 (0.027)	-0.037 (0.032)	0.000 (0.021)	-0.004 (0.018)	-0.077** (0.035)	-0.014 (0.024)
Observations	78,111	81,573	49,074	55,215	66,034	69,614	42,595	49,037
Mean	0.031	0.063	0.096	0.244	0.163	0.086	0.221	0.088
<i>Panel B: Sample 3+</i>								
First-stage	0.766*** (0.030)	0.811*** (0.032)	0.654*** (0.045)	0.741*** (0.043)	0.766*** (0.036)	0.798*** (0.037)	0.639*** (0.051)	0.751*** (0.047)
Number of children	-0.011 (0.009)	-0.011 (0.011)	0.030 (0.024)	0.051** (0.026)	-0.008 (0.021)	-0.021* (0.013)	0.017 (0.034)	-0.001 (0.018)
Observations	90,116	94,267	54,570	61,800	75,598	79,706	46,832	54,635
Mean	0.035	0.076	0.105	0.278	0.169	0.084	0.232	0.081

Notes: The table shows the twins IV estimates of the coefficient β_1 of Equation 1 by gender of the child. The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). The “2+ sample” comprises first-born nontwins aged 15-18, while the “3+ sample” includes first and second-born nontwins of the same age range (refer to Section 2.3 for further details). The labor force participation and household chores outcomes are measured only for children aged 10 to 18. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

Conversely, panel A reveals a statistically significant reduction in household chores for 15–18-year-old female children in the 2+ sample panel. These findings align with the hypothesis that, in households where financial pressure is less severe, the fertility shock can reduce girls’ domestic burdens and, in turn, facilitate their continued schooling. Instead, in larger families boys close to working age may involve in labor activities to supplement family income.

5.3 | Characteristics of the district of residence: educational supply

Table 8 shows the heterogeneous effects by a key district’s characteristics: the ratio of teachers to school-age children. Districts are divided in terciles according to the distribution of this variable. We find that the negative effects of family size on educational outcomes

for the 3+ sample, are concentrated in districts with an intermediate supply of formal education. Only in the second tertile do we find effects that are statistically significant at the 5% level, amounting to 9.9 and 5.7 percentage points for years of schooling and primary completion, respectively. For falling behind, we estimate an effect of 7.4 percentage points in the same tertile, significant at the 1% level. This suggests that in these areas, education is relatively more accessible compared to under-supplied districts, yet still scarce enough for the marginal cost of additional education to be salient for families facing resource constraints.

As expected, in under-supplied districts, the trade-off seems to be largely absent. In these areas, the structural barriers to accessing education may be so severe that even small families may find it difficult to educate their children. The addition of another child does not significantly alter this baseline. In contrast, in districts with high educational provision, we also observe a statistically insignificant trade-off. In these contexts, education is abundant and more easily accessible. Families, regardless of size, may be better able to accommodate the costs of educating all children, especially if public provision reduces the marginal cost of schooling.

The positive effect for the 2+ sample becomes present—in the second tercile column. This pattern aligns with the hypothesis that family size effects, particularly the potential gains from household-level economies of scale, are more likely to emerge in areas with a moderate supply of formal education. In areas where educational supply is low or high, the arrival of a new sibling has little to no effect.

Overall, these results suggest that local educational supply is a relevant mediator in the impact of fertility shocks on children's schooling. The trade-off between family size and education may be context-dependent, emerging where education is neither highly limited nor abundant, underscoring the role of both household constraints and institutional capacity in shaping child outcomes.

6 | NON-LINEAR MODELS

So far, we have presented evidence for the quantity-quality trade-off based on the assumption of a linear relationship between family size and child human capital formation. In other words, this IV setting imposes that any additional child in the family has the same marginal effect on the educational outcomes of the first or the first and second born. However, there are arguments to think of non-linearities in the arrival of siblings. Recent literature has shown compensation behaviour in parental investment, challenging the assumption that resources and endowments are homogeneous between siblings. But even if this assumption holds, we should not discard that moving from 1 to 2 children has a different marginal effect than moving from 2 to 3, from 3 to 4 and so on.

This discussion is essential for a correct understanding of our empirical results. As shown in Angrist and Imbens (1995) the linear IV estimator can be decomposed into a weighted average of underlying marginal effect. If non-linear patterns underlie the Q-Q trade-off, the marginal effect we report in the previous section would be an average that could mask different marginal effects that vary from child to child.

| Generated instruments

In this section, we relax the classical assumption of constant marginal effect of family size in the IV estimation allowing for the coefficient β_1 in model 1 to be different across siblings. This implies adopting an unrestricted model in family size, replacing the linear family size

TABLE 8 IV estimates by education supply of the district of residence (terciles)

Dependent variable:	Yrs. School. (z-score)			Attend			Behind			Primary		
	q1 (1)	q2 (2)	q3 (3)	q1 (4)	q2 (5)	q3 (6)	q1 (7)	q2 (8)	q3 (9)	q1 (10)	q2 (11)	q3 (12)
<i>Panel A: Sample 2+</i>												
First-stage	0.728*** (0.031)	0.690*** (0.034)	0.738*** (0.021)	0.729*** (0.031)	0.691*** (0.034)	0.737*** (0.021)	0.710*** (0.034)	0.657*** (0.038)	0.722*** (0.023)	0.659*** (0.043)	0.611*** (0.054)	0.670*** (0.033)
Number of children	0.028 (0.041)	0.101* (0.052)	0.003 (0.032)	0.017 (0.015)	-0.033* (0.019)	-0.005 (0.010)	-0.029 (0.022)	-0.029 (0.025)	0.010 (0.015)	0.014 (0.022)	0.074** (0.032)	-0.013 (0.019)
Observations	100,023	92,829	184,476	99,495	92,658	184,402	86,220	78,964	164,825	59,305	50,208	99,940
Mean	0.083	0.116	0.258	0.834	0.835	0.908	0.294	0.302	0.196	0.807	0.780	0.859
<i>Panel B: Sample 3+</i>												
First-stage	0.733*** (0.036)	0.824*** (0.038)	0.799*** (0.022)	0.730*** (0.036)	0.824*** (0.038)	0.799*** (0.022)	0.719*** (0.040)	0.818*** (0.041)	0.791*** (0.024)	0.699*** (0.047)	0.762*** (0.056)	0.752*** (0.030)
Number of children	0.011 (0.046)	-0.099** (0.041)	-0.015 (0.030)	-0.013 (0.019)	-0.012 (0.016)	-0.003 (0.010)	-0.013 (0.024)	0.074*** (0.021)	0.022 (0.015)	-0.023 (0.025)	-0.057** (0.026)	0.004 (0.015)
Observations	110,952	111,492	193,814	110,173	111,252	193,716	96,777	96,954	176,441	65,723	61,483	109,556
Mean	-0.079	0.003	0.173	0.797	0.815	0.893	0.371	0.356	0.243	0.745	0.732	0.827

Notes: The table shows the twins IV estimates of Equation 1 by education supply of the district (in terciles). The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). "q1" refers to the first tercile, "q2" refers to the second tercile, and "q3" refers to the third tercile. Education supply is measured as the ratio of education workers to school-age children (6-18 years) in the district. Terciles are constructed using the entire sample of districts. The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children's age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the "3+ sample", standard errors are clustered by mother's ID. *** p<0.01, ** p<0.05, * p<0.1.

Source: Data from IPUMS International samples.

variable, with dummies for each number of children in the second stage. For simplicity, we illustrate the case for the 2+ sample:

$$y_{ihjd} = \alpha + \beta_3 Nchild_{3ih} + \beta_4 Nchild_{4ih} + \beta_5 Nchild_{5ih} + \beta_6 Nchild_{6ih} + X'_i \beta_x + \mu_i, \quad (4)$$

Where $Nchild_{c ih}$ is an indicator variable for child i in family h taking the value of 1 if $Nchild_{ih} \geq c$ with $c = 3 \dots 6$. In our first stage, we now have several endogenous variables to instrument for:

$$Nchild_{c ih} = \pi_{s2} Twins_{2ih} + \pi_{s3} Twins_{3ih}^* + \pi_{s4} Twins_{4ih}^* + \pi_{s5} Twins_{5ih}^* + X'_i \pi_x + \eta_i, \quad c = 3 \dots 6 \quad (5)$$

To estimate this equation, first, we need to overcome the problem of non-defining variables or technically, a problem of partially missing instruments. Taking the case of 2+ samples (i.e. families with at least 2 children) for all these families we can define $Twins_{2ih}$ as a binary variable equal to 1 if the second birth was a twin birth. When moving one order ahead, the instruments for twins at the third birth - $Twins_{3ih}^*$ - (and any higher order) are simply not defined for families with 2 children. The same occurs, for example, for the instrument of twins at a fourth birth for families with 3 or fewer children. In resume, we can not observe for the entire sample the full set of instruments we need to instrument the endogenous variables in the unrestricted model.¹⁸

In order to construct new valid instruments we follow the methodology proposed by Angrist et al. (2010) and then extended by Mogstad and Wiswall (2012). As shown by the authors, with the same assumptions required for the linear IV, it is possible to construct valid instruments for the entire sample, defined as a linear projection of the twin instruments on the included set of control X . For our 2+sample ($c = 2$) and for the sample 3+sample ($c = 3$) the generated instruments are defined as:

$$twin_{ci}^* = \begin{cases} 0, & \text{if } Nchild_i < c, \\ twin_{ci} - E[twin_{ci}|X_i, Nchild_i \geq c], & \text{if } Nchild_i \geq c, \end{cases} \quad (6)$$

where the term $E[twin_{ci}|X_i, Nchild_i \geq c]$ is the OLS estimation of the conditional probability of twin birth given X_i in the sample where it is nonmissing.¹⁹

| Results

For the 2+ sample, the positive effect on years of schooling is only verified for the lower parity, when the third birth was a twin, whereas for higher twin orders, the coefficients turn negative, becoming statistically significant at the 10% level for the fifth birth. A similar pattern is found for the primary outcome, a small positive effect when moving from 2 to 3 children, which then turns negative and significant when moving from 3 to 4. The non-linear

¹⁸As in Mogstad and Wiswall (2016), for the estimation of non-linearities, we exclude a small number of families with more than 6 children (1.3% in 2+ sample and 2.6% in 3+ sample).

¹⁹Specifically we include as covariates a full set of dummy variables for child's age, sex, year and month of birth, mother's age and age at first child born and parent education, a mortality dummy, the interaction between education dummies with mortality, and mortality interacted with country fixed effects.

estimates also show a negative and significant negative effect on school attendance when the family experience a size shock due to a twin at the fourth birth.

In the 3+ sample, the estimates report an increasing negative effect on the probability of falling behind when comparing the twin at the fourth birth with the effect at the fifth birth. The non-linear effect is also increasing for years of schooling and school attendance, and more stable for primary enrollment, although not significantly distinguishable from zero. Due to the problem of small sample size for higher twin orders, typically these estimates suffer from high standard errors, as noted by [Mogstad and Wiswall \(2016\)](#). The coefficients become noisier at, for example, the sixth birth order, where twinning is extremely rare in the data.

TABLE 9 Standard IV estimates for the unrestricted model of family size for 2+ sample

Dependent variable:	Yrs. School. (z-score)	Attend	Behind	Primary
	(1)	(2)	(3)	(4)
Children ≥ 3	0.072** (0.034)	-0.007 (0.015)	-0.024 (0.021)	0.038* (0.020)
Children ≥ 4	-0.035 (0.035)	0.016 (0.011)	0.031 (0.020)	-0.045** (0.019)
Children ≥ 5	-0.092* (0.048)	-0.057*** (0.019)	0.031 (0.023)	-0.022 (0.030)
Children ≥ 6	-0.026 (0.110)	-0.005 (0.040)	-0.021 (0.042)	0.043 (0.044)
Joint Significance (p-value)	0.036	0.000	0.144	0.005
Observations	372,573	371,840	325,256	205,030
Mean	0.187	0.875	0.240	0.832

Notes: The table shows the Standard IV estimates for the unrestricted model of family size for the “2+ sample”. Children ≥ 3 is the marginal effect for moving from 2 to 3 children, Children ≥ 4 of moving from 3 to 4, and so on. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regressions. The p-value for the Joint Significance test refers to the null hypothesis that each parameter is zero. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

In Appendix B, we also offer estimates obtained using efficient instruments following the approach in [Mogstad and Wiswall \(2016\)](#). In line with their findings, these estimates exhibit greater precision with results that are substantially larger in magnitude. However, the interpretability of the larger estimates is less transparent, and it is not straightforward, which motivates caution when drawing substantive conclusions from this particular specification.

Despite the difficulty of obtaining precise estimates from the non-linear instruments, taken together, both tables highlight the importance of extending the linear estimates to account for heterogeneous marginal effects that may otherwise mask substantial trade-offs at higher birth orders. This is often an aspect where the literature remains silent or has been largely overlooked. Addressing the quantity–quality trade-off solely through average linear

effects may ignore potentially much stronger penalties for large families, which is an issue that may be particularly relevant in the context of Latin American countries.

TABLE 10 Standard IV estimates for the unrestricted model of family size for 3+ sample

Dependent variable:	Yrs. School. (z-score)	Attend	Behind	Primary
	(1)	(2)	(3)	(4)
Children ≥ 4	-0.037 (0.032)	-0.003 (0.011)	0.031** (0.014)	-0.026 (0.019)
Children ≥ 5	-0.066 (0.043)	-0.031 (0.021)	0.041* (0.025)	-0.024 (0.023)
Children ≥ 6	-0.032 (0.077)	-0.002 (0.028)	-0.019 (0.034)	0.039 (0.040)
Joint Significance (p-value)	0.071	0.424	0.004	0.314
Observations	405,310	404,284	359,253	227,237
Mean	0.080	0.854	0.295	0.792

Notes: The table shows the Standard IV estimates for the unrestricted model of family size for the “3+ sample”. Children ≥ 4 is the marginal effect for moving from 3 to 4 children, Children ≥ 5 of moving from 4 to 5, and so on. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regressions. The p-value for the Joint Significance test refers to the null hypothesis that each parameter is zero. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

7 | CONCLUSIONS

The causal relationship between family size and children’s human capital accumulation has been extensively analyzed worldwide, yielding surprisingly inconclusive results. In this paper, we provide new evidence on the quantity–quality trade-off in five Latin American countries over the 1990s and 2000s. The harmonized Census data allow us to precisely identify twin births, which we use as instruments for the number of children in the household.

Our results indicate that a fertility shock negatively affects the probability of falling behind in educational progression, the completion of primary education, and consequently, years of schooling attained when focusing on families with three or more children. These adverse effects are concentrated among second-born children. By contrast, we find no statistically significant effects in families with two or more children.

The robustness of our main estimates highlights the importance of accounting for positive selection into twinning among healthier mothers, as recent literature suggests. By addressing potential endogeneity in the IV strategy, we confirm the direction of our results and their interpretation as a lower bound of the true effect. We also show that the precise identification of twins in the sample materially affects the results, both quantitatively and qualitatively.

The richness of the Census data allows us to shed light on the potential mechanisms at play. Negative effects are concentrated among low-educated families, where household resources are more limited, and in districts with intermediate educational supply, where public provision of schooling appears to mediate the impact of fertility shocks. Importantly,

the trade-off is more pronounced for male children, affecting both school attendance and the probability of falling behind. Our findings are consistent with gender norms and parental preferences that may lead boys to enter the labor market at early ages. For girls in the 2+ sample, the fertility shock may reduce domestic chore burdens, thereby facilitating continued schooling.

The policy implications highlight the role of gender norms and the provision of goods and services that complement parental efforts as crucial mechanisms in the interplay between fertility decisions and children's human capital development. The marked decline in fertility experienced by the region offers an opportunity to relax the quantity–quality trade-off, potentially contributing to narrowing human capital gaps. However, for more financially constrained families, fertility shocks may still generate negative consequences for older siblings, particularly at higher parity orders, as suggested by our non-linear estimates.

Complementing these findings with further research would help address some of the limitations of the analysis. Access to richer information on maternal health characteristics would allow us to better assess the magnitude of the estimated effects and rule out more comprehensively potential selection concerns, enhancing the instrument validity. This, in turn, would strengthen confidence in the results, as unobserved selection into twinning may partly explain the inconclusive findings in literature. Additionally, expanding the sample to include a larger set of countries would allow us to further validate and extend our conclusions in the context of the ongoing demographic transition in the region.

ACKNOWLEDGEMENTS

We thank the valuable comments from participants of conferences at Asociación Argentina de Economía Política and Society of Family and Gender Economics (GeFam), and seminar at University of San Andres. We also thank Brian Kovak, Andrew Garin and Felix Koenig for helpful comments and feedback. We are grateful to Iván Albina for helpful discussions regarding the analysis of IPUMS-I data and Juan Pantano for valuable comments. We thank Matías Maljar for outstanding research assistance. Views expressed here do not necessarily correspond to those of our affiliation.

REFERENCES

- Alidou, S. and Verpoorten, M. (2019) Family size and schooling in sub-saharan africa: testing the quantity-quality trade-off. *Journal of Population Economics*, **32**, 1353–1399.
- Angrist, J., Lavy, V. and Schlosser, A. (2010) Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, **28**, 773–824.
- Angrist, J. D. and Imbens, G. W. (1995) Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, **90**, 431–442.
- Åslund, O. and Grönqvist, H. (2010) Family size and child outcomes: Is there really no trade-off? *Labour Economics*, **17**, 130–139.
- Baez, J. E. (2008) Does more mean better? sibling sex composition and the link between family size and children's quality. *Tech. rep.*, Institute of Labor Economics (IZA).
- Becker, G. S. and Lewis, H. G. (1973) On the interaction between the quantity and quality of children. *Journal of political Economy*, **81**, S279–S288.
- Bhalotra, S. and Clarke, D. (2019) Twin birth and maternal condition. *Review of Economics and Statistics*, **101**, 853–864.

- (2020) The twin instrument: Fertility and human capital investment. *Journal of the European Economic Association*, **18**, 3090–3139.
- Bhalotra, S. and Rawlings, S. B. (2011) Intergenerational persistence in health in developing countries: The penalty of gender inequality? *Journal of Public Economics*, **95**, 286–299.
- Black, S. E., Devereux, P. J. and Salvanes, K. G. (2005) The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, **120**, 669–700.
- (2010) Small family, smart family?: family size and the iq scores of young men. *Journal of Human Resources*, **45**, 33–58.
- Cáceres-Delpiano, J. (2006) The impacts of family size on investment in child quality. *Journal of Human Resources*, **41**, 738–754.
- Clarke, D. and Matta, B. (2018) Practical considerations for questionable ivs. *The Stata Journal*, **18**, 663–691.
- Conley, D. (2008) Bringing sibling differences in: Enlarging our understanding of the transmission of advantage in families. In *Social Class: How Does it Work* (eds. A. Lareau and D. Conley), 179–200. New York, NY: Russell Sage Foundation.
- Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012) Plausibly exogenous. *Review of Economics and Statistics*, **94**, 260–272.
- Dayioğlu, M., Kirdar, M. G. and Tansel, A. (2009) Impact of sibship size, birth order and sex composition on school enrolment in urban turkey. *Oxford Bulletin of Economics and Statistics*, **71**, 399–426.
- Fitzsimons, E. and Malde, B. (2014) Empirically probing the quantity–quality model. *Journal of Population Economics*, **27**, 33–68.
- Guo, R., Yi, J. and Zhang, J. (2022) The child quantity–quality trade-off. *Handbook of labor, human resources and population economics*, 1–23.
- Hsin, A. (2012) Is biology destiny? birth weight and differential parental treatment. *Demography*, **49**, 1385–1405.
- Li, H., Zhang, J. and Zhu, Y. (2008) The quantity-quality trade-off of children in a developing country: Identification using chinese twins. *Demography*, **45**, 223–243.
- Li, J., Dow, W. H. and Rosero-Bixby, L. (2017) Education gains attributable to fertility decline: Patterns by gender, period, and country in latin america and asia. *Demography*, **54**, 1353–1373.
- Marteletto, L. J. and de Souza, L. R. (2012) The changing impact of family size on adolescents' schooling: Assessing the exogenous variation in fertility using twins in brazil. *Demography*, **49**, 1453–1477.
- de la Mata, D., Berniell, L., Schargrodsky, E., Álvarez, F. and Alves, G. (2022) Inherited inequalities: The role of skills, employment, and wealth in the opportunities of new generations.
- Mogstad, M. and Wiswall, M. (2012) Instrumental variables estimation with partially missing instruments. *Economics Letters*, **114**, 186–189.
- (2016) Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics*, **7**, 157–192.
- Nevo, A. and Rosen, A. M. (2012) Identification with imperfect instruments. *Review of Economics and Statistics*, **94**, 659–671.
- Onofri, M., Berniell, L., Fernández, R. and Menduina, A. (2026) Understanding latin america's fertility decline: Age, education, and cohort dynamics. *Tech. rep.*, National Bureau of Economic Research.

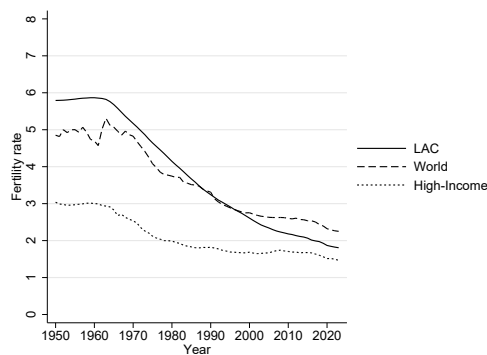
- Policy, E. and Center, D. (2018) National education profile: 2018 update. <https://www.epdc.org/taxonomy/term/90.html>.
- Ponczek, V. and Souza, A. P. (2012) New evidence of the causal effect of family size on child quality in a developing country. *Journal of Human Resources*, **47**, 64–106.
- Porta, E. and Laguna, J. R. (2007) Nicaragua: Country case study. country profile prepared for the education for all global monitoring report 2008. *Tech. rep.*, UNESCO, Paris.
- Rosenzweig, M. R. and Wolpin, K. I. (1980) Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica: journal of the Econometric Society*, 227–240.
- Rosenzweig, M. R. and Zhang, J. (2009) Do population control policies induce more human capital investment? twins, birth weight and china's "one-child" policy. *The Review of Economic Studies*, **76**, 1149–1174.
- Ruggles, S., Cleveland, L. L., Lovatón Dávila, R., Sarkar, S., Sobek, S., Burk, D., Ehrlich, D. E., Heimann, Q., Lee, J. and Merrill, N. (2025) Data from: Ipums international: Version 7.6. <https://doi.org/10.18128/D020.V7.6>.
- Ruggles, S., McCaa, R., Sobek, M. and Cleveland, L. (2015) The ipums collaboration: integrating and disseminating the world's population microdata. *Journal of Demographic Economics*, **81**, 203–216.
- United Nations, Department of Economic and Social Affairs, Population Division (2024) World population prospects 2024, online edition.

A | APPENDIX

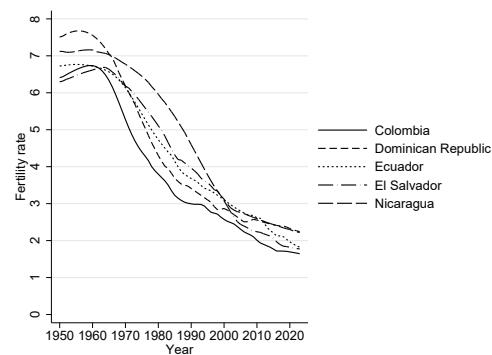
TABLE A.1 Fertility rate by decades

Year:	1960	1970	1980	1990	2000	2010	Var % 2010-1960
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Colombia	6.3	4.5	3.3	2.9	2.3	1.8	-71.1
Dominican Republic	7.1	5.3	3.8	3.1	2.7	2.5	-65.3
Ecuador	6.5	5.5	4.2	3.4	2.8	2.3	-64.9
El Salvador	6.6	5.7	4.5	3.6	2.6	2.1	-68.7
Nicaragua	7.0	6.5	5.4	3.9	2.8	2.5	-65.1
Latin America and the Caribbean (UN)	5.7	4.7	3.7	2.9	2.4	2.1	-63.4
High-income countries	2.8	2.2	1.9	1.7	1.7	1.7	-41.2
World	4.9	4.2	3.6	2.9	2.7	2.5	-48.7

Notes: The table reports fertility rates for Latin American countries, high-income countries, and worldwide, as well as for the selected Latin American countries, across each decade from 1960 to 2010 (Columns 1–6). Column 7 shows the percentage change between 1960 and 2010. The fertility rate refers to the average number of live births a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates of a given period and if they were not subject to mortality. It is expressed as live births per woman. Bold numbers represent country-decades that we observe in our census data. *Source:* Data from [United Nations, Department of Economic and Social Affairs, Population Division \(2024\)](#).



(a) Regions



(b) LAC countries

FIGURE A.1 Fertility rate by regions and LAC countries. *Notes:* The graphs show the evolution of the fertility rate in Latin American countries, in high-income countries, in worldwide (Panel A), and in the selected Latin American countries (Panel B). The fertility rate refers to the average number of live births a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates of a given period and if they were not subject to mortality. It is expressed as live births per woman. *Source:* Data from [United Nations, Department of Economic and Social Affairs, Population Division \(2024\)](#).

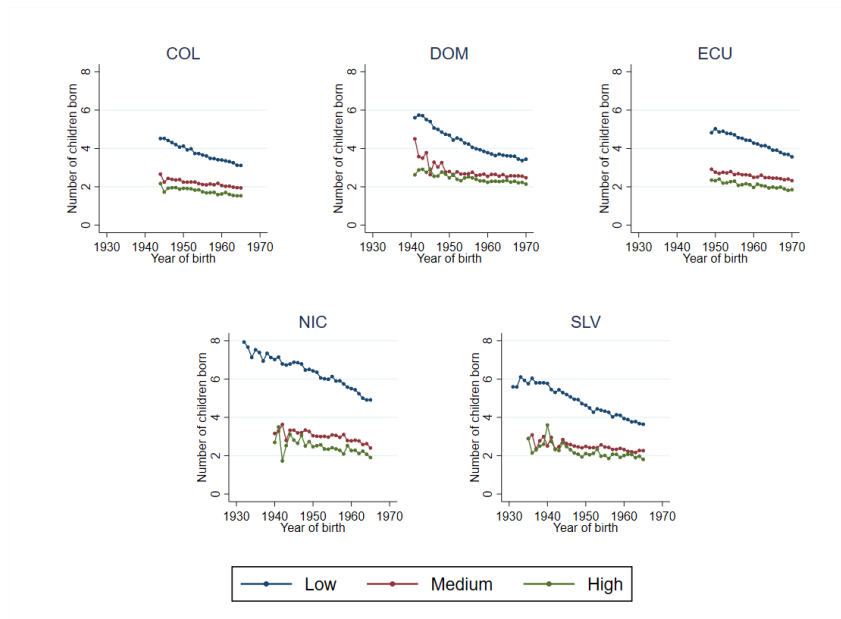
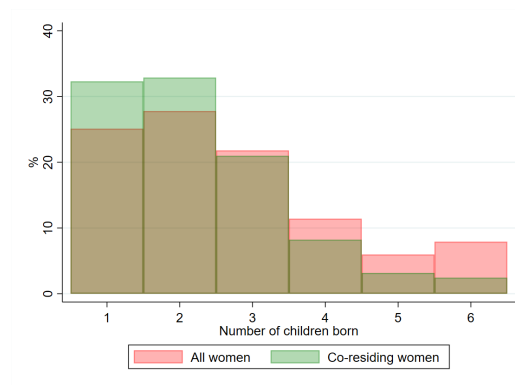
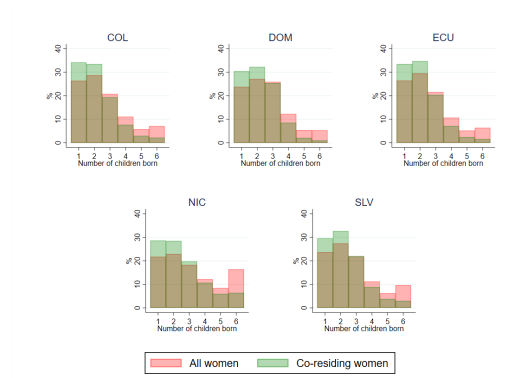


FIGURE A.2 Number of surviving children born to women between 40 and 60 years old at census date, by education level. *Notes:* The graphs show the total number of children born to women between 40 and 60 years old at the time of the census (Colombia 2005, Ecuador 2010, El Salvador 1992 and 2007, Nicaragua 1995 and 2005, and Dominican Republic 2002 and 2010) by country and by year of birth and educational attainment. “Low” education refers to at most primary education complete or lower; “Medium” education refers to secondary education complete; “High” education refers to university complete. *Source:* Data from IPUMS International samples.

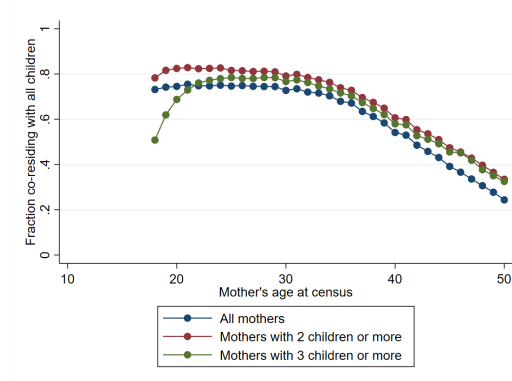


(a) Pool

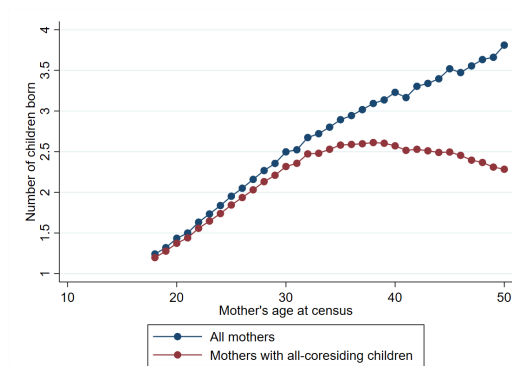


(b) By country

FIGURE A.3 Notes: The graphs show the distribution of the number of surviving children born to women between 18 and 50 years old at the census date, by co-residence status (Colombia 2005, Ecuador 2010, El Salvador 1992 and 2007, Nicaragua 1995 and 2005, and Dominican Republic 2002 and 2010). Panel A presents the pooled distribution across all observations, while Panel B displays the distribution separately for each of the selected countries. Number of children above 6 are collapsed at the value 6. Co-residing families correspond to those where all children live with the mother. Source: Data from IPUMS International samples.



(a) Fraction of mothers living with all their children



(b) Number of children by co-residence status

FIGURE A.4 Co-residence rate and number of children born to women between 18 and 50 years old at census date. *Notes:* The graphs show the fraction of mothers that co-resides with all their children (Panel A) and the average number of surviving children born to women by co-residence status (Panel B) by mother's age at census (Colombia 2005, Ecuador 2010, El Salvador 1992 and 2007, Nicaragua 1995 and 2005, and Dominican Republic 2002 and 2010). *Source:* Data from IPUMS International samples.

TABLE A.2 Descriptive statistics of children in families that have twins in the *n*th birth

	2+ Sample	3+ Sample	
	First-born (1)	First-born (2)	Second-born (3)
Age (years)	11.9	12.8	11.6
Female (%)	49.2	48.5	48.1
Birth month	6.8	6.5	6.6
Age difference with <i>n</i> th birth (years)	4.0	6.8	4.2
Number of children	3.5	4.4	4.4
3 children or more (%)	-	-	-
4 children or more (%)	33.1	-	-
Father or mother's spouse at home (%)	82.6	84.9	83.4
Father at home (%)	81.1	82.4	82.0
Mother's age at first child (years)	22.3	20.9	21.6
Mother's age at census (years)	34.2	33.7	35.5
Years of schooling (mother)	8.2	6.7	6.8
Primary complete (mother) (%)	52.3	63.1	63.5
Secondary complete (mother) (%)	35.1	29.6	28.3
University complete (mother) (%)	12.6	7.3	8.2
Mother's health (% have not an alive child)	5.7	6.1	6.7
Years of schooling (father)	7.6	6.6	6.7
Primary complete (father) (%)	55.7	64.7	65.3
Secondary complete (father) (%)	30.8	26.5	25.0
University complete (father) (%)	13.5	8.8	9.7
Primary complete (%) (*)	84.8	78.3	76.0
Secondary complete (%) (*)	42.9	38.6	41.1
Attending school (%)	87.9	83.6	85.4
Behind 2 years of schooling or more (%)	21.5	33.5	30.4
Behind 1 year of schooling (%)	20.9	17.0	20.6
Not behind (%)	57.6	49.5	49.0
Labor force participation (%)	16.8	25.2	21.8
Population of district of residence (in logs)	12.0	11.9	11.9
Colombia 2005 (%)	44.8	40.1	40.7
Dominican Republic 2002 (%)	11.4	13.4	12.4
Dominican Republic 2010 (%)	12.0	12.4	12.0
Ecuador 2010 (%)	14.0	11.2	12.2
Nicaragua 1995 (%)	3.2	5.8	5.7
Nicaragua 2005 (%)	4.4	5.5	5.4
El Salvador 1992 (%)	4.1	4.6	4.6
El Salvador 2007 (%)	6.1	7.0	7.1
Observations	2,606	1,595	1,625

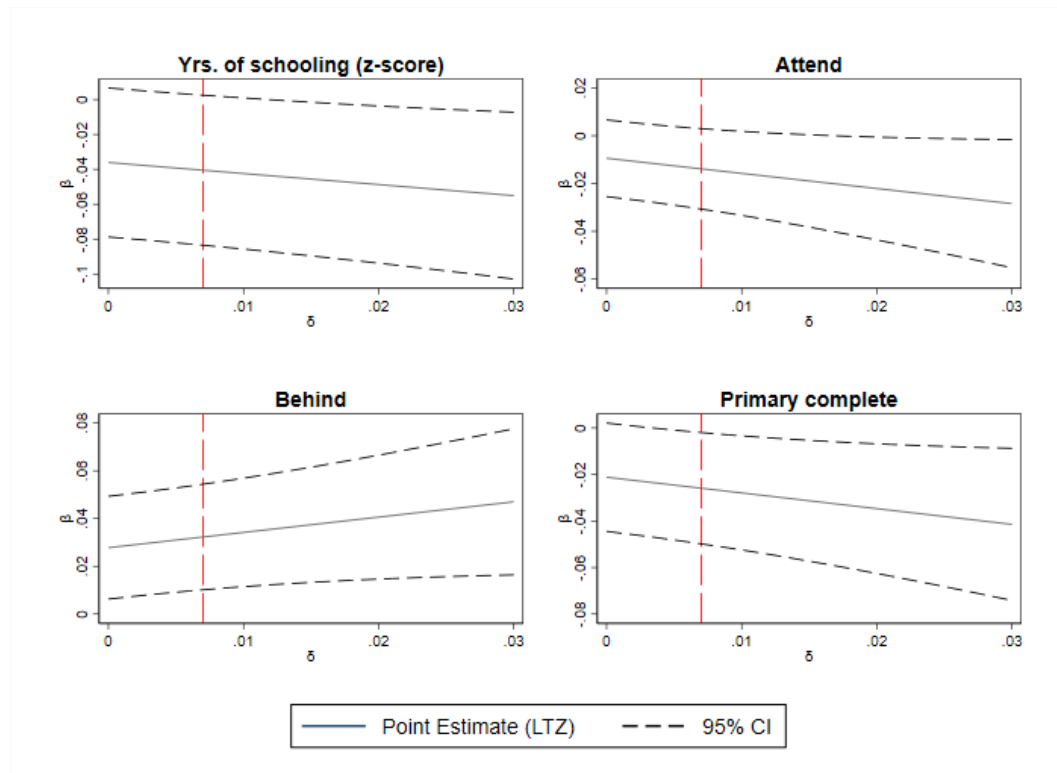
Notes: The table presents the mean values of selected characteristics for two subsamples. Column 1 shows the “2+ sample”, which includes the first non-twin child in families with two or more births. Columns 2 and 3 show the “3+ sample”, focusing on the first and second non-twin children, respectively, in families with three or more births. The two subsamples are restricted to children from families that experienced a twin birth in the *n*th birth. (*) The length of primary and secondary education differ by country. We use these country specific cut-offs to compute primary and secondary education completion on the population who is above this age cutt-offs.

Source: Data from IPUMS International samples.

TABLE A.3 Probability of giving birth to twins

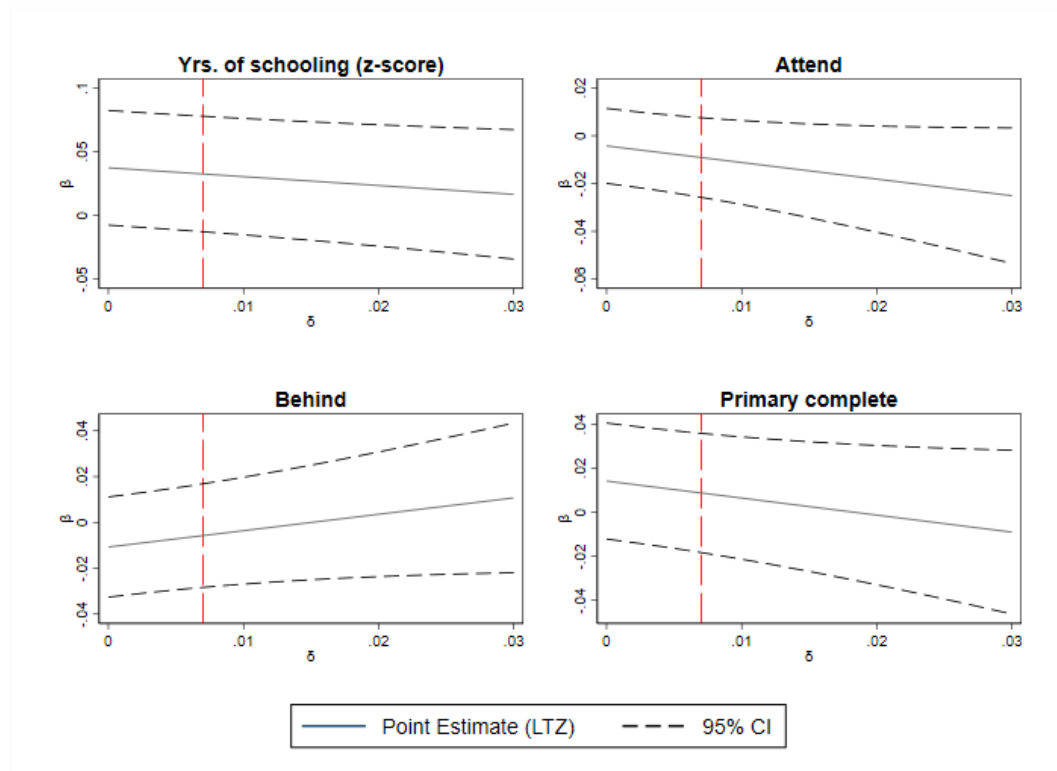
Dependent variable:	Twins
Mother Health (Mortality)	-0.001** (0.000)
Mother Years School = 1	0.002** (0.001)
Mother Years School = 2	0.002*** (0.001)
Mother Years School = 3	0.001** (0.001)
Mother Years School = 4	0.001 (0.001)
Mother Years School = 5	0.001* (0.000)
Mother Years School = 6	0.001** (0.000)
Mother Years School = 7	-0.000 (0.001)
Mother Years School = 8	0.001* (0.001)
Mother Years School = 9	0.001* (0.001)
Mother Years School = 10	0.001 (0.001)
Mother Years School = 11	0.002*** (0.001)
Mother Years School = 12	0.001** (0.001)
Mother Years School = 13	0.002** (0.001)
Mother Years School = 14	0.001* (0.001)
Mother Years School = 15	0.001 (0.001)
Mother Years School = 16	0.001 (0.001)
Mother Years School = 17	0.001 (0.001)
Mother Years School = 18	0.003*** (0.001)
Observations	615,411
Mean	0.007

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows marginal effects of a probit regression where the dependent variable takes value one if the mother had twins. The estimation sample considers all mothers in the "2+ sample" and "3+ sample". To test positive selection into twins, marginal effects for mother's education and health measures are reported. The regression includes as control variables the mother's age at the census fixed effects, mother's age at her first birth fixed effects, district of residence's population (log and log square) and census fixed effects. Robust standard errors are reported in parentheses. Source: Data from IPUMS International samples.

FIGURE A.5 Conley et al. (2012) estimates. Sensitivity analysis. 3+ sample

Note: Each panel shows sensitivity bounds for the effect of family size on each educational outcome for the "3+ sample". Estimates are computed following Conley et al. (2012) using the `plausexog` command in Stata. The horizontal axis shows values of the assumed degree of violation of the exclusion restriction (δ), while the vertical axis displays the corresponding estimated coefficient (β). The red vertical dashed line at $\delta = 0.007$ corresponds to the assumed deviation used to construct bounds in Figure 1. The shaded area represents the 95% confidence interval. The instrument used is an indicator for being a twin, and the endogenous regressor is the residualized total number of siblings from a regression including the controls, following Equation 3. *Source:* Data from IPUMS International samples.

FIGURE A.6 Conley et al. (2012) estimates. Sensitivity analysis. 2+ sample



Note: Each panel shows sensitivity bounds for the effect of family size on each educational outcome for the "2+ sample". Estimates are computed following Conley et al. (2012) using the `plausexog` command in Stata. The horizontal axis shows values of the assumed degree of violation of the exclusion restriction (δ), while the vertical axis displays the corresponding estimated coefficient (β). The red vertical dashed line at $\delta = 0.007$ corresponds to the assumed deviation used to construct bounds in Figure 1. The shaded area represents the 95% confidence interval. The instrument used is an indicator for being a twin, and the endogenous regressor is the residualized total number of siblings from a regression including the controls, following Equation 3. *Source:* Data from IPUMS International samples.

TABLE A.4 IV estimates for the number of adults in the household

Dependent variable:	Number of adults	
	OLS	IV
	(1)	(2)
<i>Panel A: Sample 2+</i>		
Number of children	0.026*** (0.001)	-0.020 (0.017)
Observations	377,328	377,328
Mean	2.012	2.012
<i>Panel B: Sample 3+</i>		
Number of children	0.019*** (0.001)	-0.002 (0.018)
Observations	416,258	416,258
Mean	2.016	2.016

Notes: The table shows the OLS and twins IV estimates of the coefficient β_1 of Equation 1 for the number of adults at home for the “2+ sample” (Panel A) and “3+ sample” (Panel B). The “2+ sample” comprises first-born non-twins aged 6-18, while the “3+ sample” includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

TABLE A.5 Twins IV estimates for the 3+ sample with and without controlling for ethnic group of the mother

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First-stage	0.785*** (0.023)	0.791*** (0.023)	0.784*** (0.023)	0.790*** (0.023)	0.778*** (0.025)	0.786*** (0.025)	0.735*** (0.030)	0.744*** (0.030)
Number of children	-0.043 (0.029)	-0.045 (0.029)	-0.011 (0.011)	-0.012 (0.011)	0.031** (0.014)	0.031** (0.014)	-0.014 (0.014)	-0.015 (0.014)
Observations	260,845	260,845	260,014	260,014	236,302	236,302	159,630	159,630
Mean	0.018	0.018	0.853	0.853	0.294	0.294	0.818	0.818
Race control	NO	YES	NO	YES	NO	YES	NO	YES

Notes: The estimations presented in the table follow the specifications of Table 2. The only difference for the right columns in each outcome is the addition of the mother's race as a control. This variable is constructed based on the information reported in IPUMS gathering the all the race and ethnic categories in four groups: white or mestizo, afro-descendant, indigenous and other races. The former one is the omitted category in the regression. The sample is smaller than our total sample, and it is composed by the censuses of Colombia (2005), Ecuador (2010) and El Salvador (2007) for which IPUMS reports the race of the individuals. The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the "3+ sample", standard errors are clustered by mother's ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

TABLE A.6 First Stages for Standard Non-Linear IV Estimates (2+ sample)

Instrument:	Children ≥ 3	Children ≥ 4	Children ≥ 5	Children ≥ 6
Twins ₂	0.511*** (0.009)	0.170*** (0.007)	0.039*** (0.005)	0.008*** (0.003)
Twins ₃ *	0.009 (0.010)	0.621*** (0.008)	0.141*** (0.006)	0.044*** (0.003)
Twins ₄ *	0.010 (0.016)	0.021 (0.014)	0.668*** (0.009)	0.102*** (0.005)
Twins ₅ *	0.017 (0.030)	0.017 (0.026)	0.026 (0.017)	0.718*** (0.010)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each column reports the first stage estimate of number of children at each parity on twin births from the standard IV regression displayed in Table 9 for the outcome "Years of schooling (z-score)". All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regression. *Source:* Data from IPUMS International samples.

TABLE A.7 First Stages for Standard Non-Linear IV Estimates (3+ sample)

Instrument:	Children ≥ 4	Children ≥ 5	Children ≥ 6
Twins ₃	0.656*** (0.008)	0.122*** (0.006)	0.039*** (0.003)
Twins ₄ *	0.019* (0.011)	0.659*** (0.008)	0.093*** (0.005)
Twins ₅ *	0.020 (0.020)	0.029* (0.015)	0.724*** (0.009)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each column reports the first stage estimate of number of children at each parity on twin births from the standard IV regression displayed in Table 10 for the outcome “Years of schooling (z-score)”. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regression. *Source:* Data from IPUMS International samples.

TABLE A.8 Theoretical entrance ages to secondary education

Census	Secondary Entrance Age	Source
COL-2005	11	https://www.epdc.org/sites/default/files/documents/EPDC_NEP_2018_Colombia.pdf
DOM-2002	12	https://www.epdc.org/sites/default/files/documents/EPDC_NEP_2018_DominicanRepublic.pdf
DOM-2010	12	https://www.epdc.org/sites/default/files/documents/EPDC_NEP_2018_DominicanRepublic.pdf
ECU-2010	12	https://www.epdc.org/sites/default/files/documents/EPDC_NEP_2018_Ecuador.pdf
NIC-1995	13	https://unesdoc.unesco.org/ark:/48223/pf0000155583
NIC-2005	13	https://unesdoc.unesco.org/ark:/48223/pf0000155583
SLV-1992	13	https://www.epdc.org/sites/default/files/documents/EPDC_NEP_2018_ElSalvador.pdf
SLV-2007	13	https://www.epdc.org/sites/default/files/documents/EPDC_NEP_2018_ElSalvador.pdf

Notes: This table presents the theoretical entrance ages to secondary education in each country-year. *Source:* Data from Policy and Center (2018) and Porta and Laguna (2007).

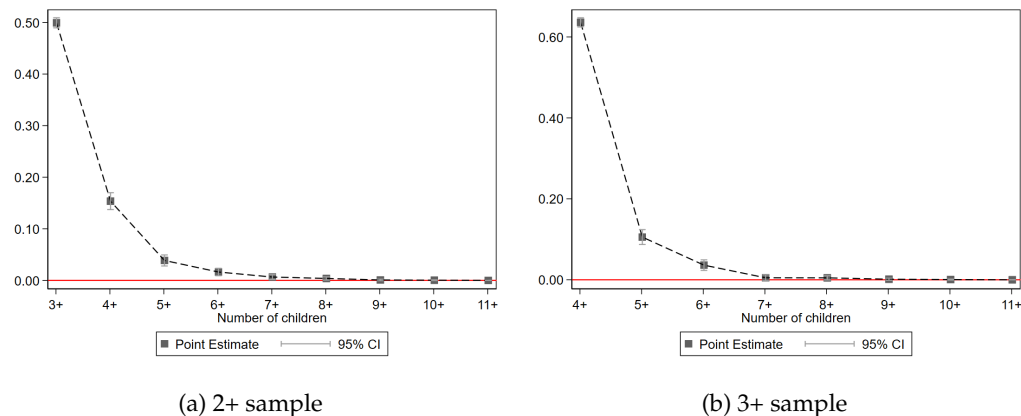


FIGURE A.7 Linear IV Weights. Notes: These panels show the impact of having twins at a specific birth order on the likelihood of exceeding a given family size (indicated on the x-axis), for the “2+ sample” and the “3+ sample”, respectively. The coefficients represent the weights assigned to each parity in the LATEs estimated using twins at a particular birth order as an IV. Point estimates are shown along with their 95% confidence intervals. *Source:* Data from IPUMS International samples.

TABLE A.10 IV estimates by gender of the child

Dependent variable:	Yrs. School. (z-score)		Attend		Behind		Primary	
	Female	Male	Female	Male	Female	Male	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: Sample 2+</i>								
First-stage	0.719*** (0.022)	0.719*** (0.022)	0.720*** (0.022)	0.718*** (0.022)	0.705*** (0.024)	0.692*** (0.024)	0.655*** (0.033)	0.638*** (0.034)
Number of children	0.067** (0.032)	0.007 (0.033)	-0.005 (0.011)	-0.005 (0.011)	-0.021 (0.015)	-0.001 (0.016)	0.026 (0.018)	0.001 (0.020)
Observations	181,839	195,489	181,462	195,093	158,925	171,084	100,175	109,278
Mean	0.254	0.104	0.886	0.857	0.208	0.283	0.851	0.802
<i>Panel B: Sample 3+</i>								
First-stage	0.763*** (0.022)	0.813*** (0.023)	0.762*** (0.022)	0.812*** (0.023)	0.752*** (0.024)	0.806*** (0.025)	0.701*** (0.031)	0.770*** (0.032)
Number of children	-0.031 (0.030)	-0.038 (0.028)	-0.004 (0.011)	-0.014 (0.011)	0.027* (0.015)	0.027* (0.014)	-0.026 (0.018)	-0.018 (0.015)
Observations	200,502	215,756	199,965	215,176	178,015	192,157	112,967	123,795
Mean	0.150	-0.024	0.864	0.831	0.261	0.348	0.812	0.750

Notes: The table shows the twins IV estimates of the coefficient β_1 of Equation 1 by gender of the child. The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). The “2+ sample” comprises first-born non-twins aged 6-18, while the “3+ sample” includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). Years of schooling are standardized as z-scores within country-year and by children’s age. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

TABLE A.11 IV estimates for Behind by gender and age of the child

Dependent variable:	Behind			
	10-14		15-18	
	Female (1)	Male (2)	Female (3)	Male (4)
<i>Panel A: Sample 2+</i>				
First-stage	0.745*** (0.034)	0.703*** (0.035)	0.590*** (0.049)	0.589*** (0.049)
Number of children	-0.034* (0.020)	-0.005 (0.023)	0.002 (0.038)	-0.005 (0.036)
Observations	78,851	82,738	49,738	56,473
Mean	0.211	0.275	0.267	0.383
<i>Panel B: Sample 3+</i>				
First-stage	0.765*** (0.030)	0.805*** (0.032)	0.660*** (0.044)	0.751*** (0.043)
Number of children	0.018 (0.020)	0.050** (0.020)	0.025 (0.034)	-0.003 (0.026)
Observations	91,013	95,651	55,414	63,353
Mean	0.270	0.347	0.318	0.444

Notes: The table shows the twins IV estimates of the coefficient β_1 of Equation 1 by gender and age of the child for the “2+ sample” (Panel A) and “3+ sample” (Panel B). The table also shows the first-stage estimates for each specification (coefficient π_1 of Equation 2). The “2+ sample” comprises first-born nontwins aged 15-18, while the “3+ sample” includes first and second-born nontwins of the same age range (refer to Section 2.3 for further details). The outcome “Behind” measures the probability of being behind at least 2 years in years of schooling compared with the mode of schooling among children of the same census-age-birth quarter. The regression for “Behind” is estimated for individuals in cohorts where the mode of schooling is two or more. This table focuses specifically on children aged 10 to 18. All regressions include all control variables detailed in the Section 3. Robust standard errors are reported in parentheses. For the “3+ sample”, standard errors are clustered by mother’s ID. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Source:* Data from IPUMS International samples.

B | EFFICIENT INSTRUMENTS

The inclusion of many endogenous variables in the unrestricted model requires the construction of this set of instruments that challenges the estimation of the parameters since it augments the imprecision of IV estimates. [Mogstad and Wiswall \(2016\)](#) offer a solution introducing the notion of efficient instruments in the family-size context. Relying on the typical assumptions for the linear IV validity, it is possible to use any function of twin_{ci} and $\mathbf{X}$ to obtain valid instruments. Whereas the optimal instrument is an unknown function of twin_{ci} and $\mathbf{X}$, the IV estimates would still be consistent under a misspecification of the functional form as long as the typical assumptions of mean-independence of the error respect to covariates and the exclusion restriction holds.

Previous works that implemented efficient instruments had found a substantial improvement in the efficiency of IV estimates.²⁰ [Mogstad and Wiswall \(2016\)](#) specifically propose a functional form for these instruments considering the particular features of family size application that we follow here. Their specification considers the predicted probability of having $n+1$ or more siblings to build efficient instruments for the from n to $n+1$ sibling margin. To be concrete, following with the case 2+ sample, to obtain the efficient instrument for moving from 1 to 2 siblings we define the predicted probability of having 2 or more siblings (equivalent to having 3 or more children) in the household, as:

$$\hat{p}_{2i} = \begin{cases} 1, & \text{if } \text{twin}_{2i} = 1, \\ f_2(\mathbf{X}_i, \hat{\theta}_2), & \text{if } \text{twin}_{2i} = 0. \end{cases} \quad (\text{B.1})$$

If the second birth was a twin, then the probability of having 2 or more siblings (i.e. 3 or more children) is obviously 1. In the case where that second birth was singleton the predicted probability will be a nonlinear function of the covariates, restricted between 0 and 1, where $\hat{\theta}_2$ are the unknown parameters.

[Mogstad and Wiswall \(2016\)](#) construct the efficient instrument for the 2 to 3, the 3 to 4, and the 4 to 5 sibling margins as the predicted probability of having 3 or more, 4 or more, and 5 or more siblings as:

$$\hat{p}_{3i} = f_3(\mathbf{X}_i, \text{twin}_{3i}^*, \hat{\theta}_3), \quad (\text{B.2})$$

$$\hat{p}_{4i} = f_4(\mathbf{X}_i, \text{twin}_{4i}^*, \hat{\theta}_4), \quad (\text{B.3})$$

$$\hat{p}_{5i} = f_5(\mathbf{X}_i, \text{twin}_{5i}^*, \text{twin}_{2i}, \hat{\theta}_5), \quad (\text{B.4})$$

Which are defined as a linear function of the same covariates that in $f_2()$ and now, of the constructed instruments from the previous section, because the original twin instrument is a missing instrument for these higher orders.

To estimate these functions we follow the authors, considering a parsimonious specification to avoid any small sample bias in the IV estimation. In concrete, the functions are

²⁰The intuition behind the improvement in the efficiency of IV estimation is that the efficient instrument is more strongly correlated with endogenous variables of family size, because of the use of a nonlinear model to predict those variables, instead of a linear model used in a standard setting. For more details refer to [Mogstad and Wiswall \(2016\)](#).

estimated using a probit model and we include as covariates the child's age (linear and quadratic term), child's sex, mom's age and mom's age at firstborn, dummies for each year of parent's education, a dummy for mortality and country fixed effects. Once we obtain the efficient instrument we replace them in the first stage specification. Following with the example for 2+ sample:

$$Nchild_{sih} = \lambda_{s2}\hat{p}_{2ih} + \lambda_{s3}\hat{p}_{3ih} + \lambda_{s4}\hat{p}_{4ih} + \lambda_{s5}\hat{p}_{5ih} + X_i'\lambda_x + \eta_i, \quad s = 2...5 \quad (B.5)$$

| Results

For the sample of children with at least two siblings (2+ sample), the nonlinear estimates suggest a non-monotonic relationship between family size and educational attainment (see Table B.1). While the marginal effect of moving from two to three children (Children ≥ 3) is positive but statistically not significant (0.046), this pattern reverses at the fourth child. Specifically, the effect of having four or more children (Children ≥ 4) turns negative and highly significant (-0.355), increasing in magnitude for five children (-0.663) and then attenuating at six or more (-0.263). This pattern suggests a turning point in the marginal effect, where initial increases in family size may reflect complementarities or sibling support, while larger family sizes likely exacerbate constraints on parental resources or time. These results nuance the linear IV estimate showing that the average effect masks considerable heterogeneity across the distribution of family size.

For the sample restricted to children with at least two older siblings (3+ sample), the evidence of nonlinear effects is even more pronounced (see Table B.2). The marginal effect of going from three to four children is already negative (-0.066), and this effect increases sharply with each additional child. Specifically, significant reductions in years of schooling are observed when moving from four to five (-0.539) and from five to six children (-0.602), both with strong statistical significance. This declining pattern is consistent with increasing pressure on household resources and diminishing marginal returns to educational outcomes. Unlike the 2+ sample, no positive effects are observed at lower margins, and the linear evidence already points to an average dominated by the negative effects of larger family sizes.

Taken together, the results underscore the importance of allowing for nonlinearities in the relationship between family size and educational outcomes. While linear IV estimates suggest a weakly negative or even mildly positive effect in the 2+ sample, the nonlinear models reveal highly heterogeneous effects: potentially positive at lower margins, but strongly negative beyond a certain threshold (four or more children). For the 3+ sample the average estimate masks much stronger penalties associated with high-parity births. In this context, the linear specification substantially underestimates the schooling loss for children in large families.

TABLE B.1 Efficient IV estimates for the unrestricted model of family size for 2+ sample

Dependent variable:	Yrs. School. (z-score)	Attend	Behind	Primary
	(1)	(2)	(3)	(4)
Children ≥ 3	0.046 (0.030)	0.018* (0.011)	-0.037** (0.016)	0.052*** (0.016)
Children ≥ 4	-0.355*** (0.038)	0.005 (0.011)	0.167*** (0.022)	-0.014 (0.017)
Children ≥ 5	-0.663*** (0.073)	-0.299*** (0.026)	0.261*** (0.027)	-0.025 (0.033)
Children ≥ 6	-0.263** (0.116)	-0.372*** (0.055)	0.125** (0.055)	-0.031 (0.041)
Joint Significance (p-value)	0.000	0.000	0.000	0.006
Observations	372,573	371,840	325,256	205,030
Mean	0.187	0.875	0.240	0.832

Notes: The table shows the Efficient IV estimates for the unrestricted model of family size for the “2+ sample”. Children ≥ 3 is the marginal effect for moving from 2 to 3 children, Children ≥ 4 of moving from 3 to 4, and so on. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regressions. The p-value for the Joint Significance test refers to the null hypothesis that each parameter is zero. *Source:* Data from IPUMS International samples.

TABLE B.2 Efficient IV estimates for the unrestricted model of family size for 3+ sample

Dependent variable:	Yrs. School. (z-score)	Attend	Behind	Primary
	(1)	(2)	(3)	(4)
Children ≥ 4	-0.066** (0.030)	0.010 (0.010)	0.046*** (0.014)	0.003 (0.018)
Children ≥ 5	-0.539*** (0.056)	-0.244*** (0.030)	0.192*** (0.029)	0.009 (0.025)
Children ≥ 6	-0.602*** (0.091)	-0.341*** (0.047)	0.214*** (0.042)	-0.039 (0.047)
Joint Significance (p-value)	0.000	0.000	0.000	0.862
Observations	405,310	404,284	359,253	227,237
Mean	0.080	0.854	0.295	0.792

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table shows the Efficient IV estimates for the unrestricted model of family size for the “3+ sample”. Children ≥ 4 is the marginal effect for moving from 3 to 4 children, Children ≥ 5 of moving from 4 to 5, and so on. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regressions. The p-value for the Joint Significance test refers to the null hypothesis that each parameter is zero. Source: Data from IPUMS International samples.

TABLE B.3 First Stages for Efficient Non-Linear IV Estimates (2+ sample)

Instrument:	Children ≥ 3	Children ≥ 4	Children ≥ 5	Children ≥ 6
$\hat{p}_2$	1.044*** (0.016)	0.304*** (0.013)	0.065*** (0.009)	0.011** (0.005)
$\hat{p}_3$	-0.008 (0.014)	0.845*** (0.012)	0.081*** (0.008)	0.010** (0.005)
$\hat{p}_4$	-0.001 (0.021)	0.203*** (0.017)	0.882*** (0.012)	0.044*** (0.007)
$\hat{p}_5$	-0.065* (0.039)	0.240*** (0.033)	0.508*** (0.022)	1.103*** (0.013)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each column reports the first stage estimate of number of children at each parity on twin births from the efficient IV regression displayed in Table B.1 for the outcome “Years of schooling (z-score)”. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regression. Source: Data from IPUMS International samples.

TABLE B.4 First Stages for Efficient Non-Linear IV Estimates (3+ sample)

Instrument:	Children ≥ 4	Children ≥ 5	Children ≥ 6
$\hat{p}_3$	0.998*** (0.012)	0.106*** (0.009)	-0.010* (0.005)
$\hat{p}_4$	0.105*** (0.015)	0.899*** (0.011)	0.062*** (0.007)
$\hat{p}_5$	0.128*** (0.025)	0.445*** (0.019)	1.079*** (0.011)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each column reports the first stage estimate of number of children at each parity on twin births from the efficient IV regression displayed in Table B.2 for the outcome “Years of schooling (z-score)”. All regressions include all control variables detailed in the Section 3. Standard errors are calculated using a cluster bootstrap, sampling each family with replacement 25 times. For each bootstrap repetition, we re-estimate the instruments and the regression. *Source:* Data from IPUMS International samples.