Bratti, Massimiliano; Fiore, Simona; Mendola, Mariapia

Working Paper

The Impact of Family Size and Sibling Structure on the Great Mexico-U.S. Migration

GLO Discussion Paper, No. 392

Provided in Cooperation with:
Global Labor Organization (GLO)

The Impact of Family Size and Sibling Structure on the Great Mexico–U.S. Migration

Massimiliano Bratti* Simona Fiore† Mariapia Mendola‡

September 1, 2019

Abstract

We investigate how fertility and demographic factors affect migration at the household level by assessing the causal effects of sibship size and structure on offspring's international migration. We use a rich demographic survey on the population of Mexico and exploit presumably exogenous variation in family size induced by biological fertility and infertility shocks. We further exploit cross-sibling differences to identify birth order, sibling-sex, and sibling-age composition effects on migration. We find that large families *per se* do not boost offspring out-migration. Yet, the likelihood of migrating is not equally distributed within a household, but is higher for sons and decreases sharply with birth order. The female migration disadvantage also varies with sibling composition by age and gender.
Keywords: International Migration, Mexico, Family Size, Sibling Structure.
JEL codes: J13 F22 O15.

*Università degli Studi di Milano (Milan, Italy); Global Labor Organization (GLO); Institute of Labor Economics (IZA, Bonn); Centro Studi Luca d'Agliano (LdA, Milan); massimiliano.bratti@unimi.it

†Università degli Studi di Bologna (Bologna, Italy); Istituto Carlo Cattaneo (Bologna, Italy), s.fiore@unibo.it

‡Università degli Studi di Milano–Bicocca (Milan, Italy); IZA (Bonn, Germany), mariapia.mendola@unimib.it

1 Introduction

Demographic pressure is generally recognized to be an important driver of international migration flows. Previous studies have shown that by increasing population cohort sizes, high fertility boosts emigration (Hanson and McIntosh, 2010, 2016). The main mechanism addressed in the literature is shocks to labor supply, which deteriorate domestic labor market opportunities and increase the incentive for migrating abroad (Borjas, 1987; Hanson, 2004). These studies focus on the effect of high demographic pressure on the returns to migration, i.e. on general equilibrium effects in the labor market. However, migration is a costly investment, necessarily borne by individuals and their families upfront (e.g. Angelucci, 2015). Hence, high fertility, which translates into a large number of siblings, may alter a household's ability to pay for this investment. In this paper, we focus on this mechanism and investigate the extent to which family size and composition affect the decision to have one or more migrant(s) at the household level. To this end, we employ a household-level approach and assess whether sibship size has a causal impact on the individual probability to migrate abroad. We further assess whether sibling structure (i.e. birth order and sibling composition by age and gender) play any role in the migration decision within a household.

Our paper complements previous studies that, in using an aggregate (i.e. cohort-level) approach, are likely to confound the effect of high fertility (and larger cohort sizes) on returns to migration (general equilibrium effects) with its effect on the ability of households to invest in migration (household-level effect). Disentangling the two effects is important in order to better understand the implications of population growth for migration. Indeed, while a decline in fertility rates, and thus less pressure on the labor market, is likely to reduce migration, it can simultaneously change the allocation of resources among household members.[1]

High fertility may decrease migration at the household level by diluting the per capita

[1] The degree to which household structure, rather than the performance of the aggregate economy, influences intra-family resource allocation in developing contexts has been documented by several seminal studies such as Rosenzweig (1988) and Rosenzweig and Stark (1989).

resources available to pay for it or because more family-work is needed at home, e.g., care for younger children (e.g. Becker and Lewis, 1973). On the other hand, in order to keep pace, large families may spur investment in the form of migrating household members. Importantly, given the costly and selective nature of international migration, household demographics may determine *who* is enabled to leave within the family (e.g. Chen, 2006). This is because the returns to migration (and perhaps some of the costs) accrue over a period of time and may depend on a child's own characteristics as well as those of his / her siblings. Indeed, in the context of limited resources and high returns to the migration investment, siblings may become rivals and some children (e.g., girls, or later borns) may have fewer economic opportunities than their siblings (Garg and Morduch, 1998; Black et al., 2005; Jayachandran and Kuziemko, 2011).[2]

Despite great interest in the determinants of international migration and growing concerns about the role of high fertility and demography in spurring migration flows—e.g., from Africa to Europe today—there is little evidence on the extent to which family size and composition affect offspring's emigration choices.[3] This is a significant gap given that migrants are typically young and gendered, come from high-fertility countries, and leave behind household members who oftentimes are siblings (Hatton and Williamson, 2003).[4] To date, the literature has mainly focused on the determinants of family migration investigating network effects (e.g. Winters et al., 2001), the effect of the number of children on parental migration (e.g. Lindstrom and Saucedo, 2007; Sarma and Parinduri, 2015) or the effect of migration on fertility (Mayer and Riphahn, 2000; Lindstrom and Saucedo, 2002). There are very few stud-

[2] A well-established theoretical literature in economics rationalizes a causal link between children's economic resources and their lifetime opportunities and adult outcomes (Becker and Tomes, 1976; Thomas, 1990).

[3] Throughout the paper we use "family size" (i.e., number of children) and "sibship size," (i.e., number of siblings) interchangeably, the former from the perspective of the parents, the latter from that of the children.

[4] The majority of migrants are young adults who are more likely to have a positive net expected return to migration due to their longer remaining life expectancy. According to recent UN figures, international migrants aged 15 to 24 account for 12.5% of total migrants worldwide, and when migrants between the ages of 25 and 34 are also considered, young migrants represent over 30% of the total (UNDESA, 2011). The proportion of young migrants is much higher in developing countries and more than doubles when internal migrants are also considered (UN, 2013).

ies on the potential impact of one childhood's family size on individual migration decisions, and they have generally looked at it while assuming exogenous family sizes. Abramitzky et al. (2013), for instance, examine the role of inheritance norms in the Norwegian context during the Age of Mass Migration (1850-1913), showing how the number of brothers (as opposed to sisters) positively affects the likelihood of (any type of) sons' emigration, but only in land owning families. Stöhr (2015) focuses on siblings' interactions in migration decisions (namely on the effect of the number of siblings who migrate), especially in relation to elderly care provision of those left behind, while controlling for the number of daughters and sons. A relatively large literature, though, has empirically investigated the role of family size, birth order and sibling composition (by age and gender) on household investments in other forms of children's human capital such as health and education (e.g. Black et al., 2005; Jayachandran and Kuziemko, 2011; Jayachandran and Pande, 2017). In general findings point to little role of family size on children's outcomes, while sibling structure and composition have more significant effects on offspring human capital investment. To the best of our knowledge, we provide the first empirical assessment of the impact of demographic characteristics of the origin household, i.e. sibship size and structure, on international migration choices.

We address this question in the context of the Great Mexico-U.S. migration over the last two decades of the 20th century. Mexico is a highly populated country that, during a demographic boom, experienced a period of mass migration to the U.S., which gradually weakened over the years. As reported in several studies using decennial U.S. Census throughout the 20th century, Mexico-U.S. migration swelled in the 1970s and continued to grow in the 1980s and 1990s, ranging from 5.2% of Mexico's national population in 1990 to a peak of 10.2% in 2005 (Hanson and McIntosh, 2010).[5] Importantly, emigration patterns differed

[5] To put the Great Mexican migration in historical perspective, it is worth noting that "as a share of Mexico's national population, the number of Mexican immigrants living in the U.S. remained at 1.5% from 1960 to 1970, before rising to 3.3% in 1980, 5.2% in 1990, and 10.2% in 2005" (see Hanson and McIntosh, 2010, p.1). In terms of absolute numbers, it is estimated that about 7 million Mexican immigrants entered the U.S. in the 1990s, 2.2 million did it legally while 4.8 million entered illegally (see Borjas and Katz, 2007; Card and Lewis, 2007). As a result, the Mexican-born population residing in the United States in 2000 was nearly 9.2 million, accounting for one third of the U.S. foreign-born population. Hence, the Mexican-born population of the late 20th Century appears historically unprecedented, being both numerically and

by age and gender, with a dominant share of young males.[6] In 1970, Mexico's fertility rate stood at about seven children per woman (Cabrera, 1994). The gradual spread of family planning practices contributed to a fertility transition and by 2005, the number of children per woman had declined to slightly more than two.[7] However, despite abundant evidence on the potentially significant implications of high fertility rates for child investment and economic outcomes, there is no systematic evidence on the impact of family size on migration decisions.

Using two waves of a large and nationally–representative demographic household survey in Mexico, we focus on the determinants of migration for young adults in the age range of 15 to 25. A crucial element in our dataset is the inclusion of information on women's completed fertility history, and hence on the total number of biological siblings ever born into a family, along with individual migration histories. Access to such information is rare, and perhaps unique, in the migration literature, particularly in combination with a large sample of young adults and the availability of sources of plausible exogenous variation in their mothers' fertility. The ideal experiment, in our context, would be to randomly assign women to different levels of fertility and then observe their children's migration outcomes. The reason lies in the potential endogeneity of parental fertility choices — which arises from the fact that families who choose to have more (or fewer) children may also be those who value child out-migration more.[8]

proportionately larger than any other immigrant influx in the former and following century (Passel et al., 2012).

[6] Using population censuses, Hanson and McIntosh (2010) report (in Figure 2) that a significant proportion of men in Mexico starts migrating around age 15, with emigration increasing sharply until approximately age 30 and decreasing thereafter, presumably as a result of return migration. By contrast, there is less youth migration among Mexican women and migration rates are relatively stable over the course of their lives.

[7] In 1974, a new population policy was designed in Mexico, with the aim of reducing population growth and promoting development. The institutional structure established at that time to ensure policy implementation (the National Population Council-CONAPO) has been expanding geographically and socially ever since (Zuniga Herrera, 2008).

[8] Costs and benefits of migration may be unevenly distributed across both families and siblings within family, and hence bias the results. Moreover, unobservable parental preferences for children and old-age support through migration may positively co-vary. Stark (1981) and Williamson (1990), for instance, postulate that heterogeneity in parental preferences for childbearing and for migration are systematically related, and in a context such as Mexico where migration cum remittances is an essential lifeline to households of origin, they are generally positively related.

Formal identification is achieved by leveraging presumably exogenous variation in women's biological fertility (e.g., miscarriage at first pregnancy) and infertility shocks (Agüero and Marks, 2008; Miller, 2011). Importantly, general equilibrium effects are controlled for by adding municipality fixed effects in our models (and in some specifications, municipality by year fixed effects), which capture the impact of demographic pressure on population size and aggregate labor supply.

Our large dataset and empirical strategy allow us to overcome the problem of separately identifying family size and birth order effects. We can identify birth order, sibling-sex, and sibling-age composition effects on migration by estimating family fixed-effects models. Here, we exploit differences across siblings within the same family, thereby eliminating concerns that birth order or sibling composition are picking up time-invariant omitted family variables also affecting family size.[9] We are also able to control for mothers' and children's birth cohorts such that birth order effects are only identified from the differential timing of births within the same family.

We find little evidence that high fertility drives migration choices at the household level. The positive correlation between the number of siblings and migration vanishes when the potential endogeneity of sibship size is addressed. Results are robust to several changes in both the estimation sample and the estimation strategy. Yet, the likelihood of migrating is not equally distributed across children within a family. We find that older siblings, especially firstborn sons, are more likely to migrate, while having more sisters than brothers may increase the chances of migration, particularly among females. This may be because investing in low parity children lengthens the period of time over which the family expects to reap the benefits of having a migrant child, hence maximizing net returns (e.g. older siblings 'cater' for the younger siblings). In addition, a son may be more valuable to send abroad as a migrant than a daughter, given unequal labor market opportunities for different genders. Indeed,

[9] Other papers using a similar estimation strategy for different child outcomes include Black et al. (2005), Kantarevic and Mechoulan (2006), Rosenzweig and Zhang (2009), Black et al. (2016) and Jayachandran and Pande (2017), among others.

labor market returns for Mexican men in the U.S. were relatively high in the 1990s (e.g., in the agriculture sector) compared to women, who were more likely to be responsible for chores and family duties at home and also more exposed to (sexual) violence at destination.[10] Overall, our results regarding the effects of sibling composition highlight the importance of the migrant's family of origin in driving migration decisions and are consistent with a household's optimal migration model in which mobility is an investment in the human agent, but private costs and rewards involve both migrants and non-migrant household members (Sjaastad, 1962; Stark, 1991).

These findings have relevant implications for both policy makers and researchers. Many observers highlight the importance of the role of demographic pressure in shaping migration flows from today's developing countries. As argued by Hanson and McIntosh (2016) while discussing the contribution of differentials in population growth to international migration in the long run, "the European immigration context today looks much like the United States did three decades ago (p.2)". The main reason for this lies in the socio-demographic features of both source and host countries, namely low living standards and high population pressure in both Africa today and Mexico in the past, high income and low fertility in both destination countries (i.e. Europe and the U.S.). Our analysis provides evidence of the first–order effects of decreasing fertility on migration choices at the household level. In particular, we show that an exogenous decrease in the number of children does not necessarily decrease (or increase) the number of migrants in a family. This finding contributes to understanding the little impact that fertility-reducing programs (e.g., investments in family planning, sex and reproductive health) may have on migration decisions at the household level. Such measures have been endorsed in many developing countries as a policy response to the apparent vicious circle of high-fertility, poverty, and economic stagnation (Schultz, 2008; Miller and Babiarz, 2016). On the other hand, welfare policies such as pension schemes or old-age support

[10] On the basis of data from the 2000 U.S. Census of Population, the annual employment rate of foreign-born Mexican men was 88.5 percent compared to 56 percent of Mexican women. Hispanic immigrant men disproportionately worked in agriculture and construction, while Hispanic immigrant women were overrepresented in manufacturing and services (Duncan et al., 2006).

measures may reduce the incentives that push one household member (e.g. the first-born son) to migrate abroad. Finally, by showing that not all children in a household have the same likelihood of migrating, our results point to the existence of an intra-household selection process that may have important implications in terms of individual (child) welfare and income distribution.

The paper unfolds as follows. Section 2 presents the data and sample selection. The methodology and empirical strategy are described in Section 3. Section 4 presents our main results on sibship size effects on migration while Section 5 reports results on the role of birth-order, gender and sibling composition. Finally, Section 6 summarizes our main findings and concludes.

2 Data and sample selection

This study uses data from the 1992 and 1997 waves of the *Encuesta Nacional de la Dinámica Demográfica* (ENADID), a cross-section survey conducted by the National Institute of Statistics and Geography (INEGI) in Mexico. Each ENADID's wave surveys more than 50,000 households from all over the country and is representative of the Mexican population. The dataset is very rich and unique, collecting comprehensive information on women's fertility as well as migration history of all household members, in addition to standard socio-economic characteristics. Importantly, by using detailed demographic information on age (month and year of birth) and gender of individuals in the same household with the same mother, we are able to identify all biological families in the sample and recover complete information on the number and gender of all siblings (including those not currently living in the household of origin).

The ENADID collects detailed information on fertility for all women aged 15 to 54 at the time of the survey. Women answer specific questions regarding the number of children they have given birth to, their gender and birth order, current and past contraceptive use, fertility preferences, as well as their socio-economic and marital status. Unlike other studies where fertility is measured as the number of co-resident children, thanks to this rich dataset, we are able to precisely measure our key explanatory variable, i.e. the total number of biological children per woman. Moreover, information in ENADID enables us to identify exogenous shocks to parental fertility induced by infertility episodes, in addition to miscarriage at first pregnancy (see Section 3.2 for more details). In line with the extant literature and the medical definition of infertility, namely the failure to conceive after a year of regular intercourse without contraception, we restrict our sample to the children of non-sterilized women who are not currently using contraception or who never did. In so doing, we identify women with infertility episodes as those who report not using contraception because of infertility problems (see Agüero and Marks, 2011). This sample selection reduces measurement error in the definition of infertility as women can only be aware of an infertility condition if they do not use contraception.[11]

Our dataset allows us to define household members' international migration experience based on three separate questions: (i) whether there is any household member who migrated *abroad* (even temporarily) during the five years prior to the survey; (ii) whether any household member has ever worked in the U.S. or looked for a job while they were in the U.S. (and the year in which this occurred); and (iii) whether the respondent reports a period of residence abroad at any point in time prior to the survey. The use of these three different sources of information for migration episodes ensures that we are able to capture a relevant part of the international migration phenomenon.[12] Overall, in 1997 (1992) almost 18 (15) percent of households in Mexico reports having a member who migrated abroad.

Since we are interested in the effect of family size on parental investment in offspring's migration, we define individual migration episodes as *non-tied* migration, i.e. we exclude

[11] We check the robustness of our results to this sample selection, though, by also including the children of sterilized women and those using contraception in our sample. We use miscarriage at first birth as a source of fertility variation in this sample and results (reported in Table A1 in online Appendix A.1) appear to be unaffected.

[12] Other papers on migration using the same data set are Hanson (2004) and Mckenzie and Rapoport (2007) among others.

children who experienced episodes of tied-migration (with their parents) and those whose parents have an international migration experience. We do so for two main reasons. First, family and individual migration are inherently very different choices and our focus is on the latter. Second, we exclude parents with migration experiences because parental absence due to migration may affect fertility and hence generate a reverse causality problem.[13]

Figure 1 reports the incidence of non-tied migration by age and gender in Mexico showing that, overall, migrants are highly concentrated (over 70%) in the age range of 15 to 25. Throughout our analysis, we therefore restrict the sample to this age group. This is also consistent with the argument that Mexican youngsters finish compulsory schooling and can potentially enter the labor market at the age of 15, and that beyond the age of 25, they are more likely to make their own lives apart from their household of origin.

[Figure 1 about here]

One limitation of the data is that, by requesting migration information only for children who are still considered household members, i.e. those currently present or those who emigrated less than five years prior,[14] ENADID may introduce a potential sample selection bias if the children for whom we have information are more (or less) likely to come from larger families. We address this concern as follows. First, by focusing on migration outcomes in the age range 15-25, we lessen concerns of household partitioning. In fact, the average age at first marriage in Mexico during the 90s was between 22 and 23 for females and about 25 for males (World Bank Gender Statistics).[15] Thus, we expect the majority of children for whom we miss information to be mostly young, married daughters who do not live in extended families.[16] Moreover, Mexico-U.S. migration during the 90s was mostly of a temporary na-

[13] We check the robustness of our findings to the inclusion of tied-migrants in the sample (about 13 percent of the sample) or those for which parents had migration experiences, adding parents' migration status among the controls, in the analysis in online Appendix A.4. The results in Table A4 are unaffected.

[14] It is worth recalling that ENADID also collects information on migration episodes for temporarily absent household members, as long as migration occurred in the five years before the survey. Thus, ENADID only lacks information on permanent or long-term migration for non-household members.

[15] http://databank.worldbank.org.

[16] This is the reason for the gender imbalance observed in our estimation sample (see Table 1).

ture, with an average duration of about two years,[17] and most migrating children may still be considered as family members by their parents. Yet, if the probability of being observed in the data is correlated with family size, the estimated effect of family size on migration may still be biased. We further investigate this issue in Section 4.3.[18]

Our final estimation sample includes 26,743 children in the age range of 15 to 25, whose mothers are 45 years of age on average. The average birth spacing between the first and last child is 13 years, which is below the minimum age of the individuals we consider (15). This ensures that, on average, our measure of fertility can be interpreted as *completed fertility* at the moment of offspring's migration.[19] In other words, in our estimation sample the child migration decision occurs when all children are already born. This is important in order to avoid potential reverse causality issues related to child migration affecting their parents' fertility (e.g., through remittances).

In our sample of individuals, 5.2 percent are migrants, with male and female migration rates of 7.07 and 2.92 percent, respectively. In Figure 2 we plot the average migration rate of boys and girls in our sample by number of children.[20] A positive association between sibship size and the migration of sons clearly emerges. Individual sample characteristics according to migration status are reported in Table 1. Migrants are mostly male (75 percent) and report significantly more brothers and sisters than non-migrants. Moreover, migrant children appear to be slightly older and live in less educated households than non-migrant children. All in all, Table 1 suggests that child migration may be more frequent in households that are less well-off, households that also have a higher number of children on average.

[Figure 2 about here]

[17] Our computation for migrants of all ages in ENADID.

[18] Moreover, in online Appendix A, we run a series of robustness checks—that include sensitivity analyses on subsamples of sons (Table A2) and younger children (Table A3)—in order to show that our results do not suffer from sample selection bias induced by new household formation.

[19] Our sample does not include children whose mothers are older than 54 years of age (9 percent of the total population aged 15-25) since fertility information was not collected from them.

[20] Since our models estimated at the individual level include household fixed effects, we can only focus on children coming from households with two or more children. Household-level estimates, including households with single children, are reported in online Appendix E.

[Table 1 about here]

3 Empirical strategy and identification

3.1 Identification of sibship size and birth order effects

We are interested in the effects of sibship size and composition on an individual's likelihood to migrate. In order to estimate the effect of sibship size, however, we need to control for the birth order of children (see, for instance, Black et al., 2005). Indeed, if parents have a preference for their first-born children (i.e. lower parities) and invest comparatively more resources in them, then a spurious negative correlation between sibship size and human capital investments may emerge simply because in larger families we find children with higher birth orders. In other words, the two variables of birth order and sibship size are highly correlated. In particular, although one can assess the effect of family size on firstborns by looking at the outcomes of firstborns from families of different sizes, it is not possible to examine, for instance, the outcome of a fourth-born child when sibship size changes from two to three, given that fourth born children are only found in larger families.

Bagger et al. (2013) have proposed a theoretically-grounded methodology to disentangle the two effects. We draw on their idea and employ a similar two-step estimation strategy. In a first step we estimate the following regression using OLS:

$$M_{ij} = \alpha_0 + \sum_{k=2}^{K} \boldsymbol{\alpha}_{1k} bo_{ijk} + \boldsymbol{\alpha}_2 \mathbf{X}_{ij} + u_j + \epsilon_{ij} \tag{1}$$

where the outcome variable M_{ij} pertains to the migration status of child i in household j and is a dichotomous indicator of either current or past migration experiences abroad. bo_{ijk} is a dichotomous indicator for the child being of birth order $k = 2, ..K$ where K is the maximum birth order of children in our sample (top coded at 10 or more) and $k = 1$ (i.e. firstborn) is the reference group; $\mathbf{X}_{ij}$ is a vector of individual covariates including child gender, age, age squared and cohort indicators (one for each year of birth).[21] u_j is a family fixed effect, and ϵ_{ij} is an idiosyncratic error.[22]

The effect of sibship size is captured in equation (1) by the household fixed effects, which control for any (observed and unobserved) difference between families. The birth order fixed effects capture the differences in the probability of migration between children of different orders within the same family. Systematic differences in ages between different parities, which are likely to affect migration choices, are controlled for by a quadratic polynomial in child age. Only within-family variation is exploited in these estimates, and the birth order effects are not contaminated by between-family differences in family sizes, i.e. the fact that children in larger families also have higher average birth orders.

In the second step, we subtract the birth order effects from the dependent variable, i.e. we compute the difference $\widehat{NM}_{ij} = M_{ij} - \sum_{k=1}^{K} \hat{\boldsymbol{\alpha}}_{1k} bo_{ijk}$ where NM stands for "netted migration", and use this as the dependent variable in the second step.[23] Hence, the following equation is estimated:

$$\widehat{NM}_{ij} = \beta_0 + \beta_1 S_{ij} + \boldsymbol{\beta}_2 \mathbf{X}_{ij} + \boldsymbol{\beta}_3 \mathbf{W}_j + v_{ij} \tag{2}$$

where S_{ij} is sibship size. The coefficient β_1 captures the effect on migration of being raised in a family with sibship size S_{ij} for the average child in that family, i.e. regardless of his / her birth order. $\mathbf{X}_{ij}$ is a vector of individual covariates defined as above and $\mathbf{W}_j$ includes family background characteristics such as the mother's and father's age and age squared, and the mother's and father's years of completed education. In some specifications, we also control for maternal health (chronic diseases), father's absence from the household (i.e. widowed and

[21] We can include a control for both age and birth cohort because we use two cross-section surveys.

[22] Another way to disentangle birth order and family size effects has been suggested by Booth and Kee (2009). They build a new birth order continuous index that purges family size from birth order and use this to test if siblings are assigned equal shares in the family's educational resources. Since we prefer to estimate birth order effects using dichotomous indicators, we follow the approach described in Bagger et al. (2013).

[23] Coefficients of all birth order indicators (including firstborns) are recovered using the method described in Suits (1984), whereby the coefficients on the dummy variables show the extent to which the behavior of each birth order deviates from the average behavior (of all birth orders).

divorced single-mother families) and municipality fixed effects. The latter capture rural vs. urban residence along with many other factors related to different local cultural influences or socio-economic conditions such as access to contraception, water sanitation, quality of health care, distance from the U.S. border, etc. Importantly, municipality indicators also capture the local population size which may be related to the demographic pressure on both local labor supply and emigration.[24] Since the dependent variable has been generated by a regression, standard errors are corrected by weighting the estimation with the inverse of the standard error of $\widehat{NM}_{ij}$ using Weighted Least Squares (WLS).[25] Throughout, standard errors are clustered at the family level so as to account for potential error correlation across siblings. We also estimated models with heteroskedasticity-robust standard errors and the results hold.

If the number of children and investment in child out-migration are both outcomes over which parents exercise some choice, then the WLS estimate of the sibship size effect in equation (2) would provide spurious evidence. In other words, parental fertility may be endogenous with respect to offspring migration.

Hence, to clearly identify the relationship between sibship size and migration, an exogenous source of variation in family size is required. The ENADID allows us to identify self-reported infertility from specific questions. Similarly to Agüero and Marks (2008) we construct an indicator variable for infertility (i.e. the inability to conceive) that takes the value of one if a woman reports she is not currently using any contraception method (including natural ones) because of infertility, and zero otherwise. Two things are worth noting. First, the fact that a woman is not currently using contraception because of her inability to get pregnant does not imply that fertility impairments were also present during most part of her reproductive life. This means that we can observe large family sizes also for women

[24] This is to say that our identification strategy is able to isolate the within-family dimension of the impact of fertility on migration from the general equilibrium effect of population size. In some more data-demanding specifications reported in online Appendix B we also control for municipality-year fixed effects.

[25] See, for instance, Lewis and Linzer (2005). We also run estimates using White robust standard errors and the results of the analysis are unaffected.

reporting infertility problems. Hence, our indicator is closer to the medical definition of 'secondary infertility', i.e. a reduction in the ability to conceive or to carry a pregnancy to a live birth, than to 'primary infertility', i.e. the condition of women who never had a live birth. Second, in our sample of women who are not using contraception due to infertility, the largest share is represented by those who never used it (89%), while the share of women who stop using it because of infertility is relatively low (11%). The ENADID also enables us to build a second indicator variable that equals one if a woman experienced a miscarriage at first pregnancy ('fertility shock') and zero otherwise.[26] These two shocks are used in an instrumental variable (IV) strategy (implemented with Two-stage least squares, 2SLS).[27] The first-stage equation is

$$S_{ij} = \gamma_0 + \gamma_1 Z_j + \boldsymbol{\gamma}_2 \mathbf{X}_{ij} + \boldsymbol{\beta}_3 \mathbf{W}_j + u_{ij} \tag{3}$$

where u_{ij} is an idiosyncratic error term and Z_j is a dichotomous indicator for secondary infertility or miscarriage experienced by the mother of the potential migrant. The second-stage equation is:

$$\widehat{NM}_{ij} = \beta_0 + \beta_1 \widehat{S}_{ij} + \boldsymbol{\beta}_2 \mathbf{X}_{ij} + \boldsymbol{\beta}_3 \mathbf{W}_j + v_{ij} \tag{4}$$

where everything is defined as for equation (2), with the exception of $\widehat{S}_{ij}$ which comes from the estimation of equation (3).

The two-step procedure reported above is based on household fixed effects and therefore

[26] Miscarriages or spontaneous abortions typically refer to any loss of pregnancy that occurs before the 20th week of pregnancy.

[27] Other studies have considered different instruments such as twin births (e.g. Rosenzweig and Wolpin, 1980; Angrist and Evans, 1998; Càceres-Delpiano, 2006) and sibling-sex composition (e.g. Angrist and Evans, 1998; Fitzsimons and Malde, 2014). Those instruments, however, are not suitable either for our data or for the Mexican context. Twin births cannot be used because we do not have administrative data, and although we make use of a large survey, we observe twin births only in 1.3 percent of families in our estimation sample. Sibling-sex composition is not suitable to the Mexican context because, for its very nature, it is likely to affect fertility of parents who desire a small number of children. The idea behind the instrument is indeed that parents have an extra child just because they are not happy with the gender of those they already have (i.e. the group of compliers). This typically happens in Mexico when early parities are all females because parents have a son bias. However, average family size in Mexico is very large in our estimation period, the probability of having at least one son is also high, hence the instrument is unlikely to be relevant for a large share of the population.

can only be applied to households with more than one child. An alternative way to proceed is to estimate the migration equation using the household instead of the individual as the unit of analysis,[28] which enables us to retain in the estimation sample single-child households.[29] In so doing we are able to check the robustness of our baseline estimates to changes in the estimation sample and the estimation strategy. Indeed, focusing on the total number of migrants in the household as a function of total fertility, we do not need to control for birth order effects and we can use a standard instrumental variables procedure. Thus, we estimate a specification as follows:

$$m_j = \gamma_0 + \gamma_1 n_j + \boldsymbol{\gamma}_2 \mathbf{W}_j + v_j \tag{5}$$

where the dependent variable is the number of children in the age range 15-25 who ever migrated in household j and the independent variable of interest is n_j, i.e. the total number of children in household j. The coefficient γ_1 captures the increase in the number of migrants associated with a unitary increase in family size. Like in the child-level estimates, $\mathbf{W}_j$ includes family background characteristics such as the mother's and the father's age, age squared, and years of completed education, mother's age at first pregnancy, an indicator for the father not being in the household and municipality fixed effects; v_j is an household-level error term. This specification is estimated both with OLS and with 2SLS in online Appendix E.

3.2 Instruments' relevance, exogeneity and exclusion restriction

For our identification strategy to be valid, the two instruments must satisfy three conditions—i.e. relevance, exogeneity, and the exclusion restriction assumption—which are discussed below.

4.2.1 Relevance

[28] More precisely, our unit of analysis are biological children in the same household.

[29] Thus, in these estimates we also include individuals who do not have siblings, and look at whether they are more (less) likely to migrate than individuals with siblings.

Infertility or sub-fertility conditions have been already used in the economic literature to estimate the effect of the number of children and fertility timing on mothers' labor market outcomes in both advanced and developing countries (see, for instance, Agüero and Marks, 2008; Schultz, 2008; Agüero and Marks, 2011). We leverage on the same source of exogenous variation in sibship size to identify its causal effect on children's migration. Table 3 reports the incidence of infertility and miscarriage shocks in our (individual and household-level) estimation samples. Data clearly show a monotonic negative association between infertility and sibship size. For instance, while 13.4 percent and 11.4 percent of women with family sizes equal to one or two, respectively, have experienced an infertility condition, the incidence falls to 3.5 percent for women with seven children or more. A negative relationship also emerges between miscarriage and sibship size, although it is non–monotonic. More direct evidence on the instruments' relevance is reported in the 2SLS first stages.

4.2.2 Exogeneity

There is evidence that infertility is largely independent of the background characteristics of infertile women. For example, variables such as the father's social status and parity have been shown to be unrelated to observed heterogeneity in fertility (Joffe and Barnes, 2000). In an article summarizing the epidemiological literature regarding the role of lifestyle factors (cigarette smoking, alcohol and caffeine consumption, exercise, BMI, and drug use) in female infertility, Buck et al. (1997) conclude that few risk factors have been assessed or identified for secondary infertility. In addition, using U.S. data, education, occupation, and race have been shown to be unrelated to impaired fecundity (Wilcox and Mosher, 1993).

Also miscarriages have been used to identify fertility *tempo* and *quantum* effects on women's labor market outcomes (Hotz et al., 2005; Miller, 2011; Bratti and Cavalli, 2014). By their nature, miscarriages should have a negative effect on total fertility, and in our context on sibship size.[30] Their exogeneity is generally supported by the medical literature. For example, a few papers using administrative data, in which rich labor market and health

[30] Casterline (1989) stresses that in most societies pregnancy losses produce a reduction of fertility of 5-10% from the levels expected in the absence of miscarriages.

data are merged, show that in general miscarrying is not significantly associated with worse labor market outcomes (e.g., work absences) before miscarriage (Karimi, 2014; Markussen and Strøm, 2015). Only two etiological factors for miscarriage are recognized by different authors in the obstetrics literature, i.e. uterine malformations and the presence of balanced chromosomal rearrangements in parents (Plouffe et al., 1992). The latter though, are unlikely to be correlated with women's attitudes towards offspring's migration.

For both biological shocks a potential threat to identification may come from mothers' general health conditions, as these conditions may affect both fertility and child out-migration. Moreover, ill health might also be related to high levels of alcohol or tobacco consumption that have been observed to correlate with miscarriage (Garcìa-Enguìanos et al., 2002) or to obesity, which might reduce fecundity (Gesink Law et al., 2007).[31] We seek to attenuate these concerns by including controls for mothers' chronic illness and disabilities in our estimation models, as women who heavily consume alcohol, tobacco, drugs or who are obese are also more likely to have developed chronic conditions.

The number of miscarriages generally increases with the number of pregnancies (which depends in turn on desired fertility) and this could potentially generate a spurious positive correlation between the number of miscarriages and observed fertility. For this reason, we consider only miscarriages that occurred at the *first* pregnancy (Miller, 2011).

A possible way to support the exogeneity assumption is to regress women's characteristics on the biological shocks and show that the latter are not statistically significant predictors of the former (see, for instance, Agüero and Marks, 2011). However, these tests are informative only if women's predetermined characteristics are considered as dependent variables. Unfortunately, given the cross-sectional nature of the data, most women's characteristics provided in ENADID are measured at the time of the survey and may be considered as outcome variables (e.g., parents' marital status, income, labor market participation, etc.). Regressing them on the biological shocks would be equivalent to running reduced form mod-

[31] Other behavioral factors mentioned in Garcìa-Enguìanos et al. (2002) are caffeine, drugs consumption, and induced abortions.

els for the effect of fertility (level or timing) on such variables. In fact, we might expect significant coefficients for many of them, as fertility is likely to affect marriage duration and dissolution (Kjaer et al., 2014; Bellido et al., 2016), earnings (Miller, 2011; Lundborg et al., 2017) or labor supply (Angrist and Evans, 1998; Agüero and Marks, 2008), just to mention some. That said, we can implement such tests for a subsample of women, namely those still living with their mothers.[32] For these women we test whether infertility and fertility shocks are systematically associated with their mothers' (or parents') background characteristics, which are predetermined to maternal biological shocks. This test is possible thanks to the ENADID large sample size, unique features of ENADID demographic data that allow us to match women's shocks with their parents' characteristics and institutional features of Mexico where living in extended households is not rare.[33] We estimate a woman's likelihood to have sub-fertility episodes or miscarriage at first pregnancy as a function of parental background variables, while controlling for women's age, birth cohort dummies and municipality indicators. Despite the limitations of this exercise, since women cohabiting with their mothers cannot be considered as a random sample, Table 2 reports regression coefficients for both shocks, which are consistent with the literature and support the argument that infertility or sub-fertility conditions are randomly assigned and independent of the characteristics of women's family of origin.

[Table 2 about here]

Finally, as a last check of exogeneity of fertility and infertility shocks, for each child we compute the average biological shock at the level of the mother's municipality of residence (excluding his / her mother's own shock) and regress the individual shocks on these averages. If infertility shocks are as good as randomly assigned to women, we expect these municipality averages not to be significant predictors of individual shocks. Indeed, municipality averages may capture factors such as local sanitation and health conditions, social norms towards

[32] That is those for which we have parents' characteristics.

[33] Alternatively, the literature has been using (rare) long-spanning longitudinal data, which allow to link the childhood (background) family characteristics to grown-up children's outcomes (Joffe and Barnes, 2000).

voluntary abortion or access to illegal abortion, all factors that may also be correlated with children's attitudes towards migration, hence undermining our identification strategy (see Fletcher and Wolfe, 2009). The coefficients of the regression of individual infertility on municipality average infertility is 0.097 ($p-$value $= 0.36$), and the one of miscarriage at first pregnancy on the municipality average of the same variable is 0.064 ($p-$value $= 0.62$).[34] In both cases, these results show that biological shocks are unlikely the be driven by time-varying municipality-level unobservables.

4.2.3 Exclusion restriction

For our instruments to be valid, in addition to exogeneity, they have to satisfy the exclusion restriction assumption, i.e. fertility and infertility shocks have an impact on children's migration only through sibship size. For this reason, in the child migration equation we control for many variables that may act as confounding factors and for those that may be affected by the instruments while also having a direct effect on children's migration. Among these variables, we include the mother's age, age at first pregnancy, education, marital status and the husband's characteristics (age, education and absence). In particular, while parental education may directly influence fertility, it also acts as a proxy for household well-being and poverty. Yet, in a set of robustness checks we include additional controls for household economic conditions, namely municipality by year (1992 or 1997) fixed effects and municipality by parental education fixed effects (see online Appendix B).

A threat to the exclusion restriction assumption comes from the fact that miscarriage is a stressful event impacting negatively on women's mental wellbeing. In principle, this may impair children's geographical mobility for two reasons. The shock may create an emotional bond inducing children to stay close to their mothers, or emotionally distressed mothers may need more support at home, in both cases negatively affecting children's likelihood of migration. Unfortunately, ENADID does not provide information on respondents' mental health status and we cannot directly control for it. At the same time, this threat is less relevant

[34] Standard errors are clustered by municipality. Only children with mothers living in municipalities for which there are at least ten women in our baseline estimation sample are included in these regressions.

in our case. Indeed, we only exploit miscarriage at first pregnancy and since migration is measured for children aged 15 or more, such negative psychological shocks on mothers must persist for more than 15 years and to several childbirths to bias our results. However, an empirical check of the existence of direct effects of mother's emotional distress on children's migration is reported in online Appendix D and shows the lack of such effects.[35]

4.2.3 Measurement error

There is a potential issue of measurement error with the miscarriage instrument, since women may be unaware of miscarriages if they happen very early in the pregnancy[36], may fail to recall them (this may hold especially for older women, although mothers in our sample are not older than 54) or just avoid reporting them as they are painful events. Misreporting may affect the strength of the instrument but we do not expect any specific pattern of correlation between it and parents' attitudes towards child out-migration conditional on the observables (including a quadratic polynomial in maternal age). Finally, as it was formulated in the ENADID, the question does not distinguish between voluntary and involuntary abortions. Thus, some of the reported abortions may be actually voluntary, even though induced abortion was illegal and Mexico had the strictest anti-abortion legislation in Latin America during the period under consideration. For women who voluntary have an abortion, the instrument would be endogenous. However, there is no evident sign in our data that a relevant share of the recorded abortions could be voluntary. For instance, Catholic women in our sample do not tend to abort significantly less than other women (information on religion is available in the 1997 wave only): incidence of abortion is 4.6 percent in the former group and 4.8 percent in the latter.[37]

[35] We test for the potential direct effects of miscarriage on child migration, via the emotional distress that a traumatic event such as miscarriage can cause to the mother, drawing on the work of van den Berg et al. (2017). The authors show that a child's death represents one the largest losses that an individual can face and has adverse effects on parents' labor income, employment status, marital status and hospitalization. Similarly, we include child death and the duration of the pregnancy that ended in a miscarriage or a stillbirth as controls in the child migration equation, but we do not find 'grief' effects on migration. This analysis is reported in online Appendix D (Table D1).

[36] In this case, however, the effect on completed fertility is probably negligible.

[37] In case the instrument is substantially contaminated by voluntary abortions, we would expect IV estimates to be biased in the same direction as OLS. Indeed, omitting subscripts and in the models without

4 Results on family size

4.1 First-step estimation of birth order effects

We start by estimating the impact of birth order on individual migration controlling for household fixed effects, as specified in equation (1). The within-family estimator sweeps out all parental- and family-level heterogeneity, including sibship size. Moreover, family fixed effects account for omitted family-specific unobservable factors simultaneously affecting fertility and child migration. The first column of Table 4 reports estimates with a linear specification of birth order on the full sample, whereas in column (2) we allow for a more flexible specification by adding birth-order-specific dichotomous indicators. Regressions control for individual age and gender plus child's birth cohort dummies (one for each year of birth).[38] Indeed, child age is correlated with birth order and it is also likely to have a (non-linear) relationship with migration (which is why we include the age quadratic term).

First, in column (1) we observe that, after controlling for household fixed effects, birth order and individual characteristics, females are significantly less likely to migrate than males by 3.6 percentage points (p.p.). Moreover, the birth order point estimate is negative and statistically significant. Column (2) shows that the effect is non-linear and starts to be economically meaningful from children of birth order 3, who are 2.1 p.p. less likely to migrate than firstborns. Although this appears to be a small effect in absolute value, it represents an approximately 40 percent decrease in the probability of migration at the sample average (5.2 percent migration rate). The coefficients for the following birth orders are larger in absolute value and peak for birth orders 9 and 10 or more (-16.6 and -20 p.p. respectively). We also estimated (1) by allowing interactions between gender and birth order indicators,

controls, if we define as $M = \beta_0 + \beta_1 S + v$ the migration equation, where M and S are child migration status and sibship size, respectively, and $S = \gamma_0 + \gamma_1 Z + u$ the sibship size equation (the first stage) and Z the instrument (abortion), $\beta_{1,OLS} = \beta_1 + Cov(S,v)/Var(S)$ while $\beta_{1,IV} = \beta_1 + Cov(Z,v)/Cov(Z,S)$, where $Cov(Z,S) < 0$ and $sign(Cov(S,v)) = -sign(Cov(Z,v))$. In case, for instance, unobserved mother's total desired fertility is positively correlated with children's migration and a substantial share of abortions are voluntary, both OLS and IV will be similarly upward biased.

[38] By including child age and cohort dummies, with household fixed effects we are also *de facto* controlling for birth spacing between siblings.

but interaction terms are never statistically significant (see Section 5 for discussion). Thus we use the specification without gender interactions to implement the two-step procedure as described below.

[Table 4 about here]

4.2 Sibship size effect: WLS and 2SLS results at the individual-level

By applying the two-step procedure described above, we now turn to the estimation of the sibship size effects. We report the WLS estimates as a benchmark model, where the dependent variable is 'netted migration' (see Section 3.1).[39] The number of siblings is tallied as the number of currently living biological brothers and sisters of each child.[40] The first column of Table 5 reports WLS results for a linear specification including sibship size. The highly significant coefficient implies that, on average and after controlling for birth order effects in the first step, adding one sibling is associated with a 1.1 p.p. higher likelihood of migrating for young adults (+17 percent at the sample mean). The same effect holds once we include individual level controls, namely child gender, age, age squared and years of birth indicators (column 2). In column (3) we estimate the same model as above by allowing for differential effects by child gender. The significant negative coefficient for the interaction term indicates that females' likelihood to migrate increases less due to sibship size with respect to males. Specifically, one extra sibling raises the migration probability more for sons than for daughters by 0.8 p.p. In columns (4) to (7), we run the same regressions above while adding further parental, household and geographical-level controls

[39] The inverse of the standard errors of 'netted migration' are used as weights.

[40] Those currently deceased are excluded from our definition of siblings. This is done for two reasons: (i) 70 percent of deceased children in our sample died before the first year of life, 90 per cent of them before the second one; (ii) the focus of our analysis is not on very young children so that we need to take into account siblings who actually 'had enough time' to compete over household resources, and exclude accordingly infant deaths. In online Appendix C (Tables C1–C4) we report robustness checks related to concerns about the endogeneity of our definition of sibship size and birth order and estimate models based on ever-born children, i.e. currently alive or deceased, and the results do not change.

in order to account for potential confounding factors of the relationship between family size and offspring's migration. Specifically, in column (4) and (5) we include parental covariates, which may predict completed fertility and affect child migration, namely mother's years of birth indicators, age at first pregnancy, chronic illness, single status (i.e. widow, divorced, single *de facto*), father's decade of birth indicators, mothers' and father's (quadratic) age and years of schooling.[41] In column (6) and (7) we further add municipality fixed effects that, conditional on family size, control for population size along with many other local factors related to different cultural or economic conditions, which may have an effect on fertility and migration (e.g., employment rates, migration intensity, access to contraception, social services, etc.). Overall, the sibiship size effect is essentially unchanged when we control for all of the aforementioned factors, and the same holds for the differential effect by gender.

[Table 5 about here]

Yet, as noted in the methodological section, the coefficients on sibship size reported in Table 5 are still likely to be biased, even when a rich set of demographic and economic controls is included. This is so as fertility may be endogenous with respect to child out-migration. Thus, we employ an IV approach and exploit the arguably exogenous fertility variation generated by episodes of infertility and miscarriage. Since these events can vary the actual family size from the desired one, we use infertility shocks and miscarriage at first pregnancy to identify the effect of sibship size on child out-migration. In Table 6 we present 2SLS estimates and the two-step methodology, as outlined above, to estimate equation (4). In column (1) we instrument sibship size with an indicator variable for infertility shocks taking value one if the woman declares she is not using contraception because she is infertile. In column (2), instead, we report results using a woman's experience of miscarriage on her first pregnancy as an instrument. Eventually, in column (3) we present results using both instruments in an

[41] We are *de facto* also controlling for mother's age at delivery, which is a linear combination of child's age and mother's age. As far as parental controls are concerned, we have more missing information for fathers than it is the case for mothers. As to keep the sample size constant, we further include a dummy variable for missing paternal information

overidentified model. Throughout all models, the first stage results point to a strong and highly significant relationship between infertility / fertility shocks and completed fertility. In particular, children whose mothers experienced an infertility shock have a reduction in their sibship size of nearly 0.5 ($t = -5.2$) with an $F-$statistic of 26.9 (column 1). The negative impact of miscarriage on sibship size is similar in magnitude (-0.437) with an $F-$statistic of 19.13 (column 2). Also the $F-$statistic of the joint significance of the instruments in the over-identified model is as high as 23.37 (column 3).[42] The sibship size effects estimated using 2SLS are always economically small. In the models using miscarriage, infertility and both instruments the effect of increasing sibship size by one changes the likelihood of migration by 0.4, -1.8 and -0.5 percentage points, respectively. Although, due to our sample size, these estimates are not 'precisely estimated zeros' (i.e. very small statistically significant coefficients) they are zeros in economic terms. The lack of significance does not appear to be due to imprecision related with a weak instruments problem. For all models, the Anderson-Rubin $F-$statistic (robust to weak instruments) cannot reject that the coefficient on the instrument is zero in the reduced form. Interestingly, the point estimate of the effect of sibship size on child migration obtained with the miscarriage instrument (which might include voluntary abortions) is lower than the one obtained with the infertility instrument, which we consider to be less affected by endogeneity issues, and much lower than the OLS estimate, a fact that is inconsistent with the premise that induced abortions include a substantial share of total abortions (see Section 3.2).

Even though in all specifications we control for parental education, in online Appendix B (Tables B1 and B2) we show that our results are robust to the inclusion of a number of additional controls for household economic conditions, namely municipality by time fixed effects and municipality by parent's education fixed effects.

[Table 6 about here]

[42] The Hansen $J-$statistic does not reject the 'validity' of the instruments (i.e. orthogonality to the the error term and correct exclusion from the main equation) in the overidentified model.

The two instruments we used may act on different compliers. We expect miscarriage at first pregnancy to hit especially women with very high desired fertility. Indeed, in Mexico in the 1990s there was generally a long time span between a woman's first pregnancy and the end of her fertile life, during which women could reach their desired number of children. Only women who desired a very large family size were prevented from attaining their target by a miscarriage experienced at first pregnancy. By contrast, compliers with the infertility instrument may be in principle more evenly distributed across different desired family sizes. Since ENADID data do not provide the exact timing of the infertility shock, we are unable to check the age it occurs and the family size' margin the instrument is mainly relevant for. Yet, Table 3 suggests that infertility is more prevalent in (ex-post) smaller families. Accordingly, finding (in Table 6) similar results using instruments with potentially different compliers is reassuring in terms of the external validity of our estimates.

In Table 7 we report results of the same 2SLS regressions as above while testing the sibship size differential effect by gender in the pooled sample with interaction terms.[43] Results do not point to any significant difference in the impact of sibiship size between boys and girls, as it turns out to be insignificant for both (columns 1-3). When using miscarriage as an instrument, though, we cannot draw strong conclusions as the $F-$statistic for the interacted endogenous variable is rather low (4.27, column 2). However, even in this case the Anderson-Rubin $F-$statistic confirms that we cannot reject the hypotheses of sibship size not affecting child migration.

[Table 7 about here]

Overall, findings in this section point to the negligible role of family size on children's migration outcomes. The comparison between the OLS and the IV estimates indicates an upward bias in the former. According to our estimates the correlation between family size and child migration observed in the data is driven by unobservable variables which make

[43] The interaction effect sibship size$\times$female is instrumented using the interaction *instrument*$\times$female, where the *instrument* is infertility or miscarriage depending on the specification.

some families more prone to both have more children and more migrant children. Such variables may include, for instance, risk aversion and preferences for income diversification. Indeed, higher fertility may be a way for parents to diversify the sources of old-age income support and child migration a way to spatially diversify household production.

As outlined in Section 3.1, household-level estimates enable us to include single-child households in the estimation sample and use a standard 2SLS procedure. They can be therefore considered as a robustness check to changing the composition of the sample and the estimation strategy. The results are reported in online Appendix E (Tables E1 and E2) and confirm those of the individual level analysis.

4.3 Sample selection bias from new household formation

In this Section, we carry out a direct check for the potential sample selection bias induced by new household formation. We are able to analyze migration decision only of children who are considered as members of the household (i.e. those who are present or emigrated less than five years ago). If a new household formation (i.e. the child leaving the parental home and start living alone or forming his / her own family) is associated with the number of siblings, our estimates would not be representative of the effect of family size on migration for the whole population of children but only for children cohabiting with their parents or who are current household members. To check if this is the case, we estimate a LPM in which the dependent variable is a dichotomous indicator for a child's not being observed in the household ('absent child'). Even for the 'absent children' some individual level information can be recovered, namely gender, birth order and age, from the mother's fertility history, so that we can include in the estimated equation exactly the same controls as in the migration equation. First, in column (1) of Table 8, we estimate a reduced form LPM in which we include all the controls of the migration equation except sibship size, but we also include the two excluded instruments used for sibship size (miscarriage and infertility conditions). We interpret the test for the joint significance of the coefficients on the two instruments as evidence about the role of family size on a child's choice to permanently leave his / her origin household. Finding statistically significant coefficients on the instruments would generate concerns that estimates could be affected by a sample selection bias. We do a similar exercise in column (2) of Table 8 in which we use 2SLS, and test for the significance of family size on the 'absent child' equation. In both cases, we can reject the null hypothesis that family size is a significant driver of a child's probability of being included in the sample used to estimate the decision to migrate abroad.

[Table 8 about here]

5 Gender and sibling composition

5.1 Migration outcomes as a function of gender and birth order

In column (1) and (2) of Table 4 (see Section 4.1), we show that an individual's probability to migrate decreases with birth order. While this is consistent with a household optimal migration model where family's returns from migration decrease with child parity, in this and the reminder sections of the paper we present more compelling tests on whether migration chances are distributed unevenly—e.g. by age and gender—across children within the same family. Indeed, within a household resources allocation framework, low-parity children may be more likely to migrate because the family has more time to reap the benefits of migration. However, it may still be argued that first-born children are better off with respect to other forms of human capital investments as well. For example, earlier parities may have benefited from higher pre-natal or post-natal parental investments, having shared household resources with fewer siblings, and this may affect the returns to migration. Thus, here we explore the gendered pattern of migration in order to test the hypothesis of the low-parity advantage.

In column (3) and (4) of Table 4 we report results by adding interaction effects between birth order and gender to the models.[44] The interactions of being female with birth order

[44] As our two-step procedure relies on household fixed effects, when estimating separate regressions by

dummies are not statistically significant, suggesting that the birth-order gradient in child migration is not statistically different between boys and girls. Yet, the latter holds for all parities but for firstborns: in column (2) the female main effect shows that female firstborns are significantly less likely to migrate than male firstborns. Overall, these estimates suggest that the chances of migration are not equally distributed across children within the same family. Low-parity children are, in general, more likely to migrate but a firstborn daughter is significantly less likely to migrate than a firstborn son by about 3 p.p. (which means a reduction in the probability of migration of roughly 60 percent at the sample average migration rate). Finding a significant effect on the interaction between first-parity and gender suggests that it is unlikely that parents decide to invest in migration of the firstborns only, irrespective of gender (first-born bias). Also gender turns out to be a significant factor, and in particular (first-born) boys may have higher migration returns than their female peers. This is consistent with a male-dominated Mexican migration phenomenon, as shown by different data (e.g., Cerrutti and Massey, 2001). Yet, while parental investment in (low-parity) boys is still a rational choice when returns to migration in the U.S. labor market are higher (or moving costs are lower) for boys than for girls, these findings are also consistent with the argument that parents may just value (low parity) sons more than daughters (preference for sons). Hence, in the next and last sub-section, we estimate the same migration equations as above by allowing for a separate effect of sibling composition from the individual gender and birth-order variables. If migration choices are driven by birth-parity or son preference, we should find no separate effect of sibling composition. On the contrary, a significant effect of sibling composition variables on the likelihood to migrate points to the existence of an intra-household migration selection process within which some children may have systematically more chances to migrate than others.

---

gender only families with at least two sons and at least two daughters can be included in the estimates for males and females, respectively. In order to avoid such a sample selection, we rather adopt a pooled estimation including interaction effects with gender.

5.2 Migration outcomes as a function of gender and sibling-sex composition

Our estimates so far show that gender is a robust predictor of migration in Mexican families and, *ceteris paribus*, boys—especially firstborns—are systematically more likely to migrate abroad than girls. In practice, this means that if migration is costly and not all children are in the position to migrate, a pro-eldest-son migration bias may lead to a situation in which children compete for household resources in order to migrate and such 'rivalry' can yield gains to having relatively more older sisters than brothers (Garg and Morduch, 1998). Thus, in order to explore the scope of sibling rivalry by age and gender, we test how sibling composition influences child migration by running two sets of regressions as reported in Table 9. First, we estimate migration equations on the full sample of children as a function of the number of their older brothers, while controlling for both family and birth order fixed effects (i.e., conditioning on the number of both siblings and older siblings), child gender, (a quadratic polynomial in) age and cohort dummies. Results in column (1) show that, *ceteris paribus*, having an older brother (sister) instead of an older sister (brother) decreases (increases) the migration probability by 1.4 p.p. ($t = 3.6$). This result points to a significant role of the gender and age composition of siblings in children's migration outcomes, consistently with a household-level migration strategy. Moreover, the sibling composition effect does not differ significantly by the gender of the child, suggesting that older siblings' sex composition equally matters for boys and girls (column 2).

[Table 9 about here]

Yet, we further exploit the gendered migration pattern and the fact that siblings are likely to migrate in order of birth to test whether females pay a toll for higher migration returns for boys. We do so by including a control for having a next-born brother in the household fixed effects regressions on the pooled-sample (with and without interactive effects), as above. If a child has at least one younger sibling, the gender of his / her next-born sibling is random and a comparison of children with next-born brothers with children with next-born sisters, while controlling for older siblings composition, can identify the effect of the sibling's gender.[45] Results in columns (3) and (4) in Table 9 show that, conditional on older siblings' composition, having a next-born brother does not play any role for sons, but reduces the likelihood to migrate for girls with respect to boys by 1.2 p.p. ($t = -2$). This result suggests that sibling composition by gender and age plays a significant role in the determinants of migration decisions within the household. More specifically, from our results it seems that a daughter with a next-born brother is less likely to migrate than a girl with a next-born sister. In other words, when parents face the costly decision whether to send a daughter abroad, they seem to prefer to invest in the migration of her next-born brother (if there is one). Similarly, by allowing children's agency, it may be that young women are less likely to offering themselves as migrants if they have a next-brother.

All in all, our results are consistent with an optimal household migration strategy where private costs and returns of migration are shared among all siblings. Indeed, a low-parity Mexican boy in the 90s may be more valuable to send as a household migrant abroad than a girl. In addition, the opportunity cost of sending girls abroad may be higher because they usually take care of chores and family duties at home or are in charge of being close to parents in their elderly age. Hence, social norms or practices combined with market returns on the migration investment may explain the male-dominated pattern of Mexico-U.S. migration and document—similarly to other developing contexts—that young females tend to have less access to human capital investment and enhancing economic opportunities than it is the case for males.

[45] A similar empirical strategy is employed by Vogl (2013) to study sibling rivalry over arranged marriages in South Asia.

6 Conclusions

In this paper we provide novel and rigorous evidence on the extent to which international labor mobility is affected by the demographic characteristics of the migrant's household of origin. Migration is largely a youth phenomenon occurring in households that seldom dispatch all or most of their children to work abroad. With capital market imperfections and high migration costs, the 'resource dilution' hypothesis predicts that a larger sibship size will decrease the chances of an offspring migrating. Yet, in relatively poor contexts, parents are likely to depend on their grown-up children for the provision of care and income, and migration opportunities can significantly contribute to the living arrangements of elderly parents.

We use data on teenagers and young adults from a rich household survey to examine the causal effects of sibship size, birth order and sibling composition on migration outcomes in Mexico. Mexican migration, mainly to the U.S., is an enduring flow that accounts for one third of total U.S. immigration and one-tenth of the entire population born in Mexico. Importantly, migration patterns in the 90s differed by age and gender, with a significant fraction of Mexican males migrating between the ages of 15 and 25.

Our large dataset allows us to overcome the limitations of small samples of children, and it includes detailed information on both women's fertility and the migration histories of household members. We find no evidence that larger families have a causal impact on migration. The positive link between family size and migration breaks down when potential endogeneity is addressed using biological fertility and infertility shocks. On the other hand, we find differences in the chances of migration between siblings within the same family. Older siblings, especially firstborn males, are more likely to migrate, while having relatively more older brothers than sisters systematically decreases the likelihood of migration for all children. Yet, girls, but not boys, are less likely to migrate when their next parity is a male.

Our findings are consistent with an optimal household migration model in which parents maximize returns to migration when deciding on whether to have one or more of their children move abroad. Large family size *per se* does not constitute a significant push factor in the migration choice, whilst gender, birth order and sibling composition have much more influence on the migration outcome. In particular, in resource-scarce contexts, out-migration of girls can be viewed as less economically rewarding and more socially costly to parents, with the result that boys end up having more economic opportunities than girls, even through migration.

These results contribute to the migration literature by shedding new light on the significant role of both the family dimension and demographic factors in the migration decision problem. Labor mobility, especially from poor to rich settings, is one of the most important ways through which young adults can expand their productivity and earning potentials. The type of family-based migration that occurred from Mexico to the U.S. during the 1990s is of considerable and growing importance for many of today's developing countries (e.g., in Asia and Africa) where both migration and fertility rates are substantial (e.g., Hatton and Williamson, 2003). Despite the easily observable association between high fertility rates and migration, we provide evidence that large families are unlikely to be a systematic driver of migration at the household level.

Understanding the link between fertility and migration is also relevant today since many governments in developing countries have attempted to curb population growth as a means of increasing the average human capital investment and possibly reducing migration (e.g., China and India, the world's two most populous countries, have experimented with different family planning policies to control family size). Yet, although our empirical findings do not point to a causal link between family size and migration, they hint to the fact that parental returns from offspring migration may play a role in lifetime fertility choices. Moreover, by showing that not all children within the family have the same chances to migrate, our findings point to the existence of an optimal intra-household selection process into migration. This is so as in contexts of scarce resources and weak formal safety nets, children may be a key social security valve for parents such that high migration opportunities to rich countries may increase the value of children (some more than others, e.g., low-parity sons). Hence, effective family welfare measures or the development of credit and insurance markets may lead to a reduction in both migration and fertility, and perhaps less gender inequality.

Eventually, it is worth noting that our analysis seeks to address endogeneity concerns related to fertility resorting to women's sub-fertility conditions, miscarriage and stillbirth events, and we have reported several pieces of evidence consistent with the validity of the instruments we used. Moreover, quite reassuringly, the evidence based on different instruments, which act on different compliers, and on an over-identified model, all lead to the same conclusion. Yet, there might some remaining concerns that our instruments are not completely exogenous, or have direct effects on child migration. For this reason, given the paucity of the research investigating the causal effects of family size on children's migration, it would be important to build more evidence, on other countries and / or using other sources of presumably exogenous variation in fertility (e.g., fertility control programs for which a treatment and a comparison group can be clearly identified) to assess the external validity of our findings.

Acknowledgements

This paper benefited from helpful comments at various stages from Michael Clemens, Pascaline Dupas, Margherita Fort, Ilyana Kuzmenko, Luca Nunziata, Giovanni Peri, Maria Perrotta, Anna Piil Damm, Erik Plug, Nicole Schneeweis, Jenny Simon, Giancarlo Spagnolo, Alessandro Tarozzi, Shqiponja Telhaj, Daniela Vuri, Alan Winters and seminar participants at GREQAM Aix-Marseille, SITE-Stockholm School of Economic, Nova School of Business and Economics (Lisbon), University of Sussex, University of Bologna, University of Turin, University of Rome "Tor Vergata," the 13th IZA Annual Migration Meeting (Bonn), the Royal Economic Society Conference (Brighton), the ESPE Conference (Izmir) and the AIEL Conference (Trento). All errors are our own.

References

Abramitzky, R., L. P. Boustan, and K. Eriksson (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration. *Journal of Development Economics 102*, 2–14.

Agüero, J. M. and M. S. Marks (2008). Motherhood and female labor force participation: Evidence from infertility shocks. *American Economic Review 98*(2), 500–504.

Agüero, J. M. and M. S. Marks (2011). Motherhood and female labor supply in the developing world: Evidence from infertility shocks. *Journal of Human Resources 46*(4), 800–826.

Angelucci, M. (2015). Migration and financial constraints: Evidence from Mexico. *Review of Economics and Statistics 97*(1), 224–228.

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

Bagger, J., J. Birchenall, H. Mansour, and S. Urzúa (2013). Education, birth order, and family size. NBER Working Papers 19111, National Bureau of Economic Research, Inc.

Becker, G. S. and H. Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2), S279–S288.

Becker, G. S. and N. Tomes (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy 84*(4), S143–62.

Bellido, H., J. A. Molina, A. Solaz, and E. Stancanelli (2016). Do children of the first marriage deter divorce? *Economic Modelling 55*(C), 15–31.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 669–700.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2016). Healthy(?), wealthy, and wise: Birth order and adult health. *Economics and Human Biology 23*, 27–45.

Booth, A. and H. Kee (2009). Birth order matters: the effect of family size and birth order on educational attainment. *Journal of Population Economics 22*(2), 367–397.

Borjas, G. (1987). Self-selection and the earnings of immigrants. *American Economic Review 77*(4), 531–53.

Borjas, G. J. and L. F. Katz (2007, May). The evolution of the mexican-born workforce in the United States. In G. J. Borjas (Ed.), *Mexican Immigration to the United States*, pp. 13–56. Chicago: University of Chicago Press.

Bratti, M. and L. Cavalli (2014). Delayed first birth and new mothers' labor market outcomes: Evidence from biological fertility shocks. *European Journal of Population 30*(1), 35–63.

Buck, G. M., L. E. Sever, R. E. Batt, and P. Mendola (1997). Life-style factors and female infertility. *Epidemiology 8*(4), 435–441.

Cabrera, G. (1994). Demographic dynamics and development: The role of population policy in Mexico. *Population and Development Review 20*, 105–120.

Càceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources 41*(4), 738–754.

Card, D. and E. Lewis (2007). The diffusion of mexican immigrants during the 1990s: Explanations and impacts. In G. J. Borjas (Ed.), *Mexican Immigration to the United States.*

Casterline, J. (1989). Collecting data on pregnancy loss: A review of evidence from the World Fertility Survey. *Studies in Family Planning 20*(2), 81–95.

Cerrutti, M. and D. Massey (2001). On the auspices of female migration from Mexico to the United States. *Demography 38*(2), 187–200.

Chen, J. J. (2006). Migration and imperfect monitoring: Implications for intra-household allocation. *American Economic Review 96*(2), 227–231.

Duncan, B., V. Hotz, and S. Trejo (2006). *Hispanics in the U.S. Labor Market.* (Washington (DC): National Academies Press (US) ed.).

Fitzsimons, E. and B. Malde (2014). Empirically probing the quantity-quality model. *Journal of Population Economics 27*(1), 33–68.

Fletcher, J. M. and B. L. Wolfe (2009). Education and Labor Market Consequences of Teenage Childbearing: Evidence Using the Timing of Pregnancy Outcomes and Community Fixed Effects. *Journal of Human Resources 44*(2).

Garcìa-Enguìanos, A., M. E. Calle, J. Valero, S. Luna, and V. Domìnguez-Rojas (2002). Risk factors in miscarriage: A review. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 102*(2), 111–119.

Garg, A. and J. Morduch (1998). Sibling rivalry and the gender gap: Evidence from child health outcomes in Ghana. *Journal of Population Economics 11*(4), 471–493.

Gesink Law, D., R. Maclehose, and M. Longnecker (2007). Obesity and time to pregnancy. *Human Reproduction 22*(2), 414–420.

Hanson, G. (2004). Illegal migration from Mexico to the United States. *Journal of Economic Literature Vol. 44*(4), 869–924.

Hanson, G. and C. McIntosh (2016, Fall). Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run. *Journal of Economic Perspectives 30*(4), 57–82.

Hanson, G. H. and C. McIntosh (2010). The Great Mexican migration. *The Review of Economics and Statistics 92*(4), 798–810.

Hatton, T. J. and J. G. Williamson (2003). Demographic and economic pressure on emigration out of Africa. *Scandinavian Journal of Economics 105*(3), 465–486.

Hotz, V. J., S. W. McElroy, and S. G. Sanders (2005). Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *Journal of Human Resources 40*(3), 683–715.

Jayachandran, S. and I. Kuziemko (2011). Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India. *The Quarterly journal of economics 126*(3), 1485–1538.

Jayachandran, S. and R. Pande (2017). Why are Indian children so short? *American Economic Review 107*(9), 2600–2629.

Joffe, M. and I. Barnes (2000). Do parental factors affect male and female fertility? *Epidemiology 11*(6), 700–705.

Kantarevic, J. and S. Mechoulan (2006). Birth order, educational attainment, and earnings: An investigation using the psid. *Journal of Human Resources 41*(4), 755–777.

Karimi, A. (2014, August). Effects of the timing of births on women's earnings – Evidence from a natural experiment. Working Paper Series 2014:17, IFAU – Institute for Evaluation of Labour Market and Education Policy.

Kjaer, T., V. Albieri, A. Jensen, S. K. Kjaer, C. Johansen, and S. O. Dalton (2014). Divorce or end of cohabitation among Danish women evaluated for fertility problems. *Acta Obstetricia Et Gynecologica Scandinavica 93*(3), 269–276.

Lewis, J. B. and D. A. Linzer (2005). Estimating regression models in which the dependent variable is based on estimates. *Political Analysis 13*(4), 345–364.

Lindstrom, D. P. and S. G. Saucedo (2002). The short-and long-term effects of us migration experience on mexican women's fertility. *Social Forces 80*(4), 1341–1368.

Lindstrom, D. P. and S. G. Saucedo (2007). The interrelationship between fertility, family maintenance, and mexico-us migration. *Demographic Research 17*.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017, June). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review 107*(6), 1611–37.

Markussen, S. and M. Strøm (2015). The effects of motherhood. Technical report, MEMORANDUM No 19/2015, Department of Economics, University of Oslo.

Mayer, J. and R. T. Riphahn (2000). Fertility assimilation of immigrants: Evidence from count data models. *Journal of population Economics 13*(2), 241–261.

Mckenzie, D. and H. Rapoport (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics 84*(1), 1–24.

Miller, A. (2011). The effect of motherhood timing on career path. *Journal of Population Economics 24*(3), 1071–1100.

Miller, G. and K. S. Babiarz (2016). Family planning program effects: Evidence from microdata. *Population and Development Review 42*(1), 7–26.

Passel, J., D. Cohn, and A. Gonzalez-Barrera (2012). Net migration from Mexico falls to zero—and perhaps less. *Pew Research Centre, Hispanic Trends April 23*.

Plouffe, L. J., T. S. White E. W., C. Sweet, L. Layman, G. Whitman, and P. McDonough (1992). Etiologic factors of recurrent abortion and subsequent reproductive performance of couples: Have we made any progress in the past 10 years? *American Journal of Obstetrics and Gynecology 167*(2), 313–321.

Rosenzweig, M. R. (1988). Risk, implicit contracts and the family in rural areas of low-income countries. *Economic Journal 98*(393), 1148–1170.

Rosenzweig, M. R. and O. Stark (1989). Consumption smoothing, migration, and marriage: Evidence from rural India. *Journal of Political Economy 97*(4), 905–26.

Rosenzweig, M. R. and K. I. Wolpin (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica 48*(1), 227–240.

Rosenzweig, M. R. and J. Zhang (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's 'one-child' policy. *Review of Economic Studies 76*(3), 1149–1174.

Sarma, V. J. and R. A. Parinduri (2015). Children and maternal migration: evidence from exogenous variations in family size. *Applied Economics Letters 22*(15), 1184–1187.

Schultz, T. P. (2008). Population policies, fertility, women's human capital, and child quality. Volume 4, Chapter 52, pp. 3249–3303. Amsterdam: Elsevier.

Sjaastad, L. (1962). The costs and returns of human migration. *Journal of Political Economy 70*(5), 80–93.

Stark, O. (1981). The asset demand for children during agricultural moderninzation. *Population and Development Review 4*(3), 671–675.

Stark, O. (1991). *The migration of labour.* Basil Blackwell, Cambridge.

Stöhr, T. (2015). Siblings' interaction in migration decisions: who provides for the elderly left behind? *Journal of Population Economics 28*(3), 593–629.

Suits, D. B. (1984). Dummy variables: Mechanics v. interpretation. *The Review of Economics and Statistics 66*(1), 177–80.

Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of Human Resources 25*(4), 635–664.

UN (2013). World youth report. Technical report, United Nations.

UNDESA (2011). The age and sex of migrants. Technical report, United Nation, Department of Economics and Social Affairs, Population Division.

van den Berg, G. J., P. Lundborg, and J. Vikstrm (2017). The economics of grief. *The Economic Journal 127*(604), 1794–1832.

Vogl, T. S. (2013). Marriage institutions and sibling competition: Evidence from south Asia. *The Quarterly Journal of Economics 128*(3), 1017–1072.

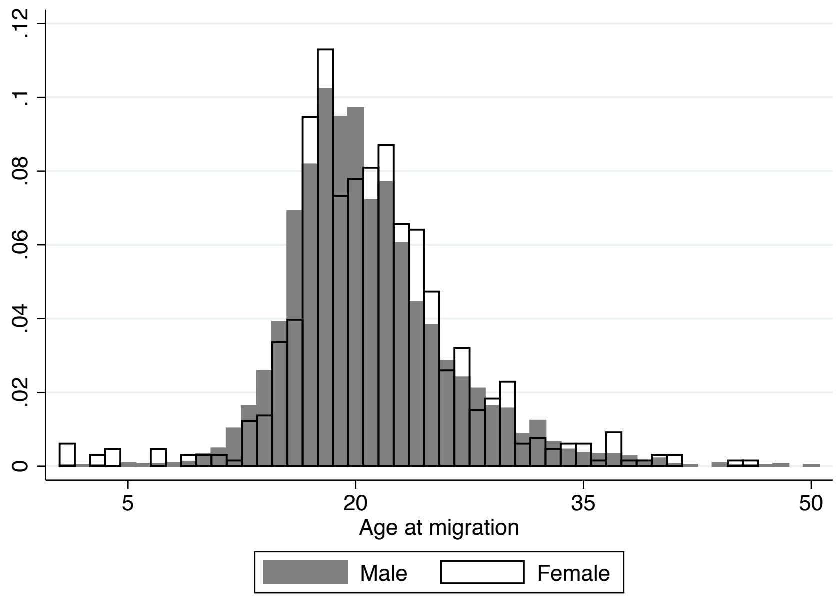
Wilcox, L. S. and W. D. Mosher (1993). Use of infertility services in the United States. *Obstetrics & Gynecology 82*(1), 122–127.

Williamson, J. (1990). *Coping with City Growth during the British Industrial Revolution.* Cambridge, MA: Harvard University Press.

Winters, P., A. De Janvry, and E. Sadoulet (2001). Family and community networks in mexico-us migration. *Journal of human resources*, 159–184.

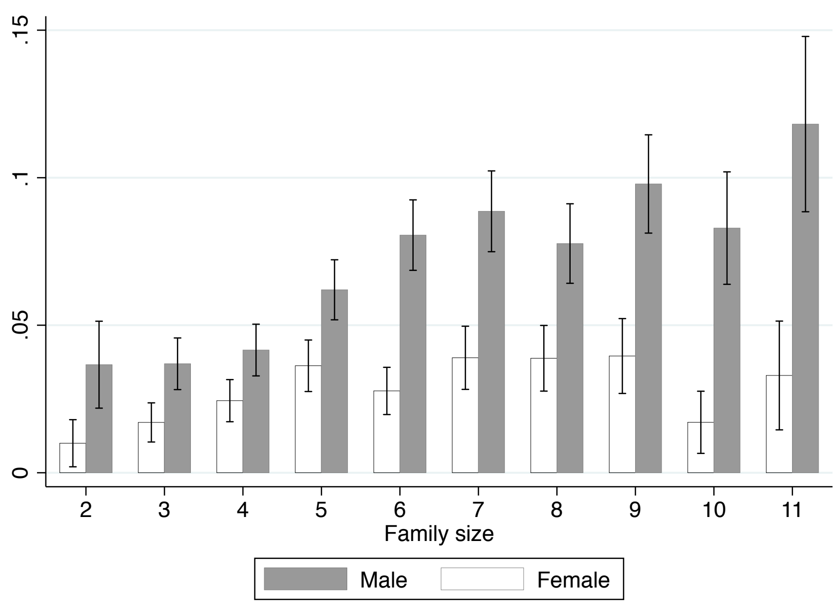
Zuniga Herrera, E. (2008). Population policy in Mexico and its relationship to development. presented at Contribution of Population Policy to the Achievement of the Internationally Development Goals including the MDGs. Panel discussion, New York, 15 December 2008.

Figure 1: Distribution of Mexican individual (non-tied) migration by age and gender



Source: Our computations on ENADID, 1992 and 1997.

Figure 2: Migration rate by family size



Source: Our computations on ENADID, 1992 and 1997. The figures report the share of migrants by family size and gender with 95% confidence intervals. Statistics are shown for the sample of individuals for which we have two or more siblings and we can include household fixed effects in the estimates.

Table 1: Sample characteristics by migration status

	Non-migrants (A)	Migrants (B)	$p-$values (A)-(B)
Individual-level characteristics			
Age	18.878	20.982	0.000
Female	0.458	0.250	0.000
N. of siblings	5.071	5.869	0.000
Birth order 1	0.181	0.192	0.300
Birth order 2	0.231	0.225	0.555
Birth order 3	0.178	0.178	0.978
Birth order 4	0.137	0.154	0.077
Birth order 5	0.102	0.102	0.993
Birth order 6	0.071	0.073	0.781
Birth order 7	0.046	0.041	0.343
Birth order 8	0.028	0.021	0.100
Birth order 9	0.014	0.009	0.121
Birth order 10+	0.011	0.006	0.107
Household-level characteristics			
Mother's age	44.612	46.171	0.000
Mother's age at first pregnancy	20.030	19.699	0.182
Mother's years of schooling	4.091	3.452	0.010
Mother chronic illness	0.023	0.008	0.131
Single mother	0.185	0.188	0.896
Father's age	48.799	52.207	0.000
Father's years of schooling	4.931	3.789	0.059

Note. Source: ENADID, 1992 and 1997. The estimation sample includes individuals aged 15–25 whose mothers are not using contraceptive methods. The sample comprises 1,394 migrants and 25,349 non-migrant individuals.

Table 2: Correlations between infertility shocks and women's background variables

	(1)	(2)
Dependent variable:	Infertility	Miscarriage at first pregnancy
Mother's age	0.009	0.013
	(0.006)	(0.010)
Mother's age squared	-0.000	-0.000
	(0.000)	(0.000)
Mother's age at first pregnancy	-0.000	0.001
	(0.000)	(0.001)
Mother's years of schooling	0.001	0.000
	(0.001)	(0.001)
Single mother	-0.006	-0.005
	(0.006)	(0.013)
Mother chronic illness	-0.026***	0.027
	(0.008)	(0.032)
Father's age	-0.001	-0.009
	(0.005)	(0.013)
Father's age squared	0.000	0.000
	(0.000)	(0.000)
Father's years of schooling	0.001	0.001
	(0.001)	(0.001)
Individual (quadratic) age	YES	YES
Individual's year of birth indicators	YES	YES
Mother's cohort dummies	YES	YES
Father's cohort dummies	YES	YES
Municipality indicators	YES	YES
Observations	4,268	4,268
R-squared	0.101	0.059

Note. The dependent variable is a dichotomous indicator of the woman having experienced an infertility episode (column 1) or miscarriage at first birth (column 2). The model is estimated using OLS. The estimation sample includes women who had at least one pregnancy and who are considered members of their mother's household in ENADID's 1992 and 1997 waves. Robust standard errors in parentheses. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table 3: Incidence of fertility and infertility shocks by sibship size

Individual sample				Household sample			
		Incidence of shock (%)				Incidence of shock (%)	
sibship size	%	infertility	miscarriage	sibship size	%	infertility	miscarriage
				0	3.69	13.37	5.05
1	4.59	11.56	6.03	1	9.84	11.43	6.80
2	12.16	8.33	5.38	2	16.88	7.48	5.20
3	14.20	5.45	4.06	3	15.96	5.16	4.02
4	14.68	5.12	4.05	4	13.66	4.30	3.86
5	13.54	4.00	5.55	5	11.20	3.73	4.71
6+	40.82	3.94	3.68	6+	28.76	3.51	3.44
	100.00				100.00		

Note. The table reports the incidence of fertility and infertility shocks in the estimation samples used in the individual-level (see Section 4.2) and the household-level analysis (see online Appendix E), respectively.

Table 4: Birth order effects on child's migration status

Variables	(1)	(2)	(3)	(4)
female	-0.036***	-0.035***	-0.032***	-0.031***
	(0.003)	(0.003)	(0.006)	(0.007)
birth order	-0.019***		-0.019***	
	(0.003)		(0.003)	
birth order 2		-0.002		0.002
		(0.005)		(0.006)
birth order 3		-0.021***		-0.023***
		(0.007)		(0.008)
birth order 4		-0.038***		-0.034***
		(0.010)		(0.011)
birth order 5		-0.068***		-0.070***
		(0.013)		(0.014)
birth order 6		-0.086***		-0.077***
		(0.016)		(0.017)
birth order 7		-0.112***		-0.103***
		(0.019)		(0.020)
birth order 8		-0.136***		-0.140***
		(0.022)		(0.023)
birth order 9		-0.161***		-0.166***
		(0.026)		(0.028)
birth order 10+		-0.199***		-0.188***
		(0.030)		(0.033)
birth order, female			-0.001	
			(0.001)	
birth order 2, female				-0.011
				(0.009)
birth order 3, female				0.005
				(0.010)
birth order 4, female				-0.010
				(0.010)
birth order 5, female				0.006
				(0.011)
birth order 6, female				-0.018
				(0.012)
birth order 7, female				-0.017
				(0.015)
birth order 8, female				0.010
				(0.018)
birth order 9, female				0.012
				(0.024)
birth order 10+, female				-0.022
				(0.027)
age	0.020**	0.021**	0.020**	0.021**
	(0.009)	(0.009)	(0.009)	(0.009)
age squared	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth indicators	YES	YES	YES	YES
Household fixed effects	YES	YES	YES	YES
Observations	26,743	26,743	26,743	26,743
R-squared	0.050	0.052	0.050	0.053

Note. The dependent variable is a dichotomous indicator of the child's migration status. The model is estimated using OLS. Sibship size is absorbed by household fixed effects. Standard errors clustered at the household level in parentheses. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table 5: Sibship size effect on child's 'netted migration' status: WLS estimates

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N. siblings	0.011***	0.011***	0.014***	0.010***	0.013***	0.010***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N. siblings × female			-0.008***		-0.007***		-0.006***
			(0.001)		(0.001)		(0.001)
female		-0.038***	-0.036***	-0.033***	-0.031***	-0.033***	-0.031***
		(0.003)	(0.006)	(0.003)	(0.006)	(0.003)	(0.003)
Individual's controls	NO	YES	YES	YES	YES	YES	YES
Mother's controls	NO	NO	NO	YES	YES	YES	YES
Father's controls	NO	NO	NO	YES	YES	YES	YES
Municipality indicators	NO	NO	NO	NO	NO	YES	YES
Observations	26,743	26,743	26,743	26,743	26,743	26,743	26,743
R-squared	0.013	0.054	0.055	0.177	0.178	0.202	0.203

Note. The dependent variable is 'netted migration' (see Section 3). The model is estimated using Weighted Least Squares (weights are the inverse of the standard errors of 'netted migration'). Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table 6: Sibship size effect on child's 'netted migration' status: 2SLS estimates

Variables	(1)	(2)	(3)
Second stage			
N. siblings	0.004	-0.018	-0.005
	(0.014)	(0.023)	(0.012)
female	-0.033***	-0.033***	-0.033***
	(0.003)	(0.003)	(0.003)
IV:	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic	0.073	0.686	0.389
	[0.787]	[0.407]	[0.678]
Hansen $J-$statistic			0.737
			[0.391]
First stage — N. siblings			
infertility	-0.494***		-0.491***
	(0.095)		(0.095)
miscarriage		-0.437***	-0.433***
		(0.10)	(0.10)
female	-0.032	-0.032	-0.033
	(0.024)	(0.024)	(0.024)
Angrist-Pischke $F-$statistic instrument(s)	26.90	19.13	23.37
Individual's controls	YES	YES	YES
Mother's controls	YES	YES	YES
Father's controls	YES	YES	YES
Municipality indicators	YES	YES	YES
Observations	26,743	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. The list of control variables is the same as in Table 5. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table 7: Child gender and sibship size effect on child's 'netted migration' status: 2SLS estimates

Variables	(1)	(2)	(3)
Second stage			
N. siblings	0.005	-0.065	-0.007
	(0.016)	(0.048)	(0.015)
N. siblings × female	-0.005	0.112	0.005
	(0.013)	(0.079)	(0.013)
female	-0.032***	-0.064***	-0.034***
	(0.004)	(0.022)	(0.005)
IV:	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic	0.074	2.210	1.150
	[0.928]	[0.110]	[0.331]
Hansen $J-$statistic			4.399
			[0.111]
First stage — N. siblings			
infertility	-0.567***		-0.564***
	(0.109)		(0.108)
infertility × female	0.168		0.169
	(0.115)		(0.115)
miscarriage		-0.453***	-0.450***
		(0.117)	(0.117)
miscarriage × female		0.037	0.038
		(0.106)	(0.105)
female	-0.041	-0.033	-0.043*
	(0.025)	(0.025)	(0.025)
Angrist-Pischke $F-$statistic instrument(s)	28.62	11.98	15.68
First stage — N. siblings × female			
infertility	0.125***		0.125***
	(0.038)		(0.038)
infertility × female	-0.694***		-0.691***
	(0.131)		(0.131)
miscarriage		-0.067	-0.068
		(0.044)	(0.043)
miscarriage × female		-0.261**	-0.254*
		(0.131)	(0.130)
female	0.292***	0.269***	0.303***
	(0.03)	(0.03)	(0.031)
Angrist-Pischke $F-$statistic instrument(s)	26.93	4.27	13.83
Individual's controls	YES	YES	YES
Mother's controls	YES	YES	YES
Father's controls	YES	YES	YES
Municipality indicators	YES	YES	YES
Observations	26,743	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. The list of control variables is the same as in Table 5. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table 8: The effect of family size on being an 'absent child'

Variables	(1) OLS	(2) 2SLS
infertility	-0.013	
	(0.010)	
miscarriage	0.002	
	(0.011)	
N. siblings		0.015
		(0.015)
$F-$statistic instruments$^{(a)}$	0.838	
IV:	-	overidentified
Anderson-Rubin $F-$statistic		0.838
		[0.433]
Hansen $J-$statistic		0.875
		[0.350]
Individual's controls	YES	YES
Mother's controls	YES	YES
Father's controls	YES	YES
Municipality indicators	YES	YES
Observations	34,852	34,852

Note. The dependent variable is a dichotomous indicator of the child's not being observed in the household of origin ('absent child'). The model is estimated using OLS in column (1) and two-stage least squares in column (2). The list of control variables is the same as in Table 5. Standard errors clustered at the household level in parentheses. $^{(a)}$ $F-$ statistic for the joint test that infertility and miscarriage are zero in the 'absent child' equation. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table 9: Siblings' composition effect on child's migration status: OLS estimates

Variables	(1)	(2)	(3)	(4)
N. older brothers	-0.014***	-0.014***	-0.017***	-0.016***
	(0.004)	(0.004)	(0.004)	(0.005)
female	-0.028***	-0.026***	-0.022***	-0.016***
	(0.003)	(0.005)	(0.005)	(0.006)
N. older brothers × female		-0.002	-0.002	-0.003
		(0.002)	(0.002)	(0.002)
Next brother			-0.005	0.001
			(0.004)	(0.004)
Next brother × female				-0.012**
				(0.006)
Age, age squared	YES	YES	YES	YES
Birth order fixed effects	YES	YES	YES	YES
Year of birth indicators	YES	YES	YES	YES
Household fixed effects	YES	YES	YES	YES
Observations	26,743	26,743	26,743	26,743
Number of hid	10,139	10,139	10,139	10,139
R-squared	0.053	0.053	0.053	0.053

Note. The dependent variable is a dichotomous indicator of the child's migration status. The model is estimated using OLS. Sibship size is absorbed by household fixed effects. Standard errors clustered at the household level in parentheses. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Appendix to "The Impact of Family Size and Sibling Structure on the Great Mexico-U.S. Migration"

A Appendix: Robustness checks

As described in Section 2, we restrict our sample to children of mothers for whom we have information on arguably exogenous variation in fertility (i.e. miscarriage and infertility shocks). Here, we run the same analysis on the full sample of women, i.e. we include children of sterilized women and of women using contraceptives, in order to address potential concerns related to sample selection. Moreover, ENADID provides information on the migration status only for children cohabiting with their parents and for those temporarily absent but still considered as household members. In order to lessen the concerns with the potential selection bias this may introduce, we make a number of further sensitivity checks by changing the composition of the estimation sample.

First, Section A.1 reports both WLS and IV estimates on the full sample, i.e. including children of sterilized women and of women using contraceptives, by using miscarriage at first birth as instrument. Table A1 shows that the estimated effect of sibship size is -0.024 (s.d.=0.017) very close to our baseline estimate of -0.018 (s.d.=0.023) in Table 6.

In Section A.2 we run a sample sensitivity check by focusing on the male (sons) subsample, since according to the data boys tend to marry and hence leave their parents' household later compared to girls. In Section A.3 we focus on a sample of individuals aged 15 to 20 as a further robustness check: only few individuals are expected to be out of their origin household in this age group. Moreover, since we are able to recover migration patterns of all individuals who left in the five years prior to the survey, our measure of migration is very precise (very few individuals leave alone before age 15 and can be considered as permanent migrants) at the cost of a smaller sample size. In both cases, the estimation results are very similar to those commented in the main text, although some coefficients are less precisely estimated.

Finally, Section A.4 addresses the biases potentially generated by the exclusion from the estimation sample of parents migrated in the past and episodes of children's tied migration. We include in the estimation sample children with parents who ever migrated abroad, adding in the regressions an extra control (i.e. a dichotomous indicator) for parental migration.

Table A4 confirms the statistically insignificant effect of sibship size on child migration when endogeneity is addressed.

A.1 Sample including sterilized women and those using contraceptives

Table A1: Sibship size effect on child's 'netted migration' status: WLS and 2SLS estimates

Variables	(1) WLS	(2) 2SLS
N. siblings	0.005***	-0.024
	(0.001)	(0.017)
female	-0.031***	-0.031***
	(0.002)	(0.002)
		(0.133)
First stage — N. siblings		
miscarriage		-0.421***
		(0.084)
Angrist-Pischke $F-$statistic instrument(s)		25.33
Individual's controls	YES	YES
Mother's controls	YES	YES
Father's controls	YES	YES
Municipality indicators	YES	YES
Observations	40,008	40,008

Note. The sample includes also children of sterilized women and women using contraceptives. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

A.2 Sons

Table A2: Sibship size effect on sons' 'netted migration' status: WLS and 2SLS estimates

Variables	(1) WLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
N. siblings	0.012***	0.004	-0.051	-0.016
	(0.001)	(0.019)	(0.036)	(0.017)
IV:	—	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic		0.0355	2.567	1.315
		[0.850]	[0.109]	[0.268]
Hansen $J-$statistic				2.175
				[0.140]
First stage — N. siblings				
infertility		-0.549***		-0.547***
		(0.110)		(0.109)
miscarriage			-0.441***	-0.438***
			(0.117)	(0.117)
Angrist-Pischke $F-$statistic instrument(s)		25.13	14.19	20.38
Individual's controls	YES	YES	YES	YES
Mother's controls	YES	YES	YES	YES
Father's controls	YES	YES	YES	YES
Municipality indicators	YES	YES	YES	YES
Observations	14,777	14,777	14,777	14,777
R-squared	0.242			

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

A.3 Age group 15-20

Table A3: Sibship size effect on children's age 15-20 'netted migration' status: WLS and 2SLS estimates

Variables	(1) WLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Second stage				
N. siblings	0.005***	0.005	-0.012	-0.002
	(0.001)	(0.014)	(0.020)	(0.011)
female	-0.024***	-0.024***	-0.025***	-0.025***
	(0.003)	(0.003)	(0.003)	(0.003)
IV:	—	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic		0.146	0.377	0.257
		[0.702]	[0.539]	[0.773]
Hansen $J-$statistic				0.515
				[0.473]
First stage — N. siblings				
infertility		-0.455***		-0.452***
		(0.102)		(0.102)
miscarriage			-0.412***	-0.408***
			(0.105)	(0.105)
Angrist-Pischke $F-$statistic instrument(s)		19.78	15.43	17.82
Individual's controls	YES	YES	YES	YES
Mother's controls	YES	YES	YES	YES
Father's controls	YES	YES	YES	YES
Municipality indicators	YES	YES	YES	YES
Observations	18,707	18,707	18,707	18,707

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

A.4 Tied and parents' migration

Table A4: Sibship size effect on child's 'netted migration' status: WLS and 2SLS estimates

Variables	(1) WLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
N. siblings	0.012***	-0.000	-0.002	-0.001
	(0.001)	(0.013)	(0.019)	(0.011)
IV:	—	infertility	miscarriage	overidentified
Anderson-Rubin F-statistic		0.000680	0.0134	0.00707
		[0.979]	[0.908]	[0.993]
Hansen J-statistic				0.006
				[0.936]
First stage — N. siblings				
infertility		-0.566***		-0.560***
		(0.089)		(0.088)
miscarriage			-0.489***	-0.482***
			(0.092)	(0.092)
Angrist-Pischke $F-$statistic instrument(s)		40.66	28.24	35.06
Individual's controls	YES	YES	YES	YES
Mother's controls	YES	YES	YES	YES
Father's controls	YES	YES	YES	YES
Municipality indicators	YES	YES	YES	YES
Observations	30,977	30,977	30,977	30,977

Note. The dependent variable is 'netted migration' (see Section 3). Compared to the estimates in Tables 5 and 6 in the main text, we retain in the estimation sample also children that experienced tied migration with their parents, or whose parents had previous migration experiences. Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness, being single and a dummy for past migration experiences; father's controls include decade of birth indicators, age and age squared, years of schooling and a dummy for past migration experiences. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

B Appendix: Poverty and further identification threats

In developing countries women's infertility conditions may partly depend on material poverty, which affects women's health. Failing to control for economic conditions may represent a threat to our IV estimates because poverty is also likely to affect children's migration status. In the baseline estimates of Section 4.2 we took into account this potential threat by including some strong correlates of individual or household poverty, such as parents' educational levels, age and municipality fixed effects. In this Section, we run supplementary checks by estimating models including municipality by (ENADID) wave fixed effects and municipality by parent's education fixed effects (years of education of the most educated parent, either the mother of the father, are interacted with municipality indicators). We report both OLS and 2SLS estimates.

Table B1 shows that the OLS estimates are not sensitive to the inclusion of additional proxies for poverty and family wealth, suggesting that the income and wealth channels are unlikely to be the main sources of the OLS upward bias.

2SLS results are reported in columns (1) and (2) of Table B2, respectively. They also serve as checks of potential concerns related to the miscarriage at first pregnancy instrument, which may also be affected by women's living standards. The results confirm the robustness of our 2SLS estimates of family size effects to including alternative proxies of household poverty.

Table B1: Robustness of WLS estimates of the effect of sibship size on child's 'netted migration' status to various proxies of poverty

Variables	(1)	(2)
N. siblings	0.006***	0.006***
	(0.001)	(0.001)
Municipality×Wave indicators	YES	NO
Municipality× parents' education indicators	NO	YES
Individual's controls	YES	YES
Mother's controls	YES	YES
Father's controls	YES	YES
Municipality indicators	YES	YES
Observations	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table B2: Robustness of 2SLS estimates of the effect of sibship size on child's 'netted migration' status to various proxies of poverty

Variables	(1)	(2)
Second stage		
N. siblings	-0.009	-0.009
	(0.011)	(0.011)
IV:	overidentified	overidentified
Anderson-Rubin $F-$statistic	0.666	0.603
	[0.514]	[0.547]
Hansen $J-$statistic	0.827	0.822
	[0.363]	[0.365]
First stage — N. siblings		
infertility	-0.552***	-0.546***
	(0..092)	(0.089)
miscarriage	-0.470***	-0.439***
	(0.096)	(0.095)
Angrist-Pischke $F-$statistic instrument(s)	30.83	30.05
Municipality×Wave indicators	YES	NO
Municipality× parents' education indicators	NO	YES
Individual's controls	YES	YES
Mother's controls	YES	YES
Father's controls	YES	YES
Municipality indicators	YES	YES
Observations	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

C Appendix: Sibship size including deceased children

Table C1: Birth order effects on child's 'netted migration' status

Variables	(1)	(2)	(3)	(4)
female	-0.035***	-0.035***	-0.031***	-0.031***
	(0.003)	(0.003)	(0.005)	(0.007)
birth order	-0.017***		-0.016***	
	(0.003)		(0.003)	
birth order×female			-0.001	
			(0.001)	
birth order 2		-0.003		0.001
		(0.005)		(0.006)
birth order 3		-0.013*		-0.016*
		(0.007)		(0.008)
birth order 4		-0.031***		-0.030***
		(0.010)		(0.011)
birth order 5		-0.057***		-0.055***
		(0.012)		(0.013)
birth order 6		-0.076***		-0.070***
		(0.015)		(0.016)
birth order 7		-0.091***		-0.081***
		(0.017)		(0.018)
birth order 8		-0.120***		-0.116***
		(0.020)		(0.022)
birth order 9		-0.133***		-0.143***
		(0.023)		(0.025)
birth order 10+		-0.164***		-0.158***
		(0.027)		(0.029)
birth order 2, female				-0.010
				(0.010)
birth order 3, female				0.006
				(0.010)
birth order 4, female				-0.002
				(0.011)
birth order 5, female				-0.003
				(0.011)
birth order6, female				-0.013
				(0.012)
birth order 7, female				-0.023*
				(0.014)
birth order 8, female				-0.007
				(0.016)
birth order 9, female				0.022
				(0.019)
birth order 10+, female				-0.012
				(0.019)
age	0.022**	0.022**	0.022**	0.022**
	(0.009)	(0.009)	(0.009)	(0.009)
age squared	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Year of birth indicators	YES	YES	YES	YES
Household fixed effects	YES	YES	YES	YES
Observations	26,743	26,743	26,743	26,743
Number of households	10,139	10,139	10,139	10,139
R-squared	0.050	0.052	0.050	0.052

Note. The dependent variable is a dichotomous indicator of the child's migration status. The model is estimated using OLS. Sibship size is absorbed by household fixed effects. Standard errors clustered at the household level in parentheses. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table C2: Sibship size effect on child's 'netted migration' status: WLS estimates

Variables	(1)	(2))
N. siblings	0.010***	0.013***
	(0.001)	(0.001)
N. siblings × female		-0.007***
		(0.001)
female	-0.032***	-0.030***
	(0.003)	(0.003)
Individual's controls	YES	YES
Mother's controls	YES	YES
Father's controls	YES	YES
Municipality indicators	YES	
Weighted	YES	YES
Observations	26,743	26,743
R-squared	0.204	0.206

Note. The dependent variable is 'netted migration' (see Section 3). The model is estimated using WLS (the weights are the inverse of the standard errors of 'netted migration'). Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table C3: Sibship size effect on child's 'netted migration' status: 2SLS estimates

Variables	(1)	(2)	(3)
Second stage			
N. siblings	0.002	-0.015	-0.005
	(0.014)	(0.024)	(0.013)
female	-0.032***	-0.033***	-0.032***
	(0.003)	(0.003)	(0.003)
IV:	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic	0.0229	0.433	0.232
	[0.880]	[0.510]	[0.793]
Hansen $J-$statistic			0.419
			[0.517]
First stage — N. siblings			
infertility	-0.475***		-0.472***
	(0.095)		(0.095)
miscarriage		-0.411***	-0.407***
		(0.10)	(0.10)
Angrist-Pischke $F-$statistic instrument(s)	25.01	17.78	21.70
Individual's controls	YES	YES	YES
Mother's controls	YES	YES	YES
Father's controls	YES	YES	YES
Municipality indicators	YES	YES	YES
Observations	26,743	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table C4: Child gender and sibship size effect on child's 'netted migration' status: 2SLS estimates

Variables	(1)	(2)	(3)
Second stage			
N. siblings	0.002	-0.065	-0.006
	(0.016)	(0.057)	(0.015)
N. siblings × female	-0.001	0.105	0.007
	(0.013)	(0.084)	(0.013)
female	-0.032***	-0.059***	-0.034***
	(0.004)	(0.022)	(0.004)
IV:	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic	0.0115	1.577	0.797
	[0.989]	[0.207]	[0.527]
Hansen $J-$statistic			2.925
			[0.232]
First stage — N. siblings			
infertility	-0.557***		-0.555***
	(0.107)		(0.107)
infertility × female	0.168		0.190*
	(0.115)		(0.114)
miscarriage		-0.390***	-0.387***
		(0.113)	(0.113)
miscarriage × female		0.046	-0.046
		(0.106)	(0.102)
Angrist-Pischke $F-$statistic instrument(s)	31.59	5.74	19.49
First stage — N. siblings × female			
infertility	0.135***		0.136***
	(0.039)		(0.038)
infertility × female	-0.689***		-0.686***
	(0.134)		(0.134)
miscarriage		-0.066	-0.068
		(0.044)	(0.044)
miscarriage × female		-0.287**	-0.280**
		(0.131)	(0.130)
Angrist-Pischke $F-$statistic instrument(s)	40.51	4.55	17.22
Individual's controls	YES	YES	YES
Mother's controls	YES	YES	YES
Father's controls	YES	YES	YES
Municipality indicators	YES	YES	YES
Observations	26,743	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). Observations are weighted by the inverse of the standard error of 'netted migration' because the dependent variable is estimated. Individual's controls include year of birth indicators, age, age squared; mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. Standard errors clustered at the household level in parentheses. $P-$values are reported in brackets. [a] The number of siblings is demeaned before taking the interaction. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

D Appendix: Pyschological effects of miscarriage

As already mentioned in Section 3.2, miscarriage may be a traumatic event, creating a special bond between a mother and her children or a higher need for care, which may reduce the likelihood of offspring migration. Although we do not have measures of mother's mental health, in this section we seek to shed light on this issue.

In a recent paper van den Berg et al. (2017) show that a child's death represents one the largest losses that an individual can face and has adverse effects on parents' labor income, employment status, marital status and hospitalization. Based on that paper, we assume that a child death should produce more negative psychological effects on mothers than a miscarriage. We also assume that a miscarriage later in the pregnancy should produce more emotional distress than an early miscarriage (i.e. the intensity of the child-mother bond depends on the duration of the interrupted pregnancy). We test for 'grief' effects by including in the 2SLS regressions an indicator variable for child death and, alternatively, the duration of the interrupted pregnancy because of stillbirth as control variables.[1] The coefficients on both variables, which are reported in column (1), (2) and (4) to (7) of Table D1, respectively, are not statistically significant. Finally, in column (3) of Table D1, we leverage on the fact that we have two excluded instruments and run an overidentification test, which is based on the validity of the infertility instrument. In particular, we include the miscarriage indicator only in the second stage of a just-identified model. The coefficient on miscarriage is not statistically significant in the second stage, and suggests that it does not have a direct effect on child migration over and above the effect on family size, identified by infertility shocks. All these checks suggest that the direct effect of miscarriage on child migration is not a major issue in our analysis.

[1] In line with the medical definition, stillbirth episodes are different from miscarriages: the former refer to a loss between the sixth and the ninth month, while the latter to a loss during the first five months of pregnancy.

Table D1: Threats to identification: Effect of 'grief' on child's 'netted migration' status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Instrument:	Infertility			Miscarriage		Overidentified model	
N. children	0.004	0.004	0.004	-0.018	-0.018	-0.005	-0.005
	(0.014)	(0.014)	(0.014)	(0.023)	(0.022)	(0.012)	(0.012)
female	-0.033***	-0.033***	-0.033***	-0.033***	-0.033***	-0.033***	-0.033***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
child death (dummy)	-0.001			-0.002		-0.002	
	(0.004)			(0.004)		(0.004)	
months of stillbirth		-0.000			-0.001		-0.000
		(0.000)			(0.001)		(0.001)
miscarriage			0.010				
			(0.011)				
Individual's controls	YES	YES	YES	YES	YES	YES	YES
Mother's controls	YES	YES	YES	YES	YES	YES	YES
Father's controls	YES	YES	YES	YES	YES	YES	YES
Municipality indicators	YES	YES	YES	YES	YES	YES	YES
Observations	26,743	26,743	26,743	26,743	26,743	26,743	26,743

Note. The dependent variable is 'netted migration' (see Section 3). In this table we investigate the psychological costs of miscarriage. In column (1) we add an indicator for a child death, in column (2) we add the duration of stillbirth, and in column (3) we leverage the fact that we have two instruments and use a just-identified model in which miscarriage is included in the second stage of 2SLS and infertility status is used as the excluded instrument. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

E Appendix: Household-level estimates

Results of the household level estimates are reported in Table E1. Column (1) shows that a unit increase in the number of children is associated with an average increase in the number of migrants of 0.02 ($t = 12.3$). Column (2) reports the 2SLS estimate using the infertility instrument. The first stage shows a reduction of -0.753 ($t = -12.1$) in the total number of children per woman who experienced an infertility shock, with an $F-$statistic of 145.4. The first-stage coefficient is a bit higher in magnitude than the one obtained in the child-level estimates (-0.5), probably because of the inclusion of one-child households in the estimation. Indeed, women with only one child are those who may have suffered from more severe subfertility conditions and for whom the instrument is likely to be stronger (see Table 3 in the main text). In spite of the higher strength of the instrument, the second stage does not show any evidence of a statistically significant effect of fertility on migration. Column (3) reports the 2SLS results using the variation in the number of children generated by miscarriage. Also in this case the first-stage coefficient is highly statistically significant and negative, with an $F-$statistic of about 45. The negative impact of miscarriage on total fertility is smaller than the one exerted by infertility, yet it is quite large and precisely estimated, i.e. -0.476 ($t = -6.7$). Like for the previous instrument, also in this case no significant effect is detected in the second stage. The same happens in the overidentified model in column (4). In Table E2 we report the estimates of the same model as above while using an indicator for the household having at least one migrant child as dependent variable and results do not change.

These findings are consistent with those reported in Section 4.2, pointing to a positive correlation between family size and migration, but excluding a causal effect of the former on the latter. Also in this case, as with individual-level estimates, the larger magnitude of OLS estimates relative to the IV ones points to an upward biased estimate because of endogeneity, i.e. families more likely to send young migrants abroad tend to have more children.

Table E1: Family size effect on the number of migrants: Household-level estimates

Variables	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Second stage				
N. children	0.020***	-0.004	-0.031	-0.011
	(0.002)	(0.015)	(0.031)	(0.014)
IV:	—	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic		0.0783	1.050	0.553
		[0.780]	[0.306]	[0.575]
Hansen $J-$statistic				0.657
				[0.418]
First stage — N. children				
infertility		-0.753***		-0.750***
		(0.062)		(0.062)
miscarriage			-0.476***	-0.469***
			(0.071)	(0.071)
Angrist-Pischke $F-$statistic instrument(s)		145.4	45.05	96.20
Mother's controls	YES	YES	YES	YES
Father's controls	YES	YES	YES	YES
Municipality indicators	YES	YES	YES	YES
Observations	18,217	18,217	18,217	18,217

Note. The dependent variable is the total number of children in the household who ever migrated. Mother's controls include year of birth indicators, age and age squared, age at first pregnancy, years of schooling, indicators for mother's chronic illness and being single; father's controls include decade of birth indicators, age and age squared, years of schooling. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

Table E2: Family size effect on having at least a migrant child: Household-level estimates

Variables	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS
Second stage				
N. children	0.012***	0.001	-0.016	-0.003
	(0.001)	(0.010)	(0.020)	(0.009)
IV:	—	infertility	miscarriage	overidentified
Anderson-Rubin $F-$statistic		0.008	0.640	0.324
		[0.928]	[0.424]	[0.723]
Hansen $J-$statistic				0.580
				[0.446]
First stage — N. children				
infertility		- 0.753***		-0.750***
		(0.062)		(0.062)
miscarriage			-0.476***	-0.469***
			(0.071)	(0.071)
Angrist-Pischke $F-$statistic instrument(s)		145.42	44.95	96.17
Mother's controls	YES	YES	YES	YES
Father's controls	YES	YES	YES	YES
Municipality indicators	YES	YES	YES	YES
Observations	18,217	18,217	18,217	18,217

Note. The dependent variable is a dummy for the household having at least one migrant child. The list of control variables is the same as in Table E1. $P-$values are reported in brackets. *,** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.

References

van den Berg, G. J., P. Lundborg, and J. Vikstrm (2017). The economics of grief. *The Economic Journal 127*(604), 1794–1832.