Population Review

Volume 62, Number 1, 2023

Type: Article pp. 20-71

# The Role of Child's Gender in Fertility and Family Structure: Evidence across Countries and Centuries*

**Author: Hamid Noghanibehambari**

**Affiliation:** Center for Demography of Health and Aging (CDHA) University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706 USA

**Abstract**

Previous studies have documented that a child's gender affects future fertility and family structure across several parts of the world. This paper use a comprehensive sets of data sources to examine its heterogeneity across countries and cohorts. Using individual-level data of 516 sample-years from 98 countries that span more than 300 years, I find that, on average, families increase their number of children if their first-born child is a girl. However, this average effect contains substantial heterogeneity. A cohort analysis using data of more developed countries suggests that these effects are concentrated in periods of fertility declines, specifically late 19th century. The effects are also primarily confined to Asia, specifically southeast Asia and Middle-East. Moreover, the fertility effects are significantly stronger among low educated mothers and those residing in countries with lower gender equal norms. Some policy implications are discussed.

**Keywords:** Son Preference, Gender Bias, Gender Discrimination, Fertility, Family Planning, Family Economics, Family Structure, Sex Selection, History, Culture

**JEL Codes**: J11, J12, J13, J16, N30, Z13, D10



* The author claims that he has no conflict of interest to report. The author would like to acknowledge financial support from the Center for Demography of Health and Aging (CDHA) at the University of Wisconsin-Madison under NIA core grant P30 AG17266.

# Introduction

The gender inequality in children's outcomes and gender biasedness in parental behavior has long concerned policymakers and triggered studies in economics among other fields. Historical economic desideratum, social elements, and cultural norms have been shown to contribute to the presence and persistence of son preference.[1] Regardless of its roots, the gender-based attitudes of parents among their children is largely established in the literature. Parents engage in sex-selective abortion, sex-selective infanticide, and spend more on prenatal care for their sons. Mothers are more likely to be unmarried if they had an Ultrasound test and their child is a girl. The sex-selective behavior persists after birth. Sons receive more parental care, are longer breastfed, get more health-related resource materials, and receive more parental time during childhood.[2] The consequences of son preference persist in adulthood and have been documented to explain the gender inequality observed in education and labor market outcomes (Fernández, 2013; Kingdon, 2002).

In this paper, I complement the previous literature on son preference by providing evidence that the gender of the first child is an important determinant of fertility and family structure. In so doing, I use a comprehensive series of surveys and census data across 98 countries spanning over 300 years. The space-time depth of the data allows me to search not only for the aggregate son preference in the world but also to search for its cross-region differences and over-decades dynamics. Another important advantage of exploiting comprehensive data to study son preference is that I can explore the factors that are more likely to be correlated or have caused son preference.

Marginal effects of the pooled sample suggest that families with a first-born girl have 0.05 more children than families with a first-born boy. This effect is equivalent to a roughly 2.1 percent

[1] For a review, refer to Lundberg (2005) and Jayachandran (2015).
[2] See, for example, Almond et al. (2013), Baker and Milligan (2016), Blau et al. (2020), Dahl and Moretti (2008), Ichino et al. (2014), Kaushal and Muchomba (2018), and Muchomba and Chatterji (2020).

rise from the mean. Nonetheless, this average treatment effect is not informative as it contains substantial heterogeneity. The effects are heterogeneous across subsamples in terms of magnitude and even sign. I show that these effects are primarily confined to Asia, specifically South and Southeast Asia and middle east. Moreover, the effects are more pronounced for the period of 1850-1900, a period of sharp demographic change with unprecedented declines in fertility (Eggleston & Fuchs, 2012; Galor & Weil, 1996). Besides, I show that low educated mothers and those residing in countries with higher gender inequality have significantly stronger fertility responses to a first-born girl. However, the link between fertility outcomes and a first-child girl has become weaker for recent cohorts. The associations are also geographically clustered and more pronounced among Asian countries specifically Southeast Asia and Middle-East. I find negative effects of a first-born girl on fertility for some developed countries.

Another outcome that reflects living arrangements and family structure is living without a father. On average, having a first-born girl is associated with 41 basis points higher likelihood that mother lives without a partner/spouse/husband, off a mean of 0.117. On the contrary to fertility behavior, the absence of the father as a response to the child's gender is stronger among the developed countries with more gender-equal norms and higher human development. A historical cross-cohort analysis among a subset of developed countries suggests that the association between a child's gender and father's absence has been intensified by time and specifically for cohorts born after 1900.

The contribution of this paper to the current literature is twofold. First, on the contrary to the previous studies that focus on specific regions or limited time frames, I use a series of data that covers many countries and over three centuries. I show the dynamics of son preference over time and across cohorts. Also, the data enables me to find the countries' correlates with son preference.

This aspect of my research has not been done previously. Second, this study provides evidence on one possible root of son preference: cultural factors related to gender equality norms. Although previous studies showed that education and human development are important determinants of gender role in society (Chung and Das Gupta, 2007; Filmer et al., 2008; Nguyen and Le, 2022; Pande and Astone, 2007), no study has investigated its correlation with son preference.

Studying and documenting son preference could have important implications for family-planning policymakers and government authorities across countries. One aspect of the gender-based preference of parents appear in adverse outcomes for girls and can explain the observed gender inequalities in other outcomes such as education. Policymakers who aim at reducing the gender gap in health, education, and labor market outcomes should, at least partly, turn their focus to parents and design policies to change the momentum of sex-selective culture and gender-based behavior among parents. Also, the fact that having a daughter has the potential to influence family structure calls for welfare policies that give higher weights to single-parent families with first-born girls. Another aspect of son preference appears in sex-selective in Vitro Fertilization. With improvements in these technologies and their ever-decreasing costs, families may turn to use these methods to influence the gender of their child toward more boys. This should concern policymakers as it causes gender imbalances in the population which comes with its own consequences (Álvarez and Miles-Touya, 2019; Rask and Tiefenthaler, 2008; Strazdins and Broom, 2004).

The rest of the paper is organized as follows: a brief literature review, discussion of the data sources and sample construction, the empirical framework, discussion of the exogeneity assumption and the potential threats to this assumption. Following that, I review the main results, conduct robustness checks, and heterogeneity across subsamples. I finish with some concluding remarks in the final section.

## Literature Review

A strand of literature in economics and sociology documents the presence of son preference across countries and over generations as well as the potential consequences of son preference for girls. In a landmark study, Dahl and Moretti (2008) provide evidence of the effects of child gender on fertility and family structure in the USA. They find that the families whose first child is a girl have more children. Mothers with a first-born girl are more likely to be divorced and be a single-mother. In addition, for the married mothers who get divorced, fathers are more likely to fight for custody of sons than daughters. Blau et al. (2020) revisit son preference in the US using an updated data. They find that the effects of first-born girls on fertility disappear for the most recent cohorts among natives while the effects of living without a father are statistically significant. Almond et al. (2013) document son preference among South and East Asian immigrants to Canada. Even after being exposed to the same institutional factors and welfare codes, the variations in home countries' gender inequality opinions can explain the respective variations among immigrants' son-preference-based behavior suggesting some cultural factors at work. Choi and Hwang (2015) show that although the sex ratio at birth in South Korea has reached its natural ratio and diverged from nearby countries the parental discrimination still exists. Parents spend more time and spend more monetary and material resources on their sons than their daughters.

Tavassoli (2021) explore the effect of the gender of the first child on future fertility of serval Southeast Asian countries and find significant and large effect of a first-born girl on number of children. Hamoudi and Nobles (2014b) explore the effects of first-child gender on divorce and ask whether this observed association is driven by in-utero selection of births. They posit that unstable marriages are more likely to put mental and emotional pressure on families. The intensified tension has the potential to cause fetal death. Neonatal death is more prevalent among male fetuses than female ones (Sanders and Stoecker, 2015). They find that a significant portion of observed divorces after a firstborn girl can be explained by in-utero selections due to maternal stress.

Kaushal and Muchomba (2018) explore son preference among immigrants of East and South Asian countries using American Time Use Surveys. They find that parents spend one hour more on their male infants and toddlers compared to their female offspring. Kishore and Spears (2014b) explore the role of son preference in household energy use behavior. They show that households whose first-born is a boy are more likely to use clean fuel for cooking compared to households with a first-child girl. Jayachandran and Pande (2017) explore the role of son preference in explaining variations in height among children in India. They show that parental investment is more concentrated among boys than girls. Moreover, they document that earlier born children have an advantage over later-born children in absorbing household resources and this advantage is intensified if they are male rather than female. Ebenstein and Leung (2010) show that part of son preference in China can be explained by the tradition of sons' support for old-age parents. They take advantage of the introduction of a government pension program in rural China as a substitute for sons' support and document that son preference and sex ratios have decreased as a response to the pension program.

Son preference and sex-selective behavior can also be reflected in prenatal development behavior. Bharadwaj and Lakdawala (2013) show that parents exercise their sex-selective behavior by differential investment in prenatal care for their boy-fetus than girl-fetus. Barcellos et al. (2014)

document that boys receive more parental investment in terms of childcare time than girls. Among children in India, boys are breastfed longer and receive more vitamin supplements.

A child's gender has implications for parental investment in health-related material inputs. Palloni (2017) shows that children are healthier if their gender matches their mother's preferred gender before their birth. Jayachandran and Lleras-Muney (2009) explore the effects of life expectancy on female education and their human capital development. They show that the progress in public health technologies and advancements in medicine introduced in Sri Lanka between the years 1946 and 1953 led to sharp declines in maternal mortality and increased females' life expectancy. They document that the education of girls relative to boys increased more in regions with higher exposure to the public health enhancements than areas with lower exposure.

## Data Sources and Sample Construction

The primary data is built upon various historical and contemporaneous census and survey data across countries gathered by Minnesota Population Center (2020) (hereafter IPUMS). I add to this data a series of US decennial census data (1910-1950) extracted from Ruggles et al. (2020) and a series of March Supplement of Current Population Survey data (1976-2020) extracted from Flood et al. (2018). While the census microdata of each country provides a specific set of variables, there are essential variables that are reported in almost all censuses such as age, gender, and geographic location. IPUMS collects the data and recode all variables in a way that the variable values contain the same meaning throughout the census years of each country and across countries. For instance, countries include in their census microdata subnational regions. However, for some country-years this variable is more aggregated than others. For the US, the subnational region is state as the 1950-onward public-use censuses do not report county variables. For prior years, census data contains both county and state. IPUMS harmonize the geographic variable so that in all years it refers to the same boundary that for the US means state as state is available in all years.

Therefore, the benefit of the IPUMS-provided data is that it has a harmonized series of variables necessary for this study. Table 1 provides a list of countries and the available years for each country in the final sample. As one would observe, the sample of countries are not balanced. Some countries appear only once in the data (Myanmar 2014) and some countries for several years (Spain 1991, 2001, 2005-2019). To account for this problem, I add to our regressions a series of subnational-region fixed effects and later in our analyses a series of subnational-region-by-year fixed effects.

As evident in Table 1, the final sample covers some countries more than others. Moreover, there are many countries for which there is only recent census data. In addition, in later sections, I discuss how the effects are heterogenous with respect to education. To the extent that the results are driven by low-educated mothers, the secular over-time rises in education may induce a sample selection issue. These facts make the interpretation of the results difficult. To address this issue, I construct a weighting scheme that follows three procedures. First, The IPUMS provides the users with a person's weight variable that makes the unit of observation in the sample a representative of the population. These weights are mainly based on the weights that was originally reported by census officials of each country. However, in cases that the weights are not available, IPUMS constructs the weights based on a sampling fraction process in which each record is assigned the inverse of the ratio of sample size with respect to the original population. This IPUMS-extracted data and mentioned weighting scheme is widely used in the literature (Aaronson et al., 2021; Almond and Mazumder, 2011b; Becquet et al., 2022; Lerch, 2019, 2020; Li et al., 2021; Permanyer et al., 2014; Reher and Requena, 2014, 2020; Sobek, 2016; Tavassoli, 2021).

**Table 1 – Country-Years in the Final Sample**

| Country | Available Years | Country | Available Years | Country | Available Years |
|---|---|---|---|---|---|
| Argentina | 1970, 1980, 1991, 2001 | Iceland | 1703, 1729, 1801, 1901, 1910 | Poland | 1978, 2002 |
| Austria | 1971, 1981, 1991, 2001 | India | 1983, 1987, 1993, 1999, 2004, 2009 | Portugal | 1981, 1991, 2001, 2011 |
| Bangladesh | 1991, 2001, 2011 | Indonesia | 1971, 1976, 1980, 1985, 1990, 1995, 2000, 2005, 2010 | Puerto Rico | 1970, 1980, 1990, 2000, 2005, 2010 |
| Armenia | 2001, 2011 | Iran | 2006, 2011 | Romania | 1977, 1992, 2002, 2011 |
| Bolivia | 1976, 1992, 2001, 2012 | Iraq | 1997 | Russia | 2002, 2010 |
| Botswana | 1981, 1991, 2001, 2011 | Ireland | 1901, 1911, 1971, 1979, 1981, 1986, 1991, 1996, 2002, 2006, 2011, 2016 | Rwanda | 1991, 2002, 2012 |
| Brazil | 1960, 1970, 1980, 1991, 2000, 2010 | Israel | 1972, 1983, 1995 | Saint Lucia | 1980, 1991 |
| Myanmar | 2014 | Italy | 2001, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019 | Senegal | 1988, 2002, 2013 |
| Belarus | 1999, 2009 | Jamaica | 1982, 1991, 2001 | Sierra Leone | 2004 |
| Cambodia | 1998, 2004, 2008, 2013 | Jordan | 2004 | Vietnam | 1989, 1999, 2009 |
| Cameroon | 1976, 1987, 2005 | Kenya | 1969, 1989, 1999, 2009 | Slovenia | 2002 |
| Canada | 1891, 1901, 1911, 2011 | Kyrgyz Republic | 1999, 2009 | South Africa | 1996, 2001, 2007, 2011, 2016 |
| Chile | 1970, 1982, 1992, 2002 | Laos | 2005 | Zimbabwe | 2012 |
| China | 1982, 1990, 2000 | Lesotho | 1996, 2006 | Spain | 1991, 2001, 2005-2019 |
| Colombia | 1973, 1985, 1993, 2005 | Liberia | 2008 | South Sudan | 2008 |
| Costa Rica | 1973, 1984, 2000, 2011 | Malawi | 1987, 1998, 2008 | Sudan | 2008 |
| Cuba | 2002, 2012 | Malaysia | 1970, 1980, 1991, 2000 | Suriname | 2012 |
| Benin | 1979, 1992, 2002, 2013 | Mali | 1987, 1998, 2009 | Sweden | 1880, 1890, 1900, 1910 |
| Denmark | 1787, 1801 | Mauritius | 1990, 2000, 2011 | Switzerland | 1970, 1980, 1990, 2000 |
| Dominican Republic | 1981, 2002, 2010 | Mexico | 1970, 1990, 1995, 2000, 2010, 2015 | Thailand | 1970, 1980, 1990, 2000 |
| Ecuador | 1974, 1982, 1990, 2001, 2010 | Mongolia | 1989, 2000 | Togo | 1960, 1970, 2010 |
| El Salvador | 1992, 2007 | Morocco | 1982, 1994, 2004, 2014 | Trinidad and Tobago | 1970, 1980, 2000, 2011 |
| Ethiopia | 1984, 1994, 2007 | Mozambique | 1997, 2007 | Turkey | 1985, 1990, 2000 |
| Fiji | 1976, 1986, 1996, 2007, 2014 | Nepal | 2001, 2011 | Uganda | 1991, 2002, 2014 |
| France | 1962, 1968, 1975, 1982, 1990, 1999, 2006, 2011 | Nicaragua | 1971, 1995, 2005 | Egypt | 1986, 1996, 2006 |
| Palestine | 1997, 2007, 2017 | Nigeria | 2006, 2007, 2008, 2009, 2010 | United Kingdom | 1851, 1861, 1871, 1881, 1891, 1901, 1911, 1991 |
| Ghana | 2000, 2010 | Norway | 1801, 1865, 1875, 1900, 1910 | Tanzania | 1988, 2002, 2012 |
| Greece | 1971, 1981, 1991, 2001 | Pakistan | 1973, 1998 | United States | 1850, 1860, 1870, 1880, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1968-2020 |
| Guatemala | 1964, 1973, 1981, 1994, 2002 | Panama | 1960, 1970, 1980, 1990, 2000, 2010 | Burkina Faso | 1996, 2006 |
| Guinea | 1983, 1996, 2014 | Papua New Guinea | 1980, 1990, 2000 | Uruguay | 1963, 1975, 1985, 1996, 2006, 2011 |
| Haiti | 1971, 1982, 2003 | Paraguay | 1962, 1972, 1982, 1992, 2002 | Venezuela | 1971, 1981, 1990, 2001 |
| Honduras | 1974, 1988, 2001 | Peru | 1993, 2007 | Zambia | 1990, 2000, 2010 |
| Hungary | 1970, 1980, 1990, 2001, 2011 | Philippines | 1990, 1995, 2000, 2010 | | |

Second, I weight each survey by its population within a country (relative to other surveys) at the time of the survey in order to make each survey representative within a given country. Third, I weight each country by its population in the year 2000. This makes each country representative within the final sample.

The data reports the age and sex of children as well as some information regarding age, sex, marital status, education, and labor force participation of parents. I restrict the sample to women as the primary outcomes, fertility and absence of the father, are related only to women. Furthermore, I restrict the sample to mothers with at least one child and that the age of their eldest child in the household is less than 12 as children tend to move outside of the households after this age. Since I am only able to locate the first child only if the child is present in the sample unit, this sample selection is important.

In addition, I restrict the sample to mothers for whom the first child can be located in the household. This means that those children whose mothers' location in the household (provided by IPUMS) does not match the person number of any mother in the household are eliminated. I also restrict the sample to mothers aged 18-45.[3] A minimum age restriction is necessary to establish the fertility effects which may appear later in life. Maximum age is also necessary as older mothers may have had births early in youth and their child could not be present in the household. I drop observations for which the marital status is missing or unknown.[4] The final sample consists of 391 census-years, 125 survey-years (a total of 516 sample-years), 98 countries, spanning 318 years (1703-2020), and including 77,184,698 observations. Figure 1 illustrates the geographic

---

[3] Restricting the sample to women reduces the sample size to 51 percent of its initial size. Among women, 31.5 percent have at least one child and that her children can technically be located in the household. In the children-mothers merged sample, restrictions on first-child's age reduces the sample size by 47 percent. In the latter sample, 89.6 percent of mothers satisfy age restrictions that I imposed on mothers. Overall, from about 1.14 billion observations in the initial IPUMS-extracted sample, the final sample suggests a 93 percent size reduction due to these restrictions.
[4] The latter selection omits less than 0.1 percent of the observations in the initial sample.

distribution of countries in the final sample. It covers all countries in Central and North America and most countries in South America as well as Asia and Europe. However, it only partially covers countries in Africa and Oceania. Table 1 reports the list of countries and years in the final sample.

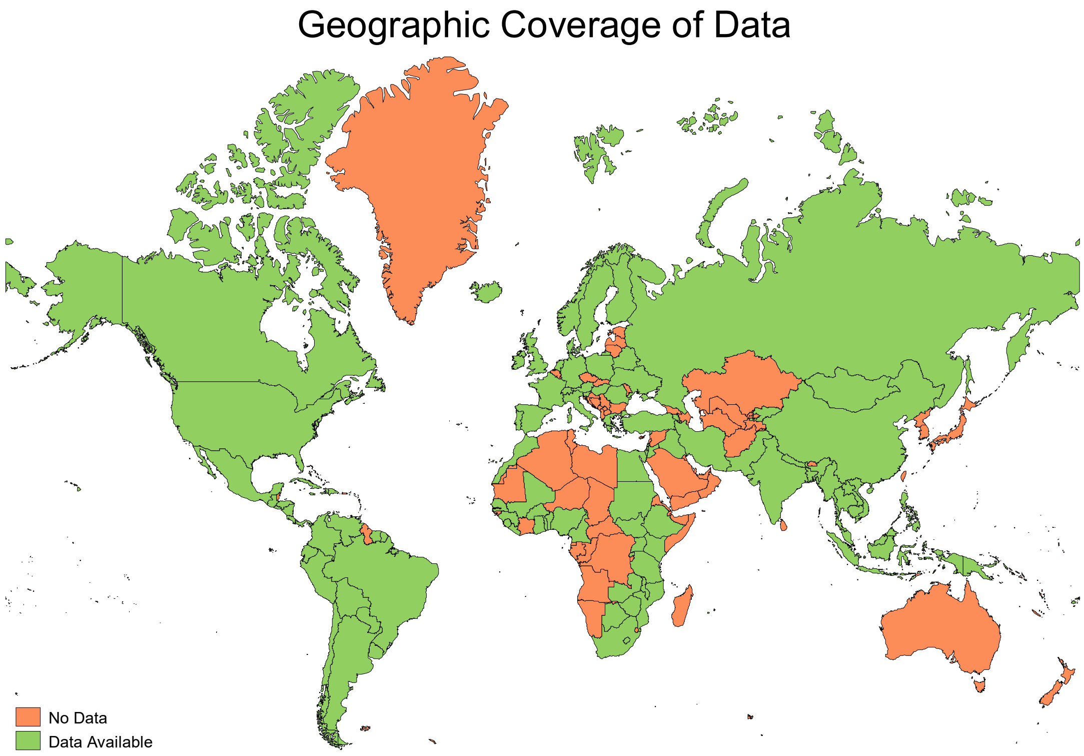



**Figure 1 - Geographic Distribution of Countries Present in the Final Sample**

Table 1 reports summary statistics of the final sample. On average, the number of children in the household is 2.17. Note that number of children is top-coded to 9 by the IPUMS. While only 0.3 percent of women have 9-and-above children I also use several dummy variables to capture the fertility of mothers. These dummies take a value of 1 if mothers have at least 2, 3, and 4 children and zero otherwise. On average, 65.1 percent of mothers have two or more children. Among those mothers with non-missing data for education and labor force, roughly 23.2 percent of mothers have more than secondary education and 33.8 percent are active in the labor force. Another important outcome related to family structure is the absence of the father. This variable is constructed using a combination of two variables. First, using the marital status of women, I assign a value of 1 to "father absent" if the mother is divorced, separated, single, never married, or spouse is absent and a value of 0 if the mother is married or in-union with a spouse. Second, if the spouse cannot be located in the household unit at the time of observation, I also assign a value of 1 to "father absent". I drop observations in analysis related to "father absent" for which these two methods provide contradictory results or I cannot infer the presence of father using either of these two methods. In the final sample, this elimination leads to exclusion of only 1.9 percent of observations. On average, 11.7 percent of children live in the absence of their father.

**Table 2 - Summary Statistics**

| Variable | Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| ***Outcomes:*** | | | | | |
| Father Absent | 75647676 | 0.1174 | 0.3219 | 0 | 1 |
| Number of Children | 77184698 | 2.1766 | 1.225 | 1 | 9 |
| Number of Children≥ 2 | 77184698 | 0.6505 | 0.4768 | 0 | 1 |
| Number of Children≥ 3 | 77184698 | 0.3098 | 0.4624 | 0 | 1 |
| Number of Children≥ 4 | 77184698 | 0.137 | 0.3438 | 0 | 1 |
| ***Parental Characteristics:*** | | | | | |
| Age | 77184698 | 29.6105 | 6.0032 | 12 | 45 |
| Birth Cohort | 77184698 | 1939.4218 | 51.7157 | 1658 | 2004 |
| Age of First Child | 77184698 | 6.4336 | 3.4524 | 1 | 12 |
| Gender of First Child | 77184698 | 0.4908 | 0.4999 | 0 | 1 |
| Dwelling Type: Owned | 77184698 | 0.414 | 0.4925 | 0 | 1 |
| Dwelling Type: Rented | 77184698 | 0.1888 | 0.3913 | 0 | 1 |
| Dwelling Type: Missing | 77184698 | 0.3973 | 0.4893 | 0 | 1 |
| School Attending | 77184698 | 0.0238 | 0.1524 | 0 | 1 |
| School Attended in the Past | 77184698 | 0.1668 | 0.3728 | 0 | 1 |
| Education: Less than Secondary | 77184698 | 0.5191 | 0.4996 | 0 | 1 |
| Education: More than Secondary | 77184698 | 0.2321 | 0.4222 | 0 | 1 |
| Education: Missing | 77184698 | 0.2488 | 0.4323 | 0 | 1 |
| Employment Status: Employed | 77184698 | 0.2832 | 0.4505 | 0 | 1 |
| Employment Status: Missing | 77184698 | 0.3344 | 0.4718 | 0 | 1 |
| Labor Force Status: Active | 77184698 | 0.3382 | 0.4731 | 0 | 1 |
| Labor Force Status Missing | 77184698 | 0.116 | 0.3202 | 0 | 1 |
| ***Country-Level Characteristics:*** | | | | | |
| Average GDP per Capita | 77184698 | 19514.246 | 19836.431 | 233.9442 | 88432.617 |
| Average Gender Inequality Index | 76888780 | 0.3056 | 0.1653 | 0.025 | 0.725 |
| Average Maternal Mortality Rate | 77108613 | 82.6256 | 113.406 | 2 | 917 |
| Average Education less than Secondary | 76924387 | 66.3824 | 22.7834 | 6.1 | 100 |
| Average Female Labor Force Participation Rate | 77139781 | 53.4013 | 11.4065 | 11.6 | 83.9 |

In the robustness of the results by subsample, I also use average cross-sectional country-level characteristics including Gender Inequality Index extracted from the United Nation's Human Development Reports (2020).[5] The Gender Inequality Index (GII) is an inequality index that is constructed based on three female outcomes: 1) maternal mortality ratio and teenage fertility. 2) Female empowerment, the share of female parliamentary seats, and female education. 3) Female labor force participation rate. In general, it captures gender equity in economic, social, health, and human development contexts. While the average GII in the sample is 0.31 it can be as low as 0.03 for Switzerland and as high as 0.73 for Papua New Guinea. Figure 2 shows the geographic distribution of countries based on their quartile rank in cross-country GII distribution in the final sample. Africa, South and Southeast Asia, Middle-East, and South America have the highest GII while most countries in Europe and North America have the lowest GII.

## Empirical Strategy

To show the association between the first-child's gender on future fertility and family structure, I use different specifications of the following form:

$$y_{irt} = \alpha FG_{it} + \beta X_{it} + \xi_r + \zeta_t + \gamma_{rt} + \varepsilon_{irt} \quad (1)$$

Where $y$ is the mother $i$ in country-specific subnational-region $r$ observed at time $t$. The parameter $FG$ is the first-child girl dummy that equals1 if the first child of the mother is a girl and zero

[5] One may truly argue that the cultural proxies such as gender inequality index could vary over time and since the coverage of data is different than the coverage of these proxies, the heterogeneity of the effects are confounded. We use the country characteristics, specifically gender inequality index, for the years 2015-2020 for two reasons. First, the data availability becomes very strict for previous decades specifically for developing countries. Second, as noted in the previous studies (Fernández, 2010; Fernandez and Fogli, 2009; Fernández and Fogli, 2006), culture evolves slowly and have momentum across generations. This fact helps me arguing that countries with higher gender inequality index today quite likely had higher gender gap decades ago. This persistence of culture is well established in economics of culture literature and specifically in culture and immigration literature (Alesina et al., 2013; Alesina & Giuliano, 2011; Gentili et al., 2017; Noghanibehambari et al., 2022).

otherwise. In vector $X$, I include limited information on mother's and household's observable characteristics including mother's education (less than secondary, more than secondary, and a

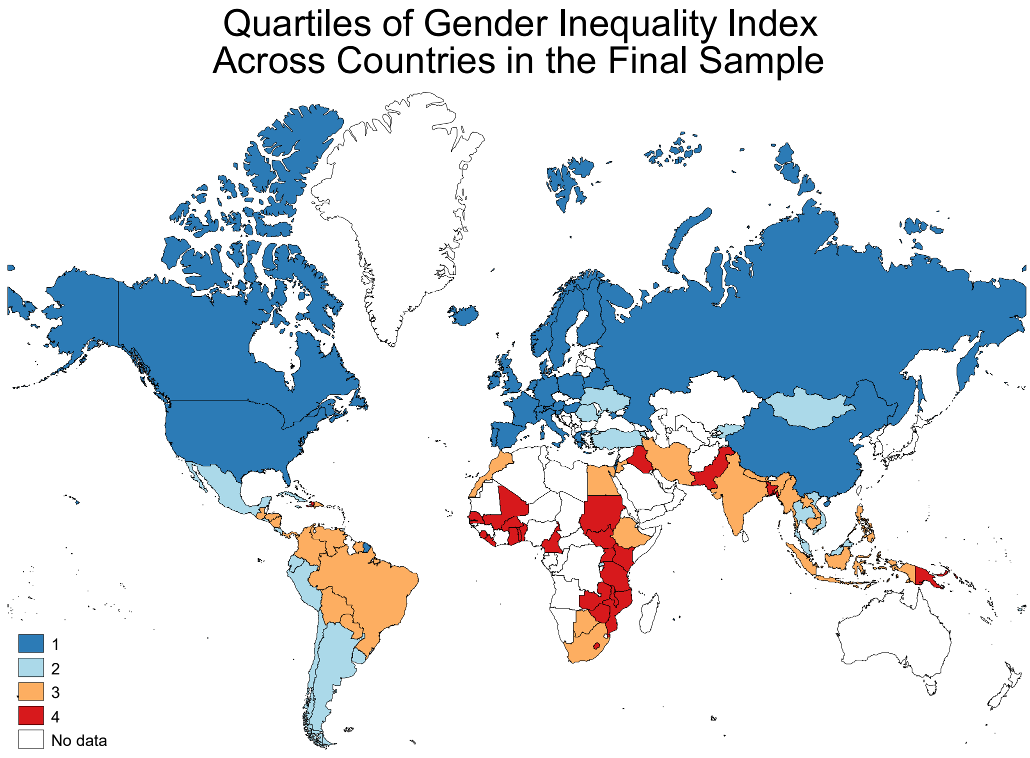



**Figure 2 - Geographic Distribution of Countries in the Sample by Quartiles of Gender Inequality Index**

missing indicator for education), employment status (employed and a missing indicator for employment), labor force status (active and a missing indicator for labor force), a cubic function of age, school attending (attending, never attended, and a missing indicator for school attendance), and dwelling type (owned, rented, and a missing indicator for dwelling type).[6] The parameters $\xi$

[6] Missing values can be related to two sources. First, there are observations within each census-year that contain missing information on these variables. Second, there are census-years that do not provide the information at all. For instance, the following country-years do not provide information on education: Canada 1852-1901; Denmark 1787-1801; Iceland 1703-1910; Ireland 1901-19860; Netherland 1960-2011; Norway 1865-1910; Sweden 1880-1910; UK 1851-2001, US 1850-1910. Similarly, the following countries do not report information on labor force activity: Botswana 1981; Canada 1852, China 2000; Ethiopia 1984-1994; Fiji 2004; Germany 1971, 1981; Hungary 1970-1980; Indonesia 2000, 2005; Ireland 1901, 19011; Israel 1983; Kenya 1964, 1979; Liberia 1974; Mexico 1960, 2005; Mongolia 1989; Netherland 1960, 1971; Philippines 1995, 200-2001; Poland 1978, 1988; Thailand 1970-2000.

and $\zeta$ represent subnational-region[7] and year fixed effects, respectively. In the preferred specification, I also include subnational-region-by-year fixed effects to control for variations in socioeconomic features of the subnational-region of the mother's residence that also vary by time. The standard errors are clustered at the subnational-region-level.[8] All regressions are weighted using a weighting scheme described in section 0. Finally, $\varepsilon$ is a disturbance term.

In this specification, $\alpha$ is the coefficient of interest. I focus on only the first-child for two important reasons. First, the choice of fertility for second-and-above birth order is endogenous. For instance, families have son preference and choose to go for the second child only if the first child is a girl. The first-born girl may also leave mothers without a present husband which in turn lowers her material resources necessary for the antenatal period. This path has the potential to influence the gender of the second child as the health environment is shown to marginally affect the sex ratio (Almond and Mazumder, 2011; Jürges, 2015; Tavassoli et al., 2020). The second reason is that the previous mechanism (that first-born girl affects the presence of spouse) also affects fertility since the fertility rates are higher among married mothers than unmarried mothers and that unmarried mothers have, on average, lower fertility. This feature is established in the literature and confirmed with the data (Blau et al., 2020; Bumpass et al., 1978; Dahl and Moretti, 2008; Nitsche and Hayford, 2020; Schultz, 1994).[9] Therefore, using the sex composition of children as the main explanatory variable causes biased estimations and generates sample selection issues.

[7] Region is the most disaggregated subnational region provided by IPUMS. For instance, region is state in the US and county in Sweden. Throughout the paper, I use region and subnational region interchangeably. Later, I disaggregate the sample based on world region. World region is totally different than the region variable in equation 1 and consists of several countries.

[8] Appendix C shows the robustness to alternative clustering levels as well as using robust standard errors.

[9] In my final sample, being married is associated with 0.294 (SE=0.013) more children, conditional on other covariates and a full set of fixed effects explained in equation 1.

The empirical method compares the fertility and family structure of mothers with a first-born girl to mothers with a first-born boy over the years. The underlying assumption in the identification strategy is that the first-child's gender is uncorrelated with other determinants of fertility and family structure. This assumption is the benchmark assumption in the literature. (Barcellos et al., 2014; Bharadwaj and Lakdawala, 2013; Blau et al., 2020; Dahl and Moretti, 2008; Palloni, 2017). However, I discuss three potential threats to the identification strategy below.

First, parents may engage in sex selection even before birth in various ways including sex-selective abortion, sex-selective infanticide, and sex-selective in Vitro Fertilization (IVF). If mothers with a preference for sons practice sex selection in the presence of a girl-fetus they are left out of the sample. In the absence of sex-selection, they would continue their fertility as a response to their first-born girl and boosts the estimated coefficient ($\alpha$) while the fact that these mothers select themselves out of the maternity ward understates the true effects. Previous studies suggest that sex-selection fertility is of lower importance for first births (Almond and Edlund, 2008; Blau et al., 2020). In Appendix A, I explore this source of endogeneity. I investigate whether the sex ratios (male/female) are within the empirical ranges of normal sex ratios among humans (1.02-1.07 male to female or equivalently 0.483-0.495 share of females in total births) suggested by the literature (Chao et al., 2019; Lobel et al., 1993; Orzack et al., 2015). The results show that there are no aggregate sex ratio changes from the normal sex ratios. The sex ratios and their 90 percent confidence intervals for both samples of all women and married women lie within the empirical ranges suggested by the literature. This fact holds even after controlling for subnational-region-by-year fixed effects and subnational-region-by-year fixed effects interacted with birth cohort fixed effects. As a further step, I check whether there is a clustering of the first-child girl for specific birth cohorts or specific countries. Figure 3 shows the average of the first-child girl for each birth cohort and country. For visualization purposes, I have censored the data to include the most recent birth cohorts (1920-2000). Visually, no birth cohort or country contains a higher ratio of the first-child girl compared to the rest. In the years 1980-2000, there are more sample-years that contain few numbers of observations. In this figure, each unit is birth-cohort-by-country. For instance, I observe some reductions from the average of sex ratio for birth cohorts of 1996 in Armenia. Indeed, sex ratio of these cohorts drop to 0.33. However, this unit of observation contains only 96 individuals. Therefore, this deviation from average is probably caused only by measurement error due to sample size and not by sex selection in our sample.

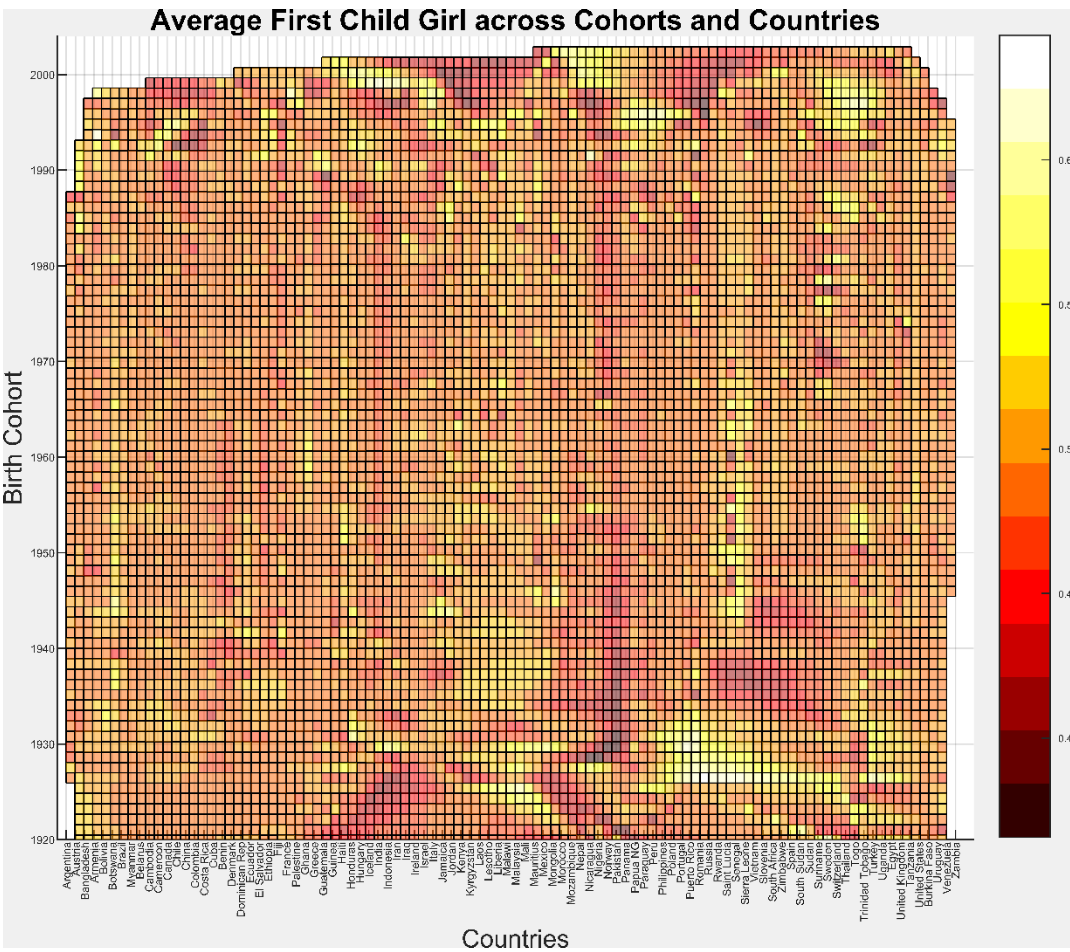



**Figure 3 - Visual Distribution of Average First Child girl across Countries and Birth Cohorts**

Second, some studies suggest that the prenatal environment and socioeconomic status of families are also correlated with the gender of the child. For instance, Almond and Mazumder (2011) show that Muslim mothers whose pregnancy period overlaps with the holy month of Ramadan, during which they restrain from eating and drinking between sunrise and sunset, are more likely to give birth to daughters. The main reason is that male embryos are more vulnerable to environmental factors and health conditions during prenatal development and are more likely to die before birth while girl fetuses have higher chances of survival. Among other factors that marginally influence sex ratio, and generally infants' health and mortality, is pollution (Sanders and Stoecker, 2011; Tavassoli et al., 2020), malnutrition (Andersson and Bergström, 2013), stress (Navara, 2010; Torche and Kleinhaus, 2011), war (Polasek et al., 2005), marital status (Almond and Edlund, 2008; Khoury et al., 1984), government policies (Bhaskar, 2011; Ebenstein, 2010; Noghanibehambari, 2022), and economic indicators (Chahnazarian, 1988; Grech, 2018; Noghanibehambari and Salari, 2020). Although I include a set of subnational-region-by-year fixed effects to control for all subnational-regional economic and demographic characteristics that may vary by time as well as a full set of observable characteristics of mothers indicating their socioeconomic status, I control for these endogenous covariates only partially as the unmeasured and unobserved mothers' characteristics that determine their fertility could also affect the gender of their first child in ways that are unobservable. In Table 3, I explore this endogeneity issue by regressing a series of observable mother's characteristics on First-Child Girl. There is no statistical evidence that the age of motherhood is correlated with the gender of the first child. Importantly, education is uncorrelated with the first-child's gender. Also, there is no evidence that the first-child's gender varies across birth cohorts. In addition, there is no difference in chances of a first-child girl among those observations who report information on education and labor market and those who do not. However, there are some statistical associations for mother's school attending and first child girl. Note that only 2.4 percent of mothers attended school at the time of observation. Mother's labor force participation is also correlated with the first-child's gender. However, the literature on the effects of female labor force participation on birth outcomes including child sex is limited and mixed. For instance, Wüst (2015) shows that mothers who work during pregnancy have lower risks of preterm birth but he did not find an effect for *small for gestational age*. On the contrary, Rossin (2011) documents that maternity leave acts could slightly improve birth outcomes including birth weight. There are mainly two mechanisms in work. First, mothers who work have higher family income and more resources during prenatal development which raises the likelihood of having a boy. On the other hand, they may experience more stress and anxiety in the workplace which in turn increase the probability of a girl. Note that the effects (positive or negative) that are documented in the literature are only residual and no study directly estimated the effects of maternal employment on child sex. However, I should claim that to the extent that employment affects the fertility and family structure the results could suffer from endogeneity.

Third, the absence of father and fertility are two outcomes that are not jointly independent. In fact, unmarried mothers have lower fertility rates and mothers with more children are more likely to be partnered.[10] Therefore, when we look into the fertility effect of the first-born girl for a sample that contains all women without considering their marital status we should be aware that we are averaging the effects among married and unmarried mothers, and that the marginal effects are probably smaller among unmarried mothers since they lack the necessary resources to respond to a first-child girl by increasing their fertility.

[10] For similar literature in this relationship, see Autor et al. (2019a), Bumpass et al. (1978), Cummins (2013), Dribe et al. (2014), Kearney and Wilson (2018), and Schultz (1994). Also, see footnote 9.

**Table 3 - First Child Girl and Mothers' Observable Characteristics**

| | ***Outcomes, Mother's Characteristics:*** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Age | School Attending | Education: Less than Secondary | Education: More than Secondary | Education: Missing | Labor Force Status: Active | Labor Force Status: Missing | Birth Cohort |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| First Child Girl | -0.008 | 0.00006 | -0.00606* | 0.00638* | -0.00032 | 0.00528*** | 0.00188 | 0.008 |
| | (0.04797) | (0.00018) | (0.00342) | (0.00363) | (0.00021) | (0.00176) | (0.00138) | (0.04797) |
| Observations | 76735244 | 76735244 | 76735244 | 76735244 | 76735244 | 76735244 | 76735244 | 76735244 |
| R-squared | 0.09593 | 0.02869 | 0.23817 | 0.2349 | 0.28017 | 0.13129 | 0.52545 | 0.82407 |
| Subnational-Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Subnational-Region-by-Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*

**** p<.01, ** p<.05, * p<.1*

To make this potential bias intuitive, note that married mothers produce more babies mostly because they have more financial support and material resources available from their husbands. Also, having a first-born girl diminishes the likelihood of a present spouse and staying married. What we observe is that these mothers have a first-born girl and reduce their fertility not because they lack son preference but because they do not have enough resources as otherwise would have had if their partners had chosen to stay with them. This fact generates a downward bias in the estimations of the first-child's gender on fertility. To partly address this endogeneity, while I show the results for the full sample of all women, I also show the results for the sample excluding unmarried mothers. The results confirm this downward bias and show larger marginal effects and percentage changes for the sample of married mothers.

# Results

## Main Results

Figure 4 shows the distribution of the number of children among mothers who have a first-born boy and those with a first-born girl. There is a visual distinction between these two groups. Families whose first child is a boy have, on average, 1.89 children (top panel) while families whose first child is a girl have, on average, 1.90 children (bottom panel). However, this visual difference is not informative as there could be various sources of bias and sample selection to underestimate/overestimate the true differences. To address this issue, we need to control for place and cohort unobserved confounders by including fixed effects, and to make the sample representative by including the weighting scheme. Specifically, I use different specifications of equation 1. Table 4 shows the main results for five outcomes including a continuous variable capturing the total number of children, dummies for having at least 2, 3, and 4 children, and a dummy to capture the father's absence. For each outcome, two columns report the regression results for specifications with/without subnational-region-by-year fixed effects. Conditional on a full set of fixed effects, family controls, and subnational-region-by-year interactions, having a first-child girl is associated with 0.05 more children, equivalent to a 2.4 percent rise from the mean number of children (column 2).

Next, I explore the effect of a first-born girl on another outcome that is related to family structure in the household, i.e., whether the father is absent in the household or not. Having a first-born girl is associated with a 41 basis points higher likelihood of living without a father, equivalent to a 3.5 percent increase from the mean.

To put these numbers into perspective, I apply the marginal effects to the world birth demographics in the year 2000.[11] Assuming that about 132.2 million births occurred in the year 2000 and that, following the sample demographics, about 28.6 percent are firstborn and the average first-child girl is 0.49, it can be inferred that about 391,000 more birth occurred in the world (in post-2000 years) as a response to child's gender of those born in the year 2000.[12] Using the same numbers in combination with the marginal effects of Table 4, one can estimate that a lower bound of about 600,000 children born in 2000 will live without a father because of their gender being a girl.

[11] I used the estimations provided by Our World in Data (2020) webpage.

[12] This number indicates that families who their first child was born in the year 2000 and that their child was a girl may move to the next child because of having a first-born girl. Their decision to have children in future reflects son-preference-induced fertility behavior. The causality in this statement (that they move to the next child as a response to the first-child's gender) comes from the regression results. Roughly 28.6 percent of births were first born. This adds the number of first-borns to be 37.8 million. Among these births, 49 percent were girls. This suggests there were 18.5 million first-born girls. The results of column 2 of Table 3 suggests an increase of 2.1 percent from the mean. Extrapolating this value to the sample of first-born girls in 2000, one can get a fertility rise of 391K.

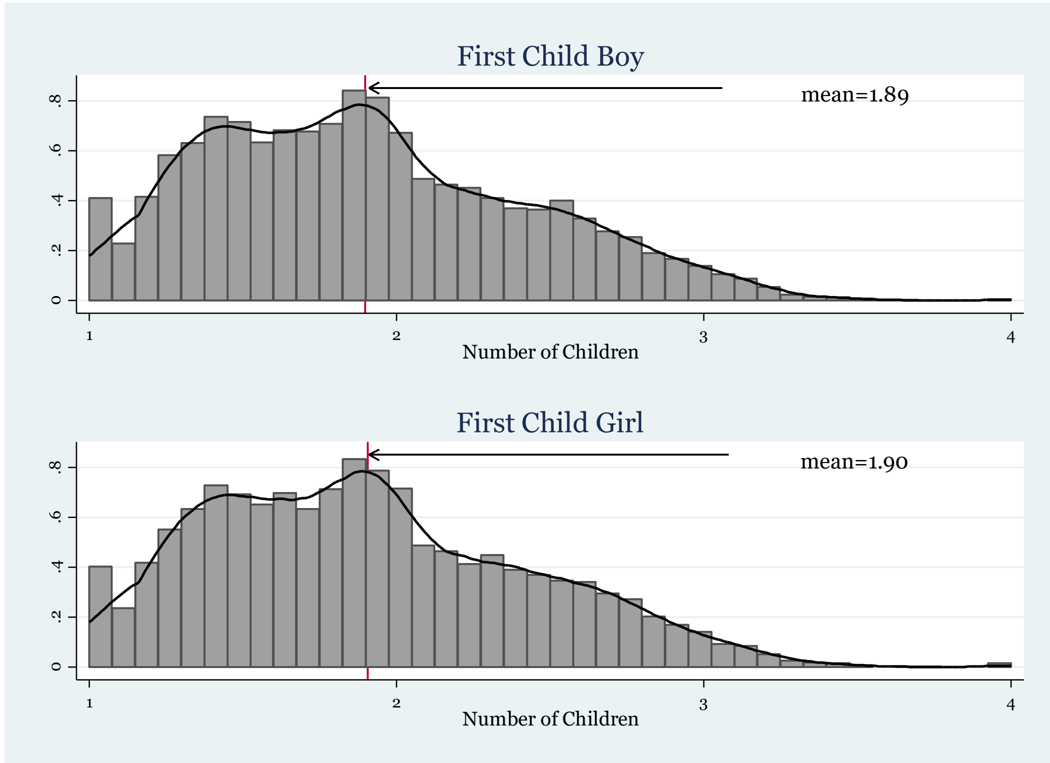



**Figure 4 - Distribution of Number of Children for Mothers with First Child Boy (Top) and First Child Girl (Bottom)**

**Table 4 - The Effects of First Child Girl on Fertility and Family Structure among All Women in the Final Sample**

| | Number of Children | | Father Absent | |
|---|---|---|---|---|
| | (1) | (2) | (9) | (10) |
| First Child Girl | 0.052*** | 0.0514*** | 0.0041*** | 0.0041*** |
| | (0.0097) | (0.0097) | (0.0012) | (0.0013) |
| Observations | 76735229 | 76735229 | 75198209 | 75198209 |
| R-squared | 0.2133 | 0.2218 | 0.1587 | 0.1696 |
| Mean DV | 2.176 | 2.176 | 0.117 | 0.117 |
| Subnational-Region FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Subnational-Region-by-Year FE | No | Yes | No | Yes |

*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*
**** p<.01, ** p<.05, * p<.1*

These fertility effects are aggregated across countries and centuries but interestingly in line with other studies that focus on limited geographical areas in limited time windows (Almond et al., 2013; Altindag, 2016; G. Andersson et al., 2006; Chai Bin Park and Nam-Hoon Cho, 1995; Chowdhury and Bairagi, 1990; Das Gupta et al., 2003; Guilmoto, 2012; Hank and Kohler, 2000; Larsen et al., 1998; Mitra, 2014; Sandström and Vikström, 2015). For instance, Almond et al. (2013) show that among first-generation Southeast Asian immigrants in Canada having a first child girl is associated with 2.2 percentage points higher probability of having a second child.

Since the first-child gender also affects the probability of being married and married mothers tend to have higher fertility for reasons such as having more resources and financial support, the fertility effects of a first-child girl could be underestimating true effects as the first-born-girl mothers have lower fertility because they are unmarried rather than not having son preference. To explore this potential confounding issue, I replicate the fertility regressions for the sample of married mothers. The results are reported in Table 5. The marginal effects and percentage changes are comparable the main results. For instance, the marginal effect of the number of children is 0.057 (compare with 0.052) and its percentage change relative to the mean is 2.6 percent (compare with 2.4).

## Alterative Specifications

In Table 6, I reexamine the main results across six alternative specifications: 1) Including only sub-national region and year fixed effects. 2) Adding country-by-year quadratic trend to control for common country-specific trends such as cultural changes and social reforms over time.

## Table 5 - The Effect of First Child Girl on Fertility among Married Mothers

| | ***Outcome:*** Number of Children | |
|---|---|---|
| | (1) | (2) |
| First Child Girl | 0.05824*** | 0.05767*** |
| | (0.01049) | (0.01046) |
| Observations | 65657764 | 65657162 |
| R-squared | 0.21234 | 0.22135 |
| Mean DV | 2.234 | 2.234 |
| Subnational-Region FE | No | Yes |
| Year FE | Yes | Yes |
| Controls | Yes | Yes |
| Subnational-Region-by-Year FE | No | Yes |

*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*

**** p<.01, ** p<.05, * p<.1*

3) Including only subnational-region-by-year fixed effects. 4) Combining specification 1 and 3 while adding country-by-birth-cohort fixed effects.[13] In this setting, the variation comes from within subnational-region-year and within country-cohort changes in the first-child's gender. Except for time-varying covariates that are controlled by subnational-region-year fixed effects, this column controls for unobserved characteristics of mothers within a country who were born at the same year and probably were exposed to the same set of country-specific social and economic conditions throughout their life. 5) Adding to specification 4 a series of year-by-birth-cohort fixed effects. This full set of fixed effects control for unobservable characteristics of each subnational-region that vary by year, each birth cohort within each country, and common (world-wide) features of birth cohorts in each year. 6) Finally, I add parental controls to specification 5 to further control for potential confounders.

The marginal effects across all models are very similar to the main results. For instance, the marginal effect of fertility is 0.051 in the full specification of column 6 which is almost identical to the main results. The first-born girl raises the likelihood of living without a father by 43 basis points that is also very similar to the marginal effect in the main results (41 basis points change). Besides, all the effects remain statistically significant at 1 percent level.

[13] From specification 3-onward, I do not include country-by-year trend as it will be absorbed by region-by-year fixed effects.

**Table 6 - Robustness of the Effects of First Child Girl on Fertility and Family Structure to Alternative Specifications**

| | Alternative Specifications | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ***Panel A. Number of Children*** | | | | | | |
| First Child Girl | 0.0462*** | 0.0461*** | 0.0458*** | 0.0497*** | 0.0508*** | 0.0518*** |
| | (0.01) | (0.01) | (0.0099) | (0.0096) | (0.0095) | (0.0095) |
| Observations | 76735229 | 76735229 | 76735229 | 76735195 | 76735179 | 76735179 |
| R-squared | 0.0952 | 0.0972 | 0.1036 | 0.1995 | 0.2236 | 0.2341 |
| Mean DV | 2.1766 | 2.1766 | 2.1766 | 2.1766 | 2.1766 | 2.1766 |
| | | | | | | |
| ***Panel E. Father Absent*** | | | | | | |
| First Child Girl | 0.0043*** | 0.0043*** | 0.0042*** | 0.0043*** | 0.0043*** | 0.0043*** |
| | (0.0012) | (0.0012) | (0.0013) | (0.0013) | (0.0013) | (0.0012) |
| Observations | 75198209 | 75198209 | 75198209 | 75198176 | 75198160 | 75198160 |
| R-squared | 0.1448 | 0.1522 | 0.1562 | 0.1746 | 0.1791 | 0.1891 |
| Mean DV | 0.1173 | 0.1173 | 0.1173 | 0.1173 | 0.1173 | 0.1173 |
| | | | | | | |
| Subnational-Region FE | Yes | Yes | No | Yes | Yes | Yes |
| Year FE | Yes | Yes | No | Yes | Yes | Yes |
| Country-by-Year Quadratic Trend | No | Yes | No | No | No | No |
| Subnational-Region-by-Year FE | No | No | Yes | Yes | Yes | Yes |
| Country-by-Cohort FE | No | No | No | Yes | Yes | Yes |
| Year-by-Cohort FE | No | No | No | No | Yes | Yes |
| Controls | No | No | No | No | No | Yes |

*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*

**** p<.01, ** p<.05, * p<.1*

## Heterogeneity

**Heterogeneity by Gender Inequality**. We start the heterogeneity analysis by showing how the effects vary by counties' Gender Inequality Index (GII). In so doing, I interact the right-hand side of equation 1 with the country-level GII. The results are reported in Table 7. Note that the main effect of GII is absorbed by subnational-region fixed effects. The interaction terms imply that the marginal effects of the first-child girl on fertility outcomes are indeed larger for countries with higher GII. Among mothers with a first-born girl compared to mothers with a first-born boy, an increase of a standard deviation of GII across countries (a rise of 0.16 units) is associated with 0.033 more children. However, column 2 implies that the gender equity increases the responsiveness of the father's absence to a first-child girl. While these effects seem counterintuitive, they are in line with some recent studies. For instance, Blau et al. (2020) explore the effect of a first-child girl on fertility and living without a father on natives and immigrants in the US. They find that among natives, those who presumably were exposed to higher gender equity norms, having a first-child girl is associated with a higher likelihood of living without a father but is correlated with negative effects on fertility. Among immigrants and specifically those from less gender-equal societies, there is no statistical association between the first-child girl and living without a father while there are strong and large associations for fertility outcomes.

**Table 7 - The Interaction Effects of First Child Girl and Gender Inequality Index on Fertility and Family Structure**

| | ***Outcomes:*** | |
|---|---|---|
| | Number of Children | Father Absent |
| | (1) | (2) |
| First Child Girl | -0.001 | 0.0031** |
| | (0.0071) | (0.0013) |
| First Child Girl×GII | 0.209*** | -0.0015 |
| | (0.0296) | (0.0032) |
| Observations | 56129809 | 56049246 |
| R-squared | 0.2456 | 0.1842 |
| Mean DV | 2.041 | 0.141 |
| Subnational-Region FE | No | Yes |
| Year FE | Yes | Yes |
| Controls | Yes | Yes |
| Subnational-Region-by-Year FE | No | Yes |

*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*
**** p<.01, ** p<.05, * p<.1*

**Heterogeneity by Education**. Female education partly reflects the gender norms and cultural elements. The gender gap in education between boys and girls has been associated with gender-based opinions and gender inequality index (Blau et al., 2013; Dhar et al., 2018; Fernández, 2010; Fernandez and Fogli, 2009; Lundborg et al., 2018). Therefore, one would expect son preference to vary by mother's education. To explore this source of heterogeneity, I interact a dummy for low education –that equals one if the mother has less than secondary education and zero if at least a secondary education- with the right-hand side covariates of equation 1. The results are reported in Table 8. Among low educated mothers compared to high educated mothers, having a first-child girl is associated with 0.03 more children. Besides, high educated mothers who have a first-born girl are more likely to live without a spouse compared to low educated mothers, although the interaction effect is insignificant. I should mention that the gender inequality index, by its construction, includes measures of female education and human capital. In the sample, 66 percent of high educated mothers live in countries with low GII and 67 percent of low educated mothers live in countries with high GII. Therefore, it is not surprising to see higher marginal effects of father absence for high educated individuals.

**Table 8 - The Interaction Effects of First Child Girl and Mother's Being Low Educated on Fertility and Family Structure**

| | ***Outcomes:*** | |
|---|---|---|
| | Number of Children | Father Absent |
| | (1) | (2) |
| First Child Girl (FCG) | 0.0322*** | 0.0051* |
| | (0.0115) | (0.0028) |
| Low Education (LE) | 0.3085*** | 0.0117*** |
| | (0.0206) | (0.0031) |
| FCG×LE | 0.0283* | -0.0014 |
| | (0.0154) | (0.0029) |
| Observations | 57528403 | 57465066 |
| R-squared | 0.2232 | 0.169 |
| Mean DV | 2.176 | 0.117 |
| Subnational-Region FE | No | Yes |
| Year FE | Yes | Yes |
| Controls | Yes | Yes |
| Subnational-Region-by-Year FE | No | Yes |

*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*
**** p<.01, ** p<.05, * p<.1*

**Heterogeneity by World Regions**. Gender norms vary substantially in different world regions of the world. Specifically, studies on son preference reveal various results and sometimes mixed evidence among different regions such as the USA (Abrevaya, 2009; Blau et al., 2020; Dahl and Moretti, 2008), UK, Italy, and Sweden (Ichino et al., 2014), Denmark, Norway, and Sweden (Andersson et al., 2006), South Korea (Kashyap and Villavicencio, 2016; Yoo et al., 2017), China (Das Gupta et al., 2003; Sun et al., 2019), India (Bharadwaj and Lakdawala, 2013; Muchomba and Chatterji, 2020), Japan (Yamamura, 2013), Pakistan (Javed and Mughal, 2019), etc. These differences suggest potential heterogeneity by region. To explore this heterogeneity, I divide the sample into 8 country-groups as follows: Southeast Asia, Middle-East and North Africa (MENA), Other Asia, Europe, South and Central America, North America, Oceania, and

Africa. The results are reported in Table 9. Regression results are reported for subsamples in columns and for outcomes in panels. Asian countries and specifically Southeast Asia reveal the largest fertility effects. The fertility response of Southeast Asian mothers implies a percentage change of 6.8 percent relative to the mean which is considerably larger than the implied percentage change of the aggregate effect in the main results (2.4 percent). The second largest impacts are driven by countries in Middle-East and North Africa (MENA). The effects across Europe, North America, Oceania, and Africa are statistically insignificant. The marginal effect of South America is statistically significant but economically very small, suggesting a 0.16 percent change (compare with 2.4 percent in the main results).

For the father-absent outcome, countries in Oceania show a negative response to a first-child girl although statistically insignificant. For other regions, the marginal effects are positive and statistically significant with the largest effects among mothers in North America and Africa.

**Further Analysis**. To examine the transition and dynamics in son preference over time and its heterogeneity across birth cohorts, I split the final sample into two country-groups. First, I explore the cross-cohort heterogeneity for countries with historical available census data in the sample. These countries include Canada, United States, United Kingdom, Denmark, Iceland, Ireland, Norway, Sweden, and Germany. The second group is the rest of the countries for which only contemporaneous (1960-onwards) census and survey data are available. The problem in pooling all birth cohorts is the sample selection issue discussed below. The countries with historically available data have lower gender inequality index and higher education. As we have seen, the lower GII is associated with different marginal effects in terms of magnitude for fertility outcomes and in terms of coefficient's sign for living without a father outcome. For instance,

**Table 9 - Heterogeneity of the Effects of First Child Girl on Fertility and Family Structure across World Regions**

| | *Subsamples by World Regions:* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Southeast Asia | MENA | Other Asia | Europe | South America | North America | Oceania | Africa |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ***Panel A. Number of Children*** | | | | | | | | |
| First Child Girl | 0.1355*** | 0.0618*** | 0.0507*** | -0.0024 | 0.0036** | -0.0021 | 0.0113 | -0.0033 |
| | (0.0062) | (0.0054) | (0.0066) | (0.0022) | (0.0017) | (0.0035) | (0.0212) | (0.014) |
| Observations | 13157499 | 4103556 | 3118830 | 20837342 | 17954670 | 10417298 | 143631 | 6960468 |
| R-squared | 0.2367 | 0.3192 | 0.2263 | 0.1498 | 0.213 | 0.1398 | 0.1679 | 0.1618 |
| Mean DV | 1.905 | 2.377 | 2.108 | 2.222 | 2.178 | 2.271 | 2.246 | 2.309 |
| ***Panel E. Father Absent*** | | | | | | | | |
| First Child Girl | 0.0014* | 0.0028*** | 0.003** | 0.0031** | 0.0051*** | 0.0048*** | -0.0008 | 0.0071** |
| | (0.0008) | (0.0006) | (0.0013) | (0.0015) | (0.0014) | (0.0018) | (0.0038) | (0.0029) |
| Observations | 13154113 | 4103345 | 3118722 | 20817436 | 17901953 | 8972660 | 143014 | 6945031 |
| R-squared | 0.0296 | 0.0265 | 0.0488 | 0.1007 | 0.1446 | 0.1457 | 0.429 | 0.1601 |
| Mean DV | 0.039 | 0.037 | 0.044 | 0.117 | 0.143 | 0.139 | 0.257 | 0.235 |
| Subnational-Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Subnational-Region-by-Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

*Notes. MENA stands for Middle-East and North Africa. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*

**** p<.01, ** p<.05, * p<.1*

pooling recent birth cohorts from countries with higher GII with earlier birth cohorts from low GII countries underestimates the effects of father-absent not because of a change in recent cohorts' behavior but as a result of more cohorts joining the sample from high GII countries for whom there is no association between living without a spouse and a first-child girl.

I show the results for countries with historically available data (nine countries mentioned above) in Appendix Table B-1 and for all other countries (1960-onward data) in Appendix Table B-2. As suggested by previous analyses of this section, the fertility effects are mostly insignificant for countries with historically available data, except for the period of 1850-1900 during which european countries experienced sharp reductions in fertility (Coale, 2017). However, the effects on father-absent are positive and similar in magnitude until 1950. It becomes substantially larger for the period of 1950-2000, suggesting a percentage change of about 2.5 percent.

For the rest of the countries, the marginal effects of fertility are positive for cohorts born after 1920 although imprecisely estimated in some cases. The effects become smaller for recent cohorts (1980-2000). For instance, the implied percentage change from the mean is 4.4, 3.3, and 0.7 for the number of children and cohorts born between the years 1940-1960, 1960-1980, and 1980-2000, respectively. For the outcome father absent, the evidence is mixed. For birth cohorts of 1940-1960, the marginal effects are negative while for cohorts of 1960-2000 they become positive and significant.

In the main results, I cluster the standard errors at the sub-national region level. I explore other clustering levels as well as using Hubert-White robust standard errors in Appendix Table C-1. The results are, without exception, statistically significant at conventional levels.

# Conclusion

There a relatively large literature to establish son preference and its implication for their children. This gender preference can be observed in a wide array of outcomes and areas. Parents may reveal their son preference and gender biased behavior by discriminating in prenatal health care use (Bharadwaj and Lakdawala, 2013; Buckles and Kolka, 2014), fuel type use in the house (Kishore and Spears, 2014), postnatal health investment (Jayachandran and Pande, 2017), future fertility (Dahl and Moretti, 2008; Jayachandran, 2015), time spent with children (Kaushal and Muchomba, 2018; Lundberg, 2005a), educational investment (Hatlebakk, 2017), and their health endowment (Mitra, 2014a; Muchomba and Chatterji, 2020). These types of discriminatory behavior widen the gender gap in future outcomes such as education, health, and labor force participation (Chao et al., 2019; Milazzo, 2018; Sun et al., 2019). Therefore, an important policy concern is the presence and dynamics of son preference among households especially if the society is biased toward a specific gender. From a policymaker's point of view, it is essential to understand the magnitude of parental son preference, its geographic differences, the potential elements that may affect this cultural norm, and its dynamics over time. This study aimed to do so.

This paper explored the effect of the first-child's gender on future fertility decisions and family structure across countries. To this purpose, I employed IPUMS-extracted census and survey data consisting of 516 sample-years, including 98 countries, and spanning over 300 years. Operating under the assumption that the gender of the first child is likely random, I found that having a first-child girl is associated with higher future fertility and an increase in the likelihood of living without a father. On average and for the pooled sample, mothers with a first-born girl have 0.052 more children, a 2.4 percent rise from the mean of the number of children. Also, having a first-born girl is associated with 41 basis points higher probability of living without a spouse, off a mean of 0.12.

While the effects appear to be quite robust in magnitude and significance across a wide range of specifications, they reveal substantial heterogeneity across subsamples. I show that these effects are primarily confined to Asia, specifically South and Southeast Asia and middle east. Moreover, the effects are more pronounced for the period of 1850-1900, a period of sharp demographic change with unprecedented declines in fertility. In addition, for the fertility outcomes, the effects are more pronounced among mothers with low education and those residing in countries with higher gender inequality index. On the contrary to these findings, having a first-born girl reduces fertility for countries in Europe and North America. On the other hand, the results of living without a father are more pronounced for the developed countries, countries with more equal gender norms, and among high educated mothers.

## References


Aaronson, D., Dehejia, R., Jordan, A., Pop-Eleches, C., Samii, C., & Schulze, K. (2021). The Effect of Fertility on Mothers' Labor Supply over the Last Two Centuries. *The Economic Journal*, *131*(633), 1–32. https://doi.org/10.1093/EJ/UEAA100

Abrevaya, J. (2009). Are there missing girls in the United States? Evidence from birth data. *American Economic Journal: Applied Economics*, *1*(2), 1–34. https://doi.org/10.1257/app.1.2.1

Alesina, A., & Giuliano, P. (2011). Family ties and political participation. *Journal of the European Economic Association*, *9*(5), 817–839.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, *128*(2), 469–530.

Almond, D., & Edlund, L. (2008). Son-biased sex ratios in the 2000 United States Census. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5681–5682. https://doi.org/10.1073/pnas.0800703105

Almond, D., Edlund, L., & Milligan, K. (2013). Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada. *Population and Development Review*, *39*(1), 75–95. https://doi.org/10.1111/j.1728-4457.2013.00574.x

Almond, D., & Mazumder, B. (2011). Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, *3*(4), 56–85. https://doi.org/10.1257/app.3.4.56

Altindag, O. (2016). Son Preference, Fertility Decline, and the Nonmissing Girls of Turkey. *Demography*, *53*(2), 541–566. https://doi.org/10.1007/s13524-016-0455-0

Álvarez, B., & Miles-Touya, D. (2019). Gender imbalance in housework allocation: a question of time?

*Review of Economics of the Household*, *17*(4), 1257–1287. https://doi.org/10.1007/s11150-019-09467-w

Andersson, G., Hank, K., Rønsen, M., & Vikat, A. (2006). Gendering family composition: Sex preferences for children and childbearing behavior in the Nordic countries. *Demography*, *43*(2), 255–267. https://doi.org/10.1353/dem.2006.0010

Andersson, G., Hank, K., & Vikat, A. (2007). Understanding parental gender preferences in advanced societies: Lessons from Sweden and Finland. *Demographic Research*, *17*, 135–156. https://doi.org/10.4054/DEMRES.2007.17.6

Andersson, R., & Bergström, S. (2013). Is Maternal Malnutrition Associated with a Low Sex Ratio at Birth? *Human Biology*, *70*(6). https://digitalcommons.wayne.edu/humbiol/vol70/iss6/10

Autor, D., Dorn, D., & Hanson, G. (2019). When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men. *American Economic Review: Insights*, *1*(2), 161–178. https://doi.org/10.1257/aeri.20180010

Baker, M., & Milligan, K. (2016). Boy-girl differences in parental time investments: Evidence from three countries. *Journal of Human Capital*, *10*(4), 399–441. https://doi.org/10.1086/688899

Barcellos, S. H., Carvalho, L. S., & Lleras-Muney, A. (2014). Child gender and parental investments in india: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, *6*(1 A), 157–189. https://doi.org/10.1257/app.6.1.157

Becquet, V., Sacco, N., & Pardo, I. (2022). Disparities in Gender Preference and Fertility: Southeast Asia and Latin America in a Comparative Perspective. *Population Research and Policy Review 2021*, 1–29. https://doi.org/10.1007/S11113-021-09692-1

Bharadwaj, P., & Lakdawala, L. K. (2013). Discrimination begins in the womb: Evidence of sex-selective prenatal investments. *Journal of Human Resources*, *48*(1), 71–113. https://doi.org/10.3368/jhr.48.1.71

Bhaskar, V. (2011). Sex selection and gender balance. *American Economic Journal: Microeconomics*, *3*(1), 214–244. https://doi.org/10.1257/mic.3.1.214

Blau, F. D., Kahn, L. M., Brummund, P., Cook, J., & Larson-Koester, M. (2020). Is there still son preference in the United States? *Journal of Population Economics*, *33*(3), 709–750. https://doi.org/10.1007/s00148-019-00760-7

Blau, F. D., Kahn, L. M., Liu, A. Y.-H., & Papps, K. L. (2013). The transmission of women's fertility, human capital, and work orientation across immigrant generations. *Journal of Population Economics*, *26*(2), 405–435.

Buckles, K., & Kolka, S. (2014). Prenatal investments, breastfeeding, and birth order. *Social Science & Medicine*, *118*(C), 66–70. https://doi.org/10.1016/J.SOCSCIMED.2014.07.055

Bumpass, L. L., Rindfuss, R. R., & Jamosik, R. B. (1978). Age and marital status at first birth and the pace of subsequent fertility. *Demography*, *15*(1), 75–86. https://doi.org/10.2307/2060491

Chahnazarian, A. (1988). Determinants of the sex ratio at birth: review of recent literature. In *Social biology* (Vol. 35, Issues 3–4, pp. 214–235). Taylor & Francis Group . https://doi.org/10.1080/19485565.1988.9988703

Chai Bin Park, & Nam-Hoon Cho. (1995). Consequences of son preference in a low-fertility society: imbalance of the sex ratio at birth in Korea. *Population & Development Review*, *21*(1), 59–84. https://doi.org/10.2307/2137413

Chao, F., Gerland, P., Cook, A. R., & Alkema, L. (2019). Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(19), 9303–9311. https://doi.org/10.1073/pnas.1812593116

Choi, E. J., & Hwang, J. (2015). Child gender and parental inputs: No more son preference in Korea? *American Economic Review*, *105*(5), 638–643. https://doi.org/10.1257/aer.p20151118

Choi, E. J., & Hwang, J. (2020). Transition of Son Preference: Evidence From South Korea. *Demography*, *57*(2), 627–652. https://doi.org/10.1007/s13524-020-00863-x

Chowdhury, M. K., & Bairagi, R. (1990). Son preference and fertility in Bangladesh. *Population & Development Review*, *16*(4), 749–757. https://doi.org/10.2307/1972966

Chung, W., & Das Gupta, M. (2007). *Why is Son Preference Declining in South Korea? The Role of Development and Public Policy, and the Implications for China and India*. https://doi.org/10.1596/1813-9450-4373

Coale, A. J. (2017). *The decline of fertility in Europe* (Vol. 4900). Princeton University Press.

Cummins, N. (2013). Marital fertility and wealth during the fertility transition: rural France, 1750-1850. *The Economic History Review*, *66*(2), 449–476. https://doi.org/10.1111/j.1468-0289.2012.00666.x

Dahl, G. B., & Moretti, E. (2008). The Demand for Sons. *Review of Economic Studies*, *75*(4), 1085–1120. https://doi.org/10.1111/j.1467-937X.2008.00514.x

Das Gupta, M., Jiang, Z., Li, B., Xie, Z., Woojin, C., & Bae, H. O. (2003). Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *Journal of Development Studies*, *40*(2), 153–187. https://doi.org/10.1080/00220380412331293807

Dhar, D., Jain, T., & Jayachandran, S. (2018). Intergenerational Transmission of Gender Attitudes: Evidence from India. *The Journal of Development Studies*, *55*(12), 1–21. https://doi.org/10.1080/00220388.2018.1520214

Dribe, M., Hacker, J. D., & Scalone, F. (2014). The impact of socio-economic status on net fertility during the historical fertility decline: A comparative analysis of Canada, Iceland, Sweden, Norway, and the USA. *Population Studies*, *68*(2), 135–149. https://doi.org/10.1080/00324728.2014.889741

Ebenstein, A. (2010). The “missing girls” of China and the unintended consequences of the one child policy. *Journal of Human Resources*, *45*(1), 87–115. https://doi.org/10.3368/jhr.45.1.87

Ebenstein, A., & Leung, S. (2010). Son Preference and Access to Social Insurance: Evidence from China’s Rural Pension Program. *Population and Development Review*, *36*(1), 47–70. https://doi.org/10.1111/j.1728-4457.2010.00317.x

Eggleston, K. N., & Fuchs, V. R. (2012). The New Demographic Transition: Most Gains in Life Expectancy Now Realized Late in Life. *Journal of Economic Perspectives*, *26*(3), 137–156. https://doi.org/10.1257/JEP.26.3.137

Fernández, R. (2010). *Does culture matter?*

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *The American Economic Review*, *103*(1), 472–500.

Fernandez, R., & Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, *1*(1), 146–177.

Fernández, R., & Fogli, A. (2006). Fertility: The role of culture and family experience. *Journal of the European Economic Association*, *4*(2–3), 552–561.

Filmer, D., Friedman, J., The, N. S., & Bank, W. (2008). *Development, Modernization, and Son Preference in Fertility Decisions*. https://doi.org/10.1596/1813-9450-4716

Flood, S., King, M., Ruggles, S., Warren, J. R., Rodgers, R., Ruggles, S., & Warren, J. R. (2018). Integrated public use microdata series, current population survey: Version 5.0. [dataset]. *Minneapolis: University of Minnesota*. https://doi.org/10.18128/D030.V5.0.

Galor, O., & Weil, D. N. (1996). The Gender Gap, Fertility, and Growth. *American Economic Review*, *86*(3), 374–387. https://doi.org/10.2307/2118202

Gangadharan, L. (2003). Testing for Son Preference in South Africa. *Journal of African Economics*, *12*(3), 371–416. https://doi.org/10.1093/jae/12.3.371

Gentili, E., Masiero, G., & Mazzonna, F. (2017). The role of culture in long-term care arrangement decisions. *Journal of Economic Behavior and Organization*, *143*, 186–200. https://doi.org/10.1016/j.jebo.2017.09.007

Grech, V. (2018). Correlation of sex ratio at birth with health and socioeconomic indicators. *Early Human Development*, *118*, 22–24. https://doi.org/10.1016/j.earlhumdev.2018.02.002

Guilmoto, C. Z. (2012). Son Preference, Sex Selection, and Kinship in Vietnam. *Population and Development Review*, *38*(1), 31–54. https://doi.org/10.1111/j.1728-4457.2012.00471.x

Hamoudi, A., & Nobles, J. (2014). Do Daughters Really Cause Divorce? Stress, Pregnancy, and Family Composition. *Demography*, *51*(4), 1423–1449. https://doi.org/10.1007/s13524-014-0305-x

Hank, K., & Kohler, H.-P. (2000a). Gender Preferences for Children in Europe. *Demographic Research*, *2*. https://doi.org/10.4054/demres.2000.2.1

Hank, K., & Kohler, H.-P. (2000b). Gender Preferences for Children in Europe. *Demographic Research*, *2*. https://doi.org/10.4054/demres.2000.2.1

Hatlebakk, M. (2017). Son Preference, Number of Children, Education and Occupational Choice in Rural Nepal. *Review of Development Economics*, *21*(1), 1–20. https://doi.org/10.1111/RODE.12237

Human Development Reports. (2020). *Gender Inequality Index (GII) | Human Development Reports*. http://hdr.undp.org/en/content/gender-inequality-index-gii

Ichino, A., Lindström, E. A., & Viviano, E. (2014). Hidden consequences of a first-born boy for mothers. *Economics Letters*, *123*(3), 274–278. https://doi.org/10.1016/j.econlet.2014.03.001

Javed, R., & Mughal, M. (2019). Have a Son, Gain a Voice: Son Preference and Female Participation in Household Decision Making. *Journal of Development Studies*, *55*(12), 2526–2548. https://doi.org/10.1080/00220388.2018.1516871

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics*, *7*(1), 63–88. https://doi.org/10.1146/annurev-economics-080614-115404

Jayachandran, S., & Lleras-Muney, A. (2009). Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines *. *Quarterly Journal of Economics*, *124*(1), 349–397. https://doi.org/10.1162/qjec.2009.124.1.349

Jayachandran, S., & Pande, R. (2017). Why Are Indian Children So Short? The Role of Birth Order and Son Preference. *American Economic Review*, *107*(9), 2600–2629. https://doi.org/10.1257/AER.20151282

Jürges, H. (2015). Ramadan fasting, sex-ratio at birth, and birth weight: No effects on Muslim infants born in Germany. *Economics Letters*, *137*, 13–16.

Kashyap, R., & Villavicencio, F. (2016). The Dynamics of Son Preference, Technology Diffusion, and Fertility Decline Underlying Distorted Sex Ratios at Birth: A Simulation Approach. *Demography*, *53*(5), 1261–1281. https://doi.org/10.1007/s13524-016-0500-z

Kaushal, N., & Muchomba, F. M. (2018). Missing time with parents: son preference among Asians in the USA. *Journal of Population Economics*, *31*(2), 397–427. https://doi.org/10.1007/s00148-017-0668-6

Kearney, M. S., & Wilson, R. (2018). Male earnings, marriageable men, and nonmarital fertility: Evidence from the fracking boom. *Review of Economics and Statistics*, *100*(4), 678–690. https://doi.org/10.1162/rest_a_00739

Khoury, M. J., Erickson, J. D., & James, L. M. (1984). Paternal effects on the human sex ratio at birth: Evidence from interracial crosses. *American Journal of Human Genetics*, *36*(5), 1103–1111. /pmc/articles/PMC1684514/?report=abstract

Kingdon, G. G. (2002). The gender gap in educational attainment in India: How much can be explained? *Journal of Development Studies*, *39*(2), 25–53. https://doi.org/10.1080/00220380412331322741

Kishore, A., & Spears, D. (2014). Having a Son Promotes Clean Cooking Fuel Use in Urban India: Women's Status and Son Preference. *Economic Development and Cultural Change*, *62*(4), 673–699. https://doi.org/10.1086/676330

Larsen, U., Chung, W., & Gupta, M. Das. (1998). Fertility and son preference in Korea. *Population Studies*, *52*(3), 317–325. https://doi.org/10.1080/0032472031000150496

Lerch, M. (2019). Regional variations in the rural-urban fertility gradient in the global South. *PLOS ONE*, *14*(7), e0219624. https://doi.org/10.1371/JOURNAL.PONE.0219624

Lerch, M. (2020). International Migration and City Growth in the Global South: An Analysis of IPUMS Data for Seven Countries, 1992–2013. *Population and Development Review*, *46*(3), 557–582.

https://doi.org/10.1111/PADR.12344
Li, C., Zhou, Z., & Lumey, L. H. (2021). Early-life exposure to the Chinese famine and tuberculosis risk: Unrecognized biases from different measures of famine intensity. *Proceedings of the National Academy of Sciences*, *118*(16). https://doi.org/10.1073/PNAS.2102809118
Lobel, S. M., Pomponio, R. J., & Mutter, G. L. (1993). The sex ratio of normal and manipulated human sperm quantitated by the polymerase chain reaction. *Fertility and Sterility*, *59*(2), 387–392. https://doi.org/10.1016/s0015-0282(16)55682-9
Lundberg, S. (2005). Sons, Daughters, and Parental Behaviour. *Oxford Review of Economic Policy*, *21*(3), 340–356. https://doi.org/10.1093/oxrep/gri020
Lundborg, P., Nordin, M., & Rooth, D. O. (2018). The intergenerational transmission of human capital: the role of skills and health. *Journal of Population Economics*, *31*(4), 1035–1065. https://doi.org/10.1007/s00148-018-0702-3
Makhlouf Obermeyer, C. (1996). Fertility norms and son preference in Morocco and Tunisia: does women's status matter? *Journal of Biosocial Science*, *28*(1), 57–72. https://doi.org/10.1017/S0021932000022082
Marcén, M., Molina, J. A., & Morales, M. (2018). The effect of culture on the fertility decisions of immigrant women in the United States. *Economic Modelling*, *70*, 15–28.
Milazzo, A. (2014). *Son Preference, Fertility and Family Structure: Evidence from Reproductive Behavior among Nigerian Women*. The World Bank. https://doi.org/10.1596/1813-9450-6869
Milazzo, A. (2018). Why are adult women missing? Son preference and maternal survival in India. *Journal of Development Economics*, *134*, 467–484. https://doi.org/10.1016/j.jdeveco.2018.06.009
Minnesota Population Center. (2020). Integrated Public Use Microdata Series, International: Version 7.3 [dataset]. *Minneapolis, MN: IPUMS*. https://doi.org/https://doi.org/10.18128/D020.V7.2
Mitra, A. (2014a). Son Preference in India: Implications for Gender Development. In *Journal of Economic Issues* (Vol. 48, Issue 4, pp. 1021–1037). Routledge. https://doi.org/10.2753/JEI0021-3624480408
Mitra, A. (2014b). Son Preference in India: Implications for Gender Development. In *Journal of Economic Issues* (Vol. 48, Issue 4, pp. 1021–1037). Routledge. https://doi.org/10.2753/JEI0021-3624480408
Muchomba, F. M., & Chatterji, S. (2020). Disability among children of immigrants from India and China: Is there excess disability among girls? *Population Studies*, *74*(2), 263–281. https://doi.org/10.1080/00324728.2020.1762911
Navara, K. J. (2010). Programming of offspring sex ratios by maternal stress in humans: Assessment of physiological mechanisms using a comparative approach. In *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* (Vol. 180, Issue 6, pp. 785–796). Springer. https://doi.org/10.1007/s00360-010-0483-9
Nguyen, M., & Le, K. (2022). Maternal education and son preference. *International Journal of Educational Development*, *89*, 102552. https://doi.org/10.1016/J.IJEDUDEV.2022.102552
Nitsche, N., & Hayford, S. R. (2020). Preferences, Partners, and Parenthood: Linking Early Fertility Desires, Marriage Timing, and Achieved Fertility. *Demography*, *57*(6), 1975–2001. https://doi.org/10.1007/s13524-020-00927-y
Noghanibehambari, H. (2022). Intergenerational health effects of Medicaid. *Economics & Human Biology*, *45*, 101114. https://doi.org/10.1016/J.EHB.2022.101114
Noghanibehambari, H., & Salari, M. (2020). Health benefits of social insurance. *Health Economics*, *29*(12), 1813–1822. https://doi.org/10.1002/hec.4170
Noghanibehambari, H., Tavassoli, N., & Noghani, F. (2022). Intergenerational Transmission of Culture Among Second-and-Higher Generation Immigrants: the Case of Age at First Birth and Nonmarital Childbirth. *Journal of Economics, Race, and Policy 2022*, 1–18. https://doi.org/10.1007/S41996-022-00103-X
Orzack, S. H., Stubblefield, J. W., Akmaev, V. R., Colls, P., Munné, S., Scholl, T., Steinsaltz, D., & Zuckerman, J. E. (2015). The human sex ratio from conception to birth. *Proceedings of the National*

*Academy of Sciences of the United States of America*, *112*(16), E2102–E2111. https://doi.org/10.1073/pnas.1416546112

*Our World in Data*. (2020). https://ourworldindata.org/grapher/births-and-deaths-projected-to-2100

Palloni, G. (2017). Childhood health and the wantedness of male and female children. *Journal of Development Economics*, *126*, 19–32. https://doi.org/10.1016/j.jdeveco.2016.11.005

Pande, R. P., & Astone, N. M. (2007). Explaining son preference in rural India: The independent role of structural versus individual factors. *Population Research and Policy Review*, *26*(1), 1–29. https://doi.org/10.1007/S11113-006-9017-2/TABLES/5

Permanyer, I., Esteve-Palos, A., Garcia, J., & Mccaa, R. (2014). Human Development Index-like Small Area Estimates for Africa Computed from IPUMS-International Integrated Census Microdata. *Journal of Human Development and Capabilities* , *16*(2), 245–271. https://doi.org/10.1080/19452829.2014.956300

Polasek, O., Kolcic, I., Kolaric, B., & Rudan, I. (2005). Sex ratio at birth and war in Croatia (1991-1995). *Human Reproduction*, *20*(9), 2489–2491. https://doi.org/10.1093/humrep/dei097

Rask, K., & Tiefenthaler, J. (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, *27*(6), 676–687. https://doi.org/10.1016/j.econedurev.2007.09.010

Reher, D., & Requena, M. (2014). The mid-twentieth century fertility boom from a global perspective. *The History of the Family* , *20*(3), 420–445. https://doi.org/10.1080/1081602X.2014.944553

Reher, D., & Requena, M. (2020). Revisiting mid-twentieth-century fertility shifts from a global perspective. *Population Studies*, 299–314. https://doi.org/10.1080/00324728.2020.1783454

Rossin, M. (2011). The effects of maternity leave on children's birth and infant health outcomes in the United States. *Journal of Health Economics*, *30*(2), 221–239. https://doi.org/10.1016/j.jhealeco.2011.01.005

Ruggles, S., Flood, S., Goeken, R., Grover, J., & Meyer, E. (2020). IPUMS USA: Version 10.0 [dataset]. *Minneapolis, MN: IPUMS*. https://doi.org/10.18128/D010.V10.0

Sanders, N. J., & Stoecker, C. (2015). Where have all the young men gone? Using sex ratios to measure fetal death rates. *Journal of Health Economics*, *41*, 30–45. https://doi.org/10.1016/J.JHEALECO.2014.12.005

Sanders, N. J., & Stoecker, C. F. (2011). *Where have all the young men gone? Using gender ratios to measure fetal death rates*.

Sandström, G., & Vikström, L. (2015). Sex preference for children in German villages during the fertility transition. *Population Studies*, *69*(1), 57–71. https://doi.org/10.1080/00324728.2014.994667

Schultz, T. P. (1994). Marital Status and Fertility in the United States: Welfare and Labor Market Effects. *The Journal of Human Resources*, *29*(2), 637. https://doi.org/10.2307/146114

Smith-Rosenberg, C. (2012). The Hysterical Woman: Sex Roles and Role Conflict in 19thcentury America. In *Women's Bodies*. DE GRUYTER SAUR. https://doi.org/10.1515/9783110976328.101/HTML

Sobek, M. (2016). *Data Prospects: IPUMS-International*. 157–174. https://doi.org/10.1007/978-94-017-7282-2_8

Stanfors, M., & Goldscheider, F. (2017). The forest and the trees: Industrialization, demographic change, and the ongoing gender revolution in Sweden and the United States, 1870-2010. *Demographic Research*, *36*(1), 173–226. https://doi.org/10.4054/DEMRES.2017.36.6

Strazdins, L., & Broom, D. H. (2004). Acts of Love (and Work). *Journal of Family Issues*, *25*(3), 356–378. https://doi.org/10.1177/0192513X03257413

Sun, A., Zhang, C., & Hu, X. (2019). Boys, Girls, and Grandparents: The Impact of the Sex of Preschool-Aged Children on Family Living Arrangements and Maternal Labor Supply. *Demography*, *56*(3), 813–833. https://doi.org/10.1007/s13524-019-00783-5

Tavassoli, N. (2021). The Transition of Son Preference: Evidence from Southeast Asian Countries. *Economics*, *9*(1), 43–67. https://doi.org/10.2478/EOIK-2021-0010

Tavassoli, N., Noghanibehambari, H., Noghani, F., & Toranji, M. (2020). Upswing in Industrial Activity and Infant Mortality during Late 19th Century US. *Journal of Environments*, *6*(1), 1–13. https://doi.org/10.20448/journal.505.2020.61.1.13

Torche, F., & Kleinhaus, K. (2012). Prenatal stress, gestational age and secondary sex ratio: The sex-specific effects of exposure to a natural disaster in early pregnancy. *Human Reproduction*, *27*(2), 558–567. https://doi.org/10.1093/humrep/der390

Wüst, M. (2015). Maternal Employment During Pregnancy and Birth Outcomes: Evidence From Danish Siblings. *Health Economics*, *24*(6), 711–725. https://doi.org/10.1002/hec.3056

Yamamura, E. (2013). Effects of sex preference and social pressure on fertility in changing Japanese families. *Journal of Socio-Economics*, *46*, 97–104. https://doi.org/10.1016/j.socec.2013.08.002

Yoo, S. H., Hayford, S. R., & Agadjanian, V. (2017). Old Habits Die Hard? Lingering Son Preference in an Era of Normalizing Sex Ratios at Birth in South Korea. *Population Research and Policy Review*, *36*(1), 25–54. https://doi.org/10.1007/s11113-016-9405-1

# Figures

## Appendix A

In this appendix, I report the sex ratios across subsamples and specifications to check whether they are in the empirical range reported by the literature and assumed to be the normal sex ratio of humans. The idea is to search for signs of sex-selective abortion and sex-selective infanticide that may drive the main results. The results are reported in Appendix Table A-1. The implications are discussed in the text in section 4.

**Appendix Table A-1 - Sex Ratio at First Birth for All Women and Married Women across Specifications**

| | Coefficient | [95% Confidence Interval] |
|---|---|---|
| ***Model 1. No Fixed Effects*** | | |
| *Full Sample* | | |
| Girl to Boy Ratio | 0.4898 | [0.4887,0.4909] |
| *Sample of Married Mothers* | | |
| Girl to Boy Ratio | 0.4889 | [0.4877,0.4902] |
| | | |
| ***Model 2. Region-by-Year FE*** | | |
| *Full Sample* | | |
| Girl to Boy Ratio | 0.4898 | [0.4887,0.4909] |
| *Sample of Married Mothers* | | |
| Girl to Boy Ratio | 0.4889 | [0.4877,0.4902] |

Notes. Confidence intervals are based on uncorrected standard errors.

## Appendix B

In this appendix, I show the heterogeneity of the results across birth cohorts. The results are reported in Appendix Table B-1 and Appendix Table B-2 Sample selection and interpretation of findings are reported in section 5.3.

The cohort analysis of Appendix Table B-1 is based on 9 countries with historically available data. These countries include Canada, United States, United Kingdom, Denmark, Iceland, Ireland, Norway, Sweden, and Germany. Appendix Table B-2 shows the heterogeneity across birth cohorts for countries excluding those studied in Appendix Table B-1 and for census years after 1960. For fertility outcomes, the effects appear to become larger and statistically significant for birth cohorts of 1940-onwards. Similar secular increases in sex ratio and son preference has been documented in the literature for the same birth cohorts (Choi and Hwang, 2020; Filmer et al., 2008).

**Appendix Table B-1 - Heterogeneity of the Effects of First Child Girl on Fertility and Family Structure across Birth Cohorts, Countries with Historical Data Available**

| | ***Disaggregation by Birth Cohort:*** | | | | |
|---|---|---|---|---|---|
| | 1650-1800 | 1800-1850 | 1850-1900 | 1900-1950 | 1950-2000 |
| | (1) | (2) | (3) | (4) | (5) |
| ***Panel A. Number of Children*** | | | | | |
| First Child Girl | 0.0078 | 0.0001 | 0.0412*** | 0.0004 | -0.0026 |
| | (0.0051) | (0.0171) | (0.0017) | (0.006) | (0.0025) |
| Observations | 178244 | 6724027 | 11706811 | 1278124 | 3499852 |
| R-squared | 0.0706 | 0.0938 | 0.0944 | 0.1419 | 0.0828 |
| Mean DV | 2.396 | 2.720 | 2.461 | 2.1216 | 1.852 |
| ***Panel E. Father Absent*** | | | | | |
| First Child Girl | 0.0015 | 0.0018 | 0.0018*** | 0.003 | 0.0067*** |
| | (0.0007) | (0.001) | (0.0004) | (0.0021) | (0.0015) |
| Observations | 178156 | 5277916 | 11698682 | 1278124 | 3499852 |
| R-squared | 0.2979 | 0.1627 | 0.1113 | 0.07 | 0.1322 |
| Mean DV | 0.0394 | 0.0577 | 0.0370 | 0.1055 | 0.2638 |
| Subnational-Region FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Subnational-Region-by-Year FE | Yes | Yes | Yes | Yes | Yes |

*Notes. Countries in the sample include Canada, Denmark, Iceland, Ireland, Norway, Sweden, United Kingdom, and United States. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*

**** p<.01, ** p<.05, * p<.1*

**Appendix Table B-2 - Heterogeneity of the Effects of First Child Girl on Fertility and Family Structure across Birth Cohorts, Countries with only Contemporaneous Data Available**

| | *Disaggregation by Birth Cohort:* | | | | |
|---|---|---|---|---|---|
| | 1900-1920 | 1920-1940 | 1940-1960 | 1960-1980 | 1980-2000 |
| | (1) | (2) | (3) | (4) | (5) |
| ***Panel A. Number of Children*** | | | | | |
| First Child Girl | -0.0275 | 0.0154 | 0.1039*** | 0.0673*** | 0.013 |
| | (0.0321) | (0.0167) | (0.0102) | (0.0123) | (0.0117) |
| Observations | 31996 | 2299374 | 11185801 | 28360619 | 11470160 |
| R-squared | 0.1557 | 0.1985 | 0.1676 | 0.2261 | 0.2448 |
| Mean DV | 2.316 | 2.742 | 2.343 | 2.0252 | 1.778 |
| ***Panel E. Father Absent*** | | | | | |
| First Child Girl | -0.0228 | 0.0029 | -0.0002 | 0.0042*** | 0.006** |
| | (0.0231) | (0.0027) | (0.0008) | (0.0015) | (0.0026) |
| Observations | 31973 | 2296261 | 11156312 | 28322069 | 11458645 |
| R-squared | 0.1916 | 0.1114 | 0.1742 | 0.152 | 0.2435 |
| Mean DV | 0.1362 | 0.0783 | 0.0872 | 0.1336 | 0.1799 |
| Subnational-Region FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Subnational-Region-by-Year FE | Yes | Yes | Yes | Yes | Yes |

*Notes. The sample excludes Canada, Denmark, Iceland, Ireland, Norway, Sweden, United Kingdom, and the United States. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*

**** p<.01, ** p<.05, * p<.1*

## Appendix C

Appendix Table C-1 reports the results for different clustering levels of standard errors

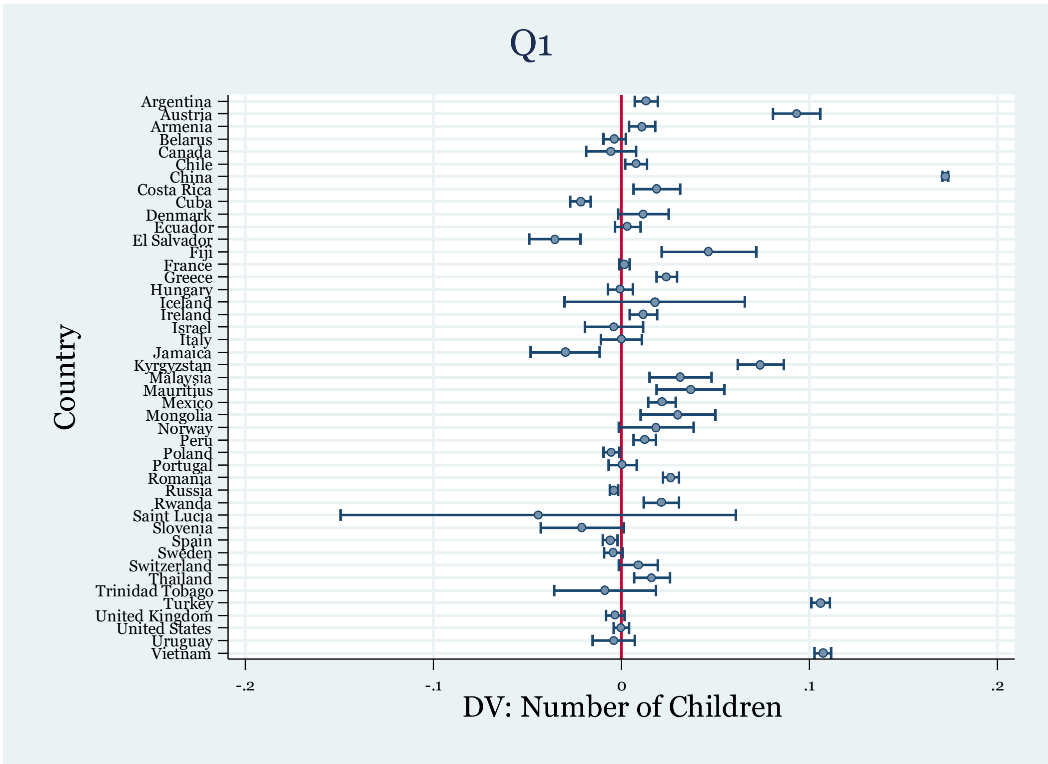
### Appendix Table C-1 - Robustness of the Main Results to Alternative Clustering Levels

| | *Correcting for Standard Errors by:* | | | | |
|---|---|---|---|---|---|
| | Cluster at Region (Results of Table 4) | Cluster at Year | Cluster at Country | Cluster at Region-by-Year | Hubert-White SE |
| | (1) | (2) | (3) | (4) | (5) |
| ***Panel A. Number of Children*** | | | | | |
| First Child Girl | 0.0515*** | 0.0515*** | 0.0515 | 0.0515*** | 0.0515*** |
| | (0.0097) | (0.0154) | (0.0357) | (0.0075) | (0.0054) |
| Observations | 76735229 | 76735229 | 76735229 | 76735229 | 76735244 |
| R-squared | 0.2218 | 0.2218 | 0.2218 | 0.2218 | 0.2218 |
| ***Panel E. Father Absent*** | | | | | |
| First Child Girl | 0.0041*** | 0.0041*** | 0.0041*** | 0.0041*** | 0.0041*** |
| | (0.0013) | (0.0011) | (0.0014) | (0.0011) | (0.0011) |
| Observations | 75198209 | 75198209 | 75198209 | 75198209 | 75198224 |
| R-squared | 0.1698 | 0.1698 | 0.1698 | 0.1698 | 0.1698 |
| Subnational-Region FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Subnational-Region-by-Year FE | Yes | Yes | Yes | Yes | Yes |

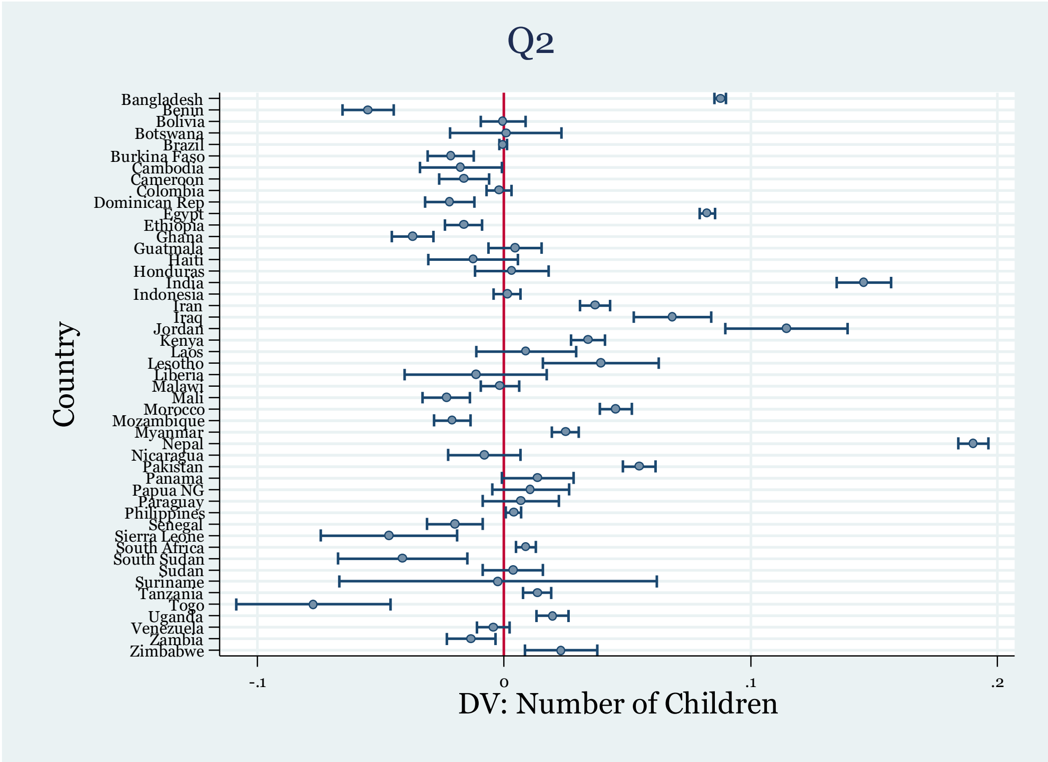
*Notes. All standard errors are in parentheses and clustered at the region-level. All regressions are weighted using weighting scheme explained in section 3. Parental controls include dummies for education, ownership of dwelling, school attendance, employment status, and labor force participation. Also, missing indicators are also included in the regressions to control for any missing values for parental controls.*
**** p<.01, ** p<.05, * p<.1*

## Appendix D

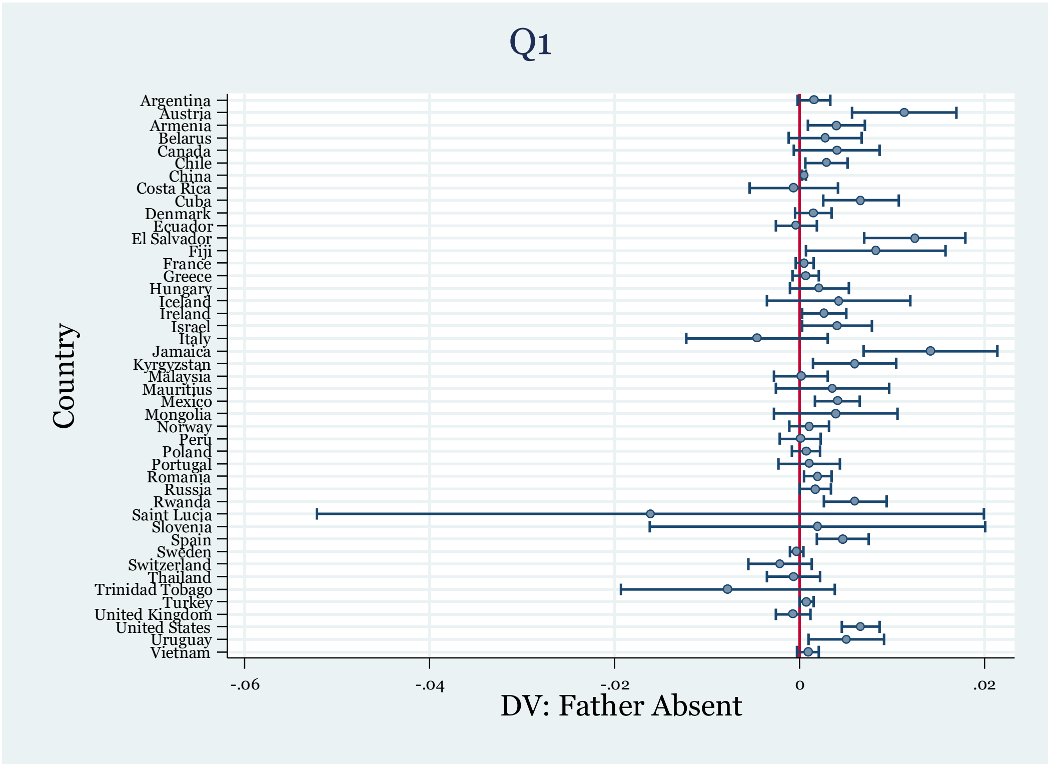
In this appendix, I replicate the main results for each individual country. Since several countries have only one sub-national region, clustering standard errors cause a problem to obtain the confidence intervals. Therefore, I use robust standard errors. I apply a full specification model and plot the marginal in Appendix Figure D-1 through Appendix Figure D-4. These results are in line with several of previous studies. For instance, Hank and Kohler (2000a) find a preference towards girls for Portugal. Andersson et al. (2006) and Andersson et al. (2007) show that while parents in Nordic countries do not reveal a son preference, Swedish parents reveal a preference over girls rather than boys. These are in line with the overall pattern of effects reported in the figures of this appendix. While the evidence for Asian countries and specifically Southeast Asian countries are established and in line with the evidence reported here, the effects are also consistent with previous studies on single countries such as South Africa (Gangadharan, 2003), Nigeria (Milazzo, 2014), Morocco (Makhlouf Obermeyer, 1996), and Mexico (Marcén et al., 2018).



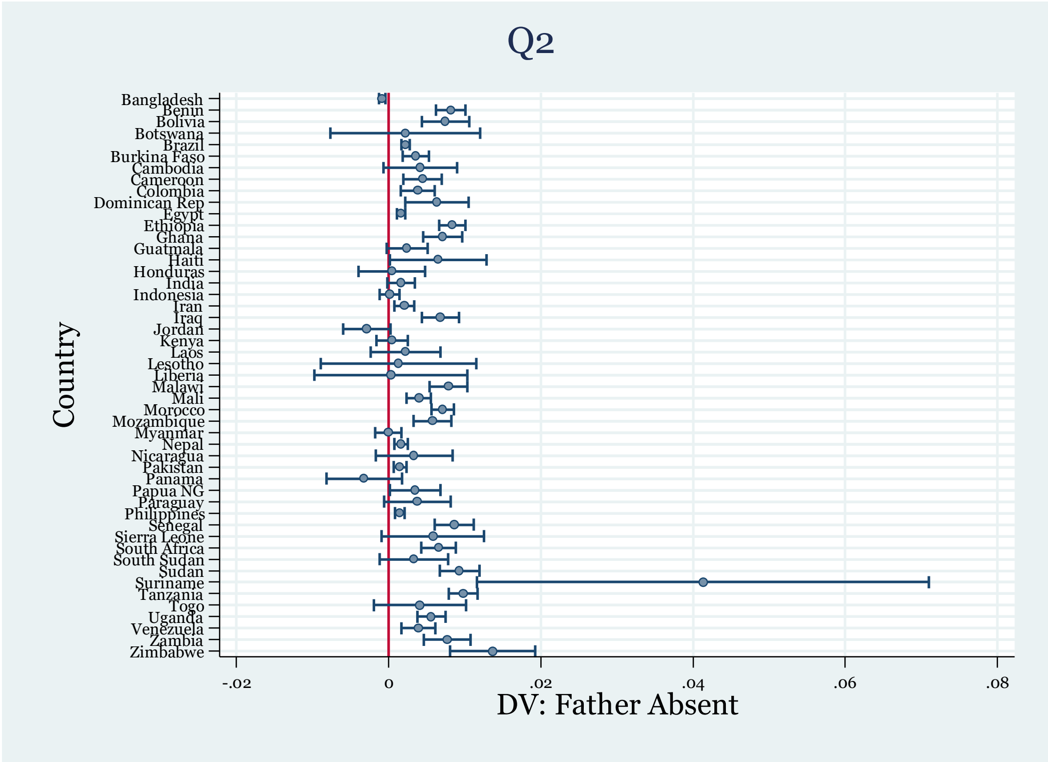

**Appendix Figure D-1 - Marginal Effects of First Child Girl on Number of Children by Country for Countries below Median GII and applying a Full Specification**




**Appendix Figure D-2 - Marginal Effects of First Child Girl on Number of Children by Country for Countries above Median GII and applying a Full Specification**




**Appendix Figure D-3 - Marginal Effects of First Child Girl on Farther Absence by Country for Countries below Median GII and applying a Full Specification**




**Appendix Figure D-4 - Marginal Effects of First Child Girl on Farther Absence by Country for Countries above Median GII and applying a Full Specification**