Multiple Experiments for the Causal Link between the Quantity and Quality of Children

Citation: Angrist, Joshua, Victor Lavy, and Analia Schlosser. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." Journal of Labor Economics 28.4 (2010): 773-824.

As Published: http://dx.doi.org/10.1086/653830

Publisher: University of Chicago Press

Persistent URL: http://hdl.handle.net/1721.1/61947

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

QQdraft_April25_2010.doc; 4/28/2010

Multiple Experiments for the Causal Link between the Quantity and Quality of Children

By,

Joshua Angrist

MIT and NBER

Victor Lavy

Hebrew University, Royal Holloway University of London, and NBER

Analía Schlosser

Tel Aviv University

March 2010

Special thanks go to the staff of the Central Bureau of Statistics in Jerusalem, whose assistance made this project possible. We thank David Autor, Oded Galor, Omer Moav, Saul Lach, Kevin Lang, Manuel Arellano, Yaacov Ritov, Yona Rubinstein, Avi Simhon, David Weil and seminar participants at the 2005 NBER Summer Institute, Boston University, Brown, Harvard, SUNY-Albany, UCL, Pompeu-Fabra, Stanford GSB, The China Center for Economic Research, The December 2005 Evaluation Conference in Paris, the 2005 SOLE meetings, and the University of Zurich for helpful discussions and comments on earlier versions of this paper. We also thank Alan Manning for helpful comments.

Multiple Experiments for the Causal Link between the Quantity and Quality of Children

This paper presents evidence on the child-quantity/child-quality trade-off using quasi-experimental variation due to twin births and preferences for a mixed sibling-sex composition, as well as ethnic differences in the effects of these variables. Our sample includes groups with very high fertility. An innovation in our econometric approach is the juxtaposition of results from multiple instrumental variables (IV) strategies, capturing the effects of fertility over different ranges for different sorts of people. To increase precision, we develop an estimator that combines different instrument sets across partially-overlapping parity-specific sub-samples. Our results are remarkably consistent in showing no evidence of a quantity-quality trade-off.

Josh Angrist
Department of Economics
MIT
angrist@mit.edu

Victor Lavy
Department of Economics
Hebrew University and Royal Holloway University of London
msvictor@huji.ac.il

Analía Schlosser
The Eitan Berglas School of Economics
Tel Aviv University
analias@post.tau.ac.il

FAMILY PLANNING: THE WAY TO PROSPERITY.
(A SLOGAN FOUND ON THE BACK OF INDONESIA'S FIVE-RUPIAH COIN)

I. Introduction

The question of how family size affects economic circumstances is one of the most enduring in social science. Beginning with Becker and Lewis (1973) and Becker and Tomes (1976), economists have developed a rich theoretical framework that sees both the number of children and parental investment per child as household choice variables that respond to economic forces. An important implication of this framework is that exogenous reductions in family size should increase parental investment in children, thereby improving human capital and welfare. By the same token, events that lead to otherwise unplanned increases in family size should reduce parental investment and therefore reduce infra-marginal "child quality."

On the policy side, the view that smaller families and slower population growth are essential for economic development motivates many international agencies and some governments to promote, or even to require, smaller families. In addition to China's One Child Policy, examples of government-sponsored family planning efforts include a forced-sterilization program in India and the aggressive public promotion of family planning in Mexico and Indonesia.[1] Bongaarts (1994) notes that by 1990, 85 percent of people in the developing world lived in countries where the government considers fertility to be too high. The Becker and Lewis (1973) model, as well as recent economic analyses of the role of the demographic transition, provide additional theoretical support for the view that large families keep living standards low (e.g., Galor and Weil, 2000; Hazan and Berdugo, 2002, and Moav, 2005).

Most of the scholarly evidence pointing to an empirical quantity-quality trade-off comes from the widely observed negative association between family size on one hand and schooling or academic achievement on the other. For example, Leibowitz (1974) and Hanushek (1992) find that children's educational attainment and achievement growth are negatively correlated with family size. Many other

[1]These episodes are recounted in Weil (2005; Chapter 4), which also mentions the anti-natalist slogan on the Indonesian Rupiah.

micro-econometric and demographic studies show similar relations.[2] The principal problem with research of this type is the likelihood of omitted variables bias in estimates of the effects of childbearing. This is highlighted by Angrist and Evans (1998), who used instrumental variables (IV) derived from multiple births and same-sex sibling pairs to estimate the causal effect of family size on mothers' labor supply. IV estimates, while still negative, are considerably smaller than the corresponding OLS estimates.

This paper provides new evidence on the quantity-quality trade-off using exogenous variation in family size in low- and high-fertility sub-samples. We begin by looking at the effect of third and higher-parity births on first- and second-born children's completed schooling, labor market status, adult earnings, and marital status and fertility. These are important long-run "quality" indicators that are likely to be affected by the home environment. Effects on marriage and fertility also play a role in some theories of the demographic transition (Lutz and Skirbekk, 2005).

Two of the instruments used here are dummies for multiple second births and a dummy for same-sex sibling pairs in families with two or more children. We also extend the sex-composition and twins identification strategies in a number of ways. First, we introduce a new source of exogenous variation in family size based on sharp differences in the effects of multiple births and sex-composition across ethnic groups in the Israeli population. Second, as an alternative to instruments based on sex-mix, we exploit preferences for boys at higher order births in some ethnic groups.[3] Third, we combine twins and sex-composition instruments at different parities to produce more precise IV estimates and increase the range of variation covered by our experiments. This parity-pooled analysis includes third and fourth-born children. Finally, we present evidence for the exclusion restrictions that justify IV by estimating reduced form effects in samples with no first stage.

[2]See, e.g., the review by Schultz (2005). Johnson (1999) notes that the relation between family size and economic well being or growth is less clear cut at the time series or cross-country level.
[3]Traditional Jewish preferences over sibling sex-composition can be traced back to the Mishna (Oral law): A man shall not stop having children until he has two. Beit Shamai (a relatively strict rabbinic tradition) says two sons, while Beit Hillel (a more forgiving rabbinic tradition) says a boy and a girl. As it is written in Genesis, 'male and female he created them.' (Mishna Nashim - Yebhamoth 6:7).

The fact that our analysis combines evidence from multiple sources of variation is important for a number of reasons. First, both twins and sex-composition instruments are potentially subject to omitted variables biases. For example twin rates vary with maternal characteristics like age at birth and race, and twin births affect child spacing and child health in a manner that seems likely to accentuate any negative effects of childbearing. Instrumental variables derived from sibling sex composition are not subject to these considerations, though sex-composition may affect outcomes due, say, to economies of scale through room- or clothes-sharing (as suggested by Rosenzweig and Wolpin, 2000). A comparison of twins and sex-composition estimates therefore provides a specification check since the omitted variables bias associated with each type of instrument should act differently. The use of instruments based on preferences for male children *per se* also provides a simple check on IV estimates derived from sex-mix.

A related consideration arises from that fact that the estimates generated by any particular IV strategy capture effects on individuals affected by that instrument (Imbens and Angrist, 1994). Moreover, in models with variable treatment intensity, IV results are specific to the range of variation induced by the instrument (Angrist and Imbens, 1995). As noted by Moffit (2005), these limitations lead to concerns about the external validity of IV estimates. Our analysis addresses these concerns by juxtaposing results from different quasi-experimental research designs. On one hand, as we show below, twins instruments identify the effect of treatment on the non-treated since compliance is perfect when a multiple birth occurs. On the other hand, the average causal response due to a twin birth reflects the impact of increasing family size only (or mostly) at the parity of occurrence. Sex-composition instruments, in contrast, shift the fertility distribution at parities as high as nine. Moreover, the ethnic composition of same-sex compliers (in the sense of Angrist, Imbens, and Rubin, 1996) tends to vary in a manner opposite to that for twins. We therefore argue that the fact that IV estimates affecting different people and inducing differing ranges of variation generates similar results, as ours do, provides considerable evidence for the external validity of our estimates.

A limitation of our quasi-experimental identification strategies is that they fail to capture effects on the marginal child. For example, we can study the effect on an older child of having an extra young sibling, but not the effect on a younger child of being born into a larger family (whether the family is large due to twinning, sex preferences, or for any other reason). At the same time, the similarity of our results across alternative identification strategies, subpopulations, and fertility increments, offers no evidence of substantially heterogeneous effects by parity or birth order.

Our paper is related to a burgeoning empirical literature that uses multiple births to estimate the causal effects of family size. Rosenzweig and Wolpin (1980) appear to have been the first to use twins to estimate a child-quantity/child-quality trade-off. Other estimates using multiple births include Duflo (1998), who looks at effects on child mortality in Indonesia, Caceres (2006), who looks at effects on private schooling and grade retention in US Census data, and Black, Devereux, and Salvanes (2005), who use twins to estimate family-size effects on education and earnings in Norway. As in our paper, Black, *et al* (2005) look at human capital variables with a large administrative sample. In contrast with the original Rosenzweig and Wolpin study, this literature has uncovered surprisingly little evidence for an adverse effect of family size on human capital. On the other hand, a recent paper by Rosenzweig and Zhang (2009), using the effect of twins instruments on twins themselves to bound quantity-quality effects, suggests there is a trade-off. Because twins probably differ from non-twins for reasons both observed and unobserved, we prefer empirical strategies that look at the effects of twins on older siblings.[4] Nevertheless, we also briefly discuss results from the Rosenzweig and Zhang (2009) approach.

To the best of our knowledge, none of this previous work has attempted to combine or reconcile evidence from multiple natural experiments. Our paper also differs from Caceres (2004) and Black, *et al* (2005) in that we study a higher-fertility population with demographic and social characteristics much closer to developing country populations. Of particular interest is the Asia-Africa (AA) subsample, that

[4] Similarly, when using sex-mix as an instrument we look only at older siblings, because the outcomes of the last child born come from an endogenously selected sample if fertility is endogenous.

is, Sephardic Jews of North African and Middle Eastern origin. Sephardic Jews are poor relative to the Israeli average and typically have larger families.[5]

On the methodological side, our paper has features in common with Oreopoulos (2006), which compares IV estimates of the returns to schooling using changes in compulsory schooling laws in different countries. Oreopoulos argues that this comparison can be used to gauge the importance of treatment-effect heterogeneity when the size of the compulsory-schooling first stage varies. A final contribution stems from the relative precision of our estimates. Having established that different instruments and samples generate broadly similar effects, we develop a simple two-stage least squares (2SLS) procedure that combines parity-specific instrumental variables estimates into a single estimate that is more precise than the twins estimates reported by Black, *et al* (2005).

The next section describes the data and the construction of the analysis samples, as well as data quality issues such as the relationship between instruments and match rates. Section III discusses the first-stage estimates and their implications for treatment effect heterogeneity and nonlinearity, while Section IV presents the main OLS and 2SLS results. On balance, the result reported here offer little evidence for an effect of family size on schooling, work, or earnings, though we do find some effects on girls marital status, age at marriage, and fertility. Section V discusses possible explanations for these findings while Section VI concludes and suggests directions for further work.

II. Data and Samples

The main sources of data used here are the 20% public-use micro-data samples from the 1995 and 1983 Israeli censuses, linked with information on parents and siblings from the population registry. The Israeli census micro files are 1-in-5 random samples that include information collected on a fairly

[5] Israel in 1975, when the subjects we study were growing up, was an upper middle income country, with GDP per-capita about like Greece and Argentina; see Heston, Summers, and Aten (2002).

detailed long-form questionnaire similar to the one used to create the PUMS files for US censuses.[6] The set of Jewish long-form respondents aged 18-60 provides our initial study sample. In the discussion that follows, we refer to these individuals as "subjects," to distinguish them from their parents and siblings, for whom we also collected data. The link from census to registry is necessary for our purposes because in a sample of adult respondents, most of whom no longer live with their parents and siblings, the census provides no information about sibship size, multiple births, or sibling sex composition.[7]

Match Rates and Sample Selection

The vast majority of our census subjects appear in the population registry. This can be seen in Table 1, the first two rows of which report starting sample sizes and subject-to-registry match rates, grouped according to whether subjects' parents were Israeli born, birth cohort, and whether subjects were Israeli-born (there are two panels in the table, one for each census). Subject-to-registry match rates range from 95-97 percent regardless of cohort and nativity. The first coverage shortfall from our point of view is the failure to obtain an administrative record for subjects' mothers. This failure arises for a number of reasons. First, subject's mothers may have been alive but not at home in 1948 when the registry was created, or a mother may have been deceased. Second, children are more likely to be linkable to parents and siblings when a subject's mother gave birth to all of her children in Israel.

The second row of each panel in Table 1 describes the impact of these record-keeping constraints on our census-to-registry match rates. The mothers of subjects with Israeli-born fathers were found 90 percent of the time for cohorts born after 1955. On the other hand, for those born before 1955, only 17

[6]Documentation can be found at the Israel Social Sciences Data Center web site: http://isdc.huji.ac.il/mainpage_e.html (data sets 115 [1995 demographic file] and 301 [1983 files]). The Census includes residents of dwellings inside the State of Israel and Jewish settlements in the occupied territories. This includes residents abroad for less than one year, new immigrants, and non-citizen tourists and temporary residents living at the indicated address for more than a year.

[7]About 80% of the Israeli population is Jewish. The study sample is limited to Jews because census-to-population-registry match rates are considerably lower for other groups. Additional information related to data set construction appears in the data appendix.

percent of mothers were found. Likewise, for those with foreign-born fathers, there is a similar age gradient in mothers' match rates. Even in this group, however, 87 percent of mothers were found for younger Israeli-born subjects in the 1995 census. The 1955 birth cohort marks a useful division for our purposes because mothers of subjects born after 1954 gave birth to most of their children in post-1948 Israel (the mothers in this group were mostly born after 1930, and, assuming childbirth starts at 18, this dates their first births at 1948 or later).

Given the match rates in Table 1, our analysis sample is weighted towards post-1955 cohorts (i.e., 40 or younger in 1995). This accounts for about two-thirds of the 1995 population aged 18-60. Among the children of immigrant fathers, we're also much more likely to find mothers of the Israeli-born. The coverage rates for post-1955 Israeli-born cohorts seem high enough that we are likely to have information on mothers for a representative sample of younger cohorts regardless of fathers' nativity. We also used information on mothers in the matched sample to discard any remaining mothers who were born before 1930 (as the match rates for this group appeared to be very low anyway). Subjects with mothers whose first birth was before age 15 or after age 45 were also dropped. These restrictions eliminate almost all subjects born before 1955, primarily because most of those born earlier have mothers born before the 1930 maternal age cutoff. We also restricted the sample of subjects with foreign-born mothers to those whose mothers arrived 1948 or later and before the age of 45 (in this case so that an immigrant mother with children is likely to have come with all her children, who would then have been included in the registry, either in the first census, or at the time IDs were issued to the family).

The final sample restriction retains only first and second-born subjects since these are the people exposed to the natural experiments exploited by the twins and sex-composition research designs. Note that the restriction to first and second born subjects naturally eliminates a higher percentage of younger rather than older cohorts. This restriction also has a bigger effect on the Israeli-born children of foreign-born fathers than on other nativity groups, probably because these children were disproportionately likely to have been born to immigrant fathers who arrived with a large wave of immigrants from Asia and

Africa in the 1950s. Immigrants from this group typically formed large families after arrival and will therefore have contributed more higher-parity births to the sample.

A key issue for the internal validity of our study is whether the instruments affect the likelihood of a successful match between Census and registry files. Specifically, the censuses might constitute a nonrandom sample that is selected in some manner related to the key variables in our study. We explored this by looking at match rates from registry to Census as a function of family size and our instruments. The maximum theoretical match rate here is 20 percent. The reverse match finds about 15 percent of registrants alive in 1983 and about 17 percent of registrants alive in 1995. Discrepancies between theoretical and actual match rates can arise due differences in the definition of the base population in the Registry and the Census, missing or invalid IDs on Census records, and Census non-response rates.[8] The means and first stage regressions coming out of the reverse match and of the sample of all Israeli citizens alive on census date and registered at the population registry, are almost identical to those generated by our working extract.[9] Most important, there is no significant relation between the twins and same-sex instruments and the probability of appearing in either the 1983 or 1995 census data. This is documented in Table A1 in the appendix, which is also discussed in the data appendix. There is a small statistically significant decrease in the probability someone from a larger family appears in the 1983 Census, which can be explained by the higher Census non-response rates of ultra-orthodox households (Central Bureau of Statistics, 1985). Nevertheless, since the bulk of our sample is from 1995, the results are similar in the two data sets, and there is no differential selection as a function of the instruments, it seems unlikely that differential match rates affect the results.

[8] For methodological aspects of the 1983 and 1995 Census see Central Bureau of Statistics (1985 and 2001).

[9] For technical reasons related to CBS data handling protocols, the population of registrants used for the reverse match differs slightly from the registry population used to construct our main extract. In practice, these differences have no bearing on our analysis.

Description of Analysis Samples

We work with two main analysis samples, both described in Table 2. One consists of first-born subjects in families with two or more births (the 2+ sample, N=89,445). The second sample consists of first- and second-born subjects in families with three or more births (the 3+ sample, N=65,673 first-born and 52,964 second-born). These samples are defined conditional on the number of births instead of the number of children so that multiple-birth families can be included in the analysis samples without affecting the sample selection criteria. Twin subjects were dropped from both samples, however.[10]

Roughly three-quarters of the observations in each sample were drawn from the 1995 Census. On average, subjects were born in the mid-sixties and their mothers were in their early twenties at first birth. Because out-of-wedlock childbearing is rare in Israel, especially among the cohorts studied here, virtually all subjects in both samples were born to married mothers. Naturally, however, some marriages have since broken up and some wives have been widowed. This is reflected in the 2003 marital status variables available in the registry.[11]

The Jewish Israeli population is often grouped by ethnicity, with Jews of African and Asian origin (AA; e.g., Moroccans), distinguished from Jews of European and North American (EA) origin. The 2+ sample is about 40 percent AA (defined using father's place of birth), while the 3+ sample is over half AA. A preference for larger families in the AA population is also reflected in the statistics on numbers of children. Average family size ranges from 3.6 in the 2+ sample to 4.2 in the 3+ sample (4.3 for second-borns). In the AA subsample, however, the corresponding family sizes are about 4.3 and 4.7.

Table 2 also reports statistics on the variables used to construct instrumental variables. The twin rate was 9/10 of one percent at second birth in the 2+ sample and 1 percent at third birth in the 3+ sample,

---

[10]A 3+ sample defined as including first-born children from families with three or more children instead of three or more births would include all families with multiple second births. Likewise, sibling-sex composition can be defined across births without the need to determine which, say, of two twins, constitutes the second child.

[11]The 2+ sample of first-borns naturally includes the 3+ sample of first-borns. In the 3+ sample, about 10 percent of the first- and second-borns have the same mother (both must appear in the 20% census sample and be in the relevant age range). We therefore cluster analyses that pool parities by mothers' ID.

with similar rates in the AA and full samples.[12] As expected, about 51 percent of births are male, regardless of birth order. Consequently, about half of the 2+ sample was born into a same-sex sibling pair and about one-quarter of the 3+ sample was part of a same-sex threesome.

The outcome variables described in Table 2 measure subjects' educational attainment, labor market status and earnings, marital status and fertility. Most Israelis are high school graduates, while 20 percent are college graduates. In the AA subsample, however, the proportion of college graduates is much lower. Most of our subjects were working at the time they were interviewed and earned about 3000 shekels (about 1000 dollars) per month on average (including zeros). About 45 percent of subjects were married, though marriage rates are higher in the AA sub-sample.

III. First-stage Estimates, Interpretation of IV Estimates, and Instrument Validity

Different instruments generate different average causal effects. Of particular importance in this context are: (a) the links between first-stage effects and the subpopulations affected by each underlying natural experiment, and (b) the relation between first-stage effects and the range of variation induced by each instrument. These points are detailed below.

A. Twins First-Stages

A multiple second birth increases the average number of siblings in the 2+ sample by about half a child, a finding reported in column 1 of Table 3, which gives first-stage estimates for the twins experiment. In particular, column 1 reports estimates of the coefficient α in the equation

$$c_i = X_i'\beta + \alpha t_{2i} + \eta_i \tag{1a}$$

[12]Note that the second-birth twin rate in the 3+ sample is not comparable to the second birth twin rate in the 2+ sample or the third-birth twin rate in the 3+ sample because the 3+ sample consists of those who had three or more births. Families with a second-born twin need not have a third birth to have three or more children. Families with a second-born twin that have a third birth have at least four children, and hence are relatively rare in the 3+ sample.

where c_i is subject i's sibship size (including the subject), X_i is a vector of controls that includes a full set of dummies for subjects' and subject's mothers' ages, Mothers' age at first birth, mothers' age at immigration (where relevant), fathers' and mothers' place of birth, census year, and a dummy for missing month of birth. The variable t_{2i} (which we call twins-2) indicates multiple second births in the 2+ sample.

The Israeli twins-2 first stage is smaller than the twins-2 first stage of about .6 in the Angrist and Evans (1998) sample, reflecting the fact that Israelis typically have larger families than Americans. Multiple births result in a smaller increase in family size when families would have been large even in the absence of a multiple birth. Within Israel, however, there are marked differences in the twins first-stage by ethnicity. This can be seen in column 2 of Table 3, which reports the twins-2 main effect and an interaction term between twins-2 and a dummy for Asia-Africa ethnicity (a_i) in the equation

$$c_i = X_i'\beta + \alpha_0 t_{2i} + \alpha_1 a_i t_{2i} + \eta_i. \tag{1b}$$

The twins-2 main effect, α_0, captures the effect of a multiple birth in the non-AA population, while the interaction term, α_1, measures the AA/non-AA difference.[13] The estimates in column 2 show that non-AA family size goes up by about .63 in response to a multiple birth (similar to the AE-98 first stage), while AA family size increases by only .63-.48=.15. Both α_0 and α_1 are very precisely estimated.

The remaining columns of Table 3 report the first-stage effect of a multiple third birth in the 3+ sample. Twins-3 effects were estimated in the 3+ sample by replacing t_{2i} with t_{3i}, a dummy for multiple third births, in equations (1) and (2). These results are reported in columns 3-4 for first-borns and columns 5-6 for the pooled sample of first- and second-borns. The first stage effect of a multiple birth is bigger in the 3+ sample than in the 2+ sample because the desire to have additional children diminishes as family size increases. For the same reason, the effect of t_{3i} differs less by ethnicity in the 3+ sample than in the 2+ sample, though, as the estimates in column 6 show, there is still a significant difference by ethnicity when first and second born subjects are pooled.

[13]The a_i main effect is included in the vector of covariates, X_i. Note that the covariate effects, all labeled 'β', differ as the first-stage specification and sample change.

Heterogeneity and Non-linearity in the Response to a Multiple Birth

The difference in first stage effects across ethnic groups has a useful interpretation in the average causal response (ACR) framework laid out by Angrist and Imbens (1995). To see this, define potential endogenous variables C_{0i} and C_{1i} to be the number of children a woman would have if a generic binary instrument, Z_i, is equal to zero or one. Because we observe C_{0i} for those with Z_i equal to zero and C_{1i} for those with Z_i equal to one, the realized number of children is

$$c_i = C_{0i} + (C_{1i} - C_{0i}) Z_i.$$

For a model without covariates, the IV estimand using this instrument is the Wald estimator (see, e.g., Angrist, 1991):

$$\beta_w = \frac{E[y_i | Z_i = 1] - E[y_i | Z_i = 0]}{E[c_i | Z_i = 1] - E[c_i | Z_i = 0]}$$

where y_i is the outcome variable. The observed y_i is related to potential outcomes, $Y_i(j)$, where j indexes possible values of $c_i = 0, 1, 2, \ldots, J$; as follows:

$$y_i = Y_i(0) + \textstyle\sum_j [Y_i(j) - Y_i(j-1)] 1[c_i \geq j], \tag{2}$$

where the summation is from $j=1, \ldots, J$.

A linear constant-effects model imposes the restriction, $Y_i(j) - Y_i(j-1) = \rho$, for all i and j, in which case the Wald estimator equals this parameter. More generally, Angrist and Imbens (1995) show that

$$\beta_w = \textstyle\sum_j E[Y_i(j) - Y_i(j-1) | C_{1i} \geq j > C_{0i}] \omega(j); \tag{3}$$

where the weighting function, $\omega(j)$, is

$$\omega(j) = P[C_{1i} \geq j > C_{0i}] / \{\textstyle\sum_j P[C_{1i} \geq j > C_{0i}]\}$$

Thus, the Wald estimator is a weighted average causal response (ACR) for people from families induced by an instrument to go from having fewer than j to at least j children, weighted over j by the probability of crossing this threshold.[14]

It is straightforward to show that the denominator normalizing the weights, $\omega(j)$, is the Wald first-stage. In other words,

$$E[c_i| Z_i=1] - E[c_i| Z_i=0] = E[C_{1i}-C_{0i}] = \sum_j P[C_{1i}\geq j > C_{0i}].$$

This relation is important because we can think of individuals with $C_{1i}\geq j>C_{0i}$ for any j in the support of c_i as *compliers* in the sense of Angrist, Imbens, and Rubin (1996). In this context, the subpopulation of compliers consists of individuals who switch from having fewer than j to at least j children because of the instrument. Differences in the size of the first stage across demographic or ethnic groups measure differences in the probability of compliance between these groups.

As a practical matter, we can use the ratio of first stages for the AA and overall sample to measure the likelihood that twins-2 compliers are of AA ethnicity. To see this, note that

$$E[C_{1i}-C_{0i}| a_i = 1]/ E[C_{1i}-C_{0i}] = \sum_j (P[a_i = 1| C_{1i}\geq j>C_{0i}]/ P[a_i = 1])\omega_{j,}$$

where the weights, $\omega_j = P[C_{1i}\geq j>C_{0i}]/ \sum_j P[C_{1i}\geq j>C_{0i}]$, sum to one. Thus, the ratio of the first-stage for the AA subsample to the overall first-stage summarizes the extent to which compliers are AA, relative to the population proportion AA. The fact that AA family size increase by only .15 in response to a second twin birth while the overall first stage is .44 therefore means that the population of twins compliers is less than half as likely to be AA as the overall population. In contrast, sex-composition compliers are disproportionately likely to be AA, as we show below.

[14]The assumptions that lay behind the ACR theorem are: (a) Potential outcomes and treatment assignments are independent of the instrument; (b) The instrument moves fertility in one direction only (monotonicity), i.e., $C_{1i}\geq C_{0i},$] With covariates, the interpretation of the ACR is more elaborate, but the basic idea is preserved. Because some parents may prefer a mixed sibship while others may prefer same-sex sibships, monotonicity need not hold for sex composition instruments. As a partial check on monotonicity, we estimated the same-sex first stage separately by intervals of individual year of birth, maternal age at first birth, and ethnicity. Only 3 out of 36 cells generated negative estimates and all 16 significant estimates were positive.

A second important feature of the twins identification strategy is the fact that twins estimates capture the causal effect of childbearing in a narrow range. Figure 1, which plots first-stage estimates of the effect of twins-2 and twins-3 on $\{d_{ji} \equiv 1(c_i \geq j);\ j=1,\ldots,11\}$, along with the associated confidence bands, documents this. The normalized CDF differences plotted in Figure 1 are the ω_j in the ACR decomposition of β_w in equation (3). The figure therefore implies that twins instruments capture an average causal effect over a range of fertility variation that is close to the parity of the multiple birth. For example, a multiple third birth increase the likelihood of having a fourth child by about .35 in the Asia-Africa 3+ sub-sample, with a much smaller effect on the likelihood of having a fifth child and no significant effect at higher parities (see the lower left panel of figure 1).[15]

The last distinctive econometric feature of the twins estimates is that they generate the average causal *effect of treatment on the non-treated*, where treatment is defined as a dummy for having another child. Specifically, the subpopulation of compliers affected by the twins-2 instrument is the entire population with two children. This is a consequence of the causal treatment-effects framework outlined in Angrist, Imbens, and Rubin (1996), which divides the population into three types of responders affected by a Bernoulli instrument: always-takers who always get treated, never-takers who never get treated, and compliers who get treated when the instrument is switched on but not otherwise. The treated consist of always-takers plus compliers with the instrument switched on while the non-treated consist of never-takers plus compliers with the instrument switched off. But with twins there are no never-takers, so the non-treated consist only of compliers with the twins instrument switched off. Because twinning is as good as randomly assigned, causal effects for the latter population are the same as causal effects on all

[15]The twins instrument engenders small shifts in fertility at parities beyond the twinning parity because a multiple birth leads to tighter spacing, thereby lengthening the biological window for continued childbirth. This is most likely to relevant for the ultra-orthodox minority who have very high fertility.

compliers. Form this we conclude that the parameter identified by twins instruments is the average effect on the non-treated.[16]

Sex-composition instruments identify average causal effects that differ in two ways from the effects captured by twins. On one hand, the compliers population is less complete; not all the non-treated are affected by sex-composition. On the other hand, as we show below, the range of fertility variation induced is often quite a bit wider. In particular, sex-composition instruments (including a dummy for 3rd-born male children) shift the fertility distribution over a wider range than does a multiple birth, especially in the event of an all-female sibship.

B. Sibling-Sex Composition First-stages

Sex-composition first stages in the 2+ sample were estimated using the following two models:

$$c_i = X_i'\beta + \gamma_1 b_{1i} + \gamma_2 b_{2i} + \pi_s s_{12i} + \eta_i \tag{4a}$$

$$c_i = X_i'\beta + \gamma_1 b_{1i} + \pi_b b_{12i} + \pi_g g_{12i} + \eta_i \tag{4b}$$

where b_{1i} (boy-first) and b_{2i} (boy-second) are dummies for boys born at first and second birth, the variable

$$s_{12i} = b_{1i}b_{2i} + (1-b_{1i})(1-b_{2i}),$$

is a dummy for same-sex sibling pairs, and

$$b_{12i} = b_{1i}b_{2i} \text{ and } g_{12i} = (1-b_{1i})(1-b_{2i})$$

indicate two boys and two girls. Note also that b_{1i} indicates the subject's sex in the 2+ sample, and that $s_{12i} = b_{12i}+g_{12i}$. The first model controls for boy-first and boy-second main effects, while the excluded instrument is a same-sex effect common to boy and girl pairs. The second model allows the effect of two

[16]Here is a more formal argument: Note first that $P[C_{1i} \geq 3 > C_{0i}] = P[C_{0i} = 2]$, since $C_{1i} \geq 3$ and $C_{0i} \geq 2$ for everybody in the 2+ sample. Moreover, $P[C_{1i} \geq j > C_{0i}]$, is close to zero for $j > 3$ since a multiple second birth has little effect on childbearing at higher parities. Therefore, $\beta_w = E[Y_i(3)-Y_i(2)|\ C_{0i}=2]$. Finally, because Z_i is independent of potential outcomes and potential treatment assignments, $\beta_w = E[Y_i(3)-Y_i(2)|\ C_{0i}=2, Z_i=0]$. But this is the same as $E[Y_i(3)-Y_i(2)|\ c_i=2]$, because all those with two children have singleton births and $C_{0i}=2$. A similar line of reasoning leads to the conclusion that the twins-3 estimator in the 3+ sample identifies $E[Y_i(4)-Y_i(3)|\ c_i=3]$.

boys and two girls to differ, though one of the boy main effects must be dropped since $\{b_{1i}, b_{2i}, b_{12i}, g_{12i}\}$ are linearly dependent.[17] We also report results from models with AA interaction terms, as in Table 3.

The first-stage effect of s_{12i} in the 2+ sample, reported in column 1 of Table 4, is .073 children. The AA interaction term in this case is essentially zero, so that in contrast with the twins first-stage, the overall sex-composition effect in the 2+ sample is the same for the AA and non-AA populations.

In models with common effects across ethnic groups, two girls increases family size by .11 (s.e.=.015) while the effect of two boys is .039 (s.e.=.015). This can be seen in columns 3 and 4 of Table 4, which report estimates of π_b and π_g in equation (4b). Models allowing different coefficients by ethnicity generate a two-girls effect equal to .088 (s.e.=.017) in the non-AA population, while the effect of two girls in the AA sample is larger by .051 (s.e.=.028). In contrast, the two boys effect is only .055 (s.e.=.016) in the non-AA population, and the AA two-boys effect is *smaller* by .038 (s.e.=.026). As a result, the AA population appears to increase childbearing in response to the birth of two girls but not in response to the birth of two boys.

The sex-composition first-stage in the 3+ sample captures the effect of an all-boy or all-girl triple on first- and second-born subjects, controlling for the sex-composition of earlier births. The first-stage therefore conditions on b_{12i} and g_{12i}, as well as a subject-sex main effect and a birth order dummy. Additional variables included in these models are dummies for the sex of the third child, an effect which is defined conditional on a *mixed-sex* sibling pair at first and second birth (because for families with b_{12i}=1, the boy-third effect is the same as having an all-male triple, while for families with g_{12i}=1, the boy-third effect is the same as having an all-female triple). The resulting model can be written as follows (we spell out notation only for the model that allows for separate all-male and all-female effects):

$$c_i = X_i'\beta + \gamma_1 b_i + \delta_b b_{12i} + \delta_g g_{12i} + \gamma_3(1-s_{12i})b_{3i} + \lambda_b b_{123i} + \lambda_g g_{123i} + \eta_i, \tag{5}$$

[17]For example, $g_{12i} = 1-b_{1i}-b_{2i} + b_{12i}$. Control for boy-first and boy-second main effects is motivated by the fact that the same-sex interaction term is, in principle, correlated with the main effects (Angrist and Evans, 1998) when the probability of male birth exceeds .5. In practice, however, this matters little because both the correlation is small and because the main effects are small.

where b_{123i} and g_{123i} are indicators for all-male and all-female triples and b_i is subject sex (i.e., b_{1i} for first-borns and b_{2i} for second-borns).[18] The term b_{3i} (boy-3) is also used as an instrument, though we postpone a discussion of the associated first stage for the moment. The sex-composition effects in this model are reported in columns 5 through 12 of Table 4.

The overall same-sex effect in the 3+ sample is .12 among first- and second-borns. This can be seen in column 9 of Table 4 (results for first-borns only, reported in column 5-8, are similar). The AA interaction term generates a large ethnic differential in sex-composition effects. For example, the same-sex effect among first- and second-born non-AA subjects, reported in column 10 of Table 4, is .070 (s.e.=.019), while the AA subsample responds to a same-sex triple by more than twice as much. This again contrasts with the twins estimates, where first-stage effects are smaller in the AA subsample.

First-stage effects in the 3+ sample show large differences when stratified by both sex and ethnicity, as can be seen in columns 7-8 and 11-12 of Table 4. The overall effect of three girls on first- and second-borns is 0.183 (s.e.=.022), almost triple the corresponding effect of three boys, 0.065 (s.e.=.021). The effect of three girls is also much larger in the AA population. The estimate for non-AA in column 12 is .072 (s.e.=.027) and the increment for AA is .217 (s.e.=.043), so that the effect of three girls in the first- and second-born AA subsample is .29 (.26 for first-borns only). This is considerably larger than the twins effect on AA subjects in the 2+ sample.

Heterogeneity and Non-linearity in the Response to Sibling-sex Composition

The difference in first-stage effects by AA status documented in Table 4 shows that the population of sex-composition compliers is disproportionately more likely to be of AA background. This is especially true for the response to an all-girl sibship. For example, the two-girl effect on AA fertility is

[18]This model is almost saturated in the sense that it controls for all lower-order interaction terms in the estimation of the effects of the two samesex triples except for one: in the $(1\text{-}s_{12i})b_{3i}$ term, we don't distinguish mixed sibling pairs according to whether a boy or girl was born first. A saturated model can be obtained by replacing the single term, $(1\text{-}s_{12i})b_{3i}$, with two terms, $b_{1i}(1\text{-}b_{2i})b_{3i}$ and $b_{2i}(1\text{-}b_{1i})b_{3i}$. In practice, this substitution matters little.

.14, while the EA effect is about .09. The AA differential in the effects of sex-composition on family size is largest for the response to same-sex triples. This pattern stands in marked contrast to the composition of twins-compliers, among which the AA subsample is under-represented. Thus, any comparison of twins and sex-composition IV estimates is implicitly a comparison for very different groups.

A second noteworthy distinction between the sex-composition and twins first-stages is in the different ranges of effects traced out by the two types of instruments. As we noted above, the twins-2 instrument in the 2+ sample increases family size from 2 to 3 with relatively little effect at higher parities, while the twins-3 in the 3+ sample primarily increases family size from 3 to 4, with virtually no other impact on fertility. In contrast, a same-sex sibship leads some families to keep having children at higher parities in pursuit of a more balanced sex composition.

The distribution shift due to sex-composition in the 2+ sample is documented in Figure 2, which reports first-stage estimates of effects of b_{12i} and g_{12i} on $d_{ji} \equiv 1(c_i \geq j)$, for j up to 11, along with the associated confidence bands. In the AA population, b_{12i} increases the likelihood that families have 3 or more children, with no significant effects at higher-order births. In contrast, the effect of two girls on d_{ji} increases from j=2 to j=3, and then tails off gradually, with a marginally significant effect on the likelihood of having 7 or more children. Effects in the non-AA population drop off more sharply as the number of children increases, and are similar for two boys and two girls. If anything, the non-AA population seems to increase childbearing more sharply in response to two boys than to two girls.

The CDF differences plotted in Figure 2 imply that sex-composition instruments capture an average causal effect which reflects the effect of having as many as seven children in the AA population and as many as six children in the non-AA population. The *range* of fertility variation induced by sex composition is even wider in the 3+ sample. This can be seen in Figure 3, which reports CDF differences in response to b_{123i} and g_{123i}, along with the associated confidence bands. The figure shows that, in the AA population, b_{123i} increases the likelihood of having 4 or more children, with a small and marginally significant effect on the likelihood of having 5 or more children. The effect of three boys is similar in the

AA and non-AA population. In contrast, the effect of three girls differs considerably by ethnicity, reaching .29 for three girls in the AA sample. Also in the AA population, the effect of g_{123i} increases from k=3 to k=4 and then diminishes gradually for higher values of k, remaining marginally significant even at k=10. In the non-AA population, in contrast, the effect of g_{123i} is considerably smaller and differs little from the effect of b_{123i}.

C. The Boy-3 Instrument

The bottom rows of columns 5-12 in Table 4 show the effect of having a boy at third birth in families with a mixed-sex sibship at first and second birth. We expect the boy-3 instrument to operate through preferences for male children that are common in more traditional Israeli households. In addition to providing additional variation, the boy-3 instrument is useful because it is implicitly used only for families with a mixed-sex sibship at parities one and two. The boy-3 instrument is therefore unlikely to be subject to the same violations of the exclusion restriction as instruments derived from sex-mix.

A boy at third birth reduces childbearing in the families of first- and second-borns with a mixed-sex sibship by .077 (s.e.=.015). Models allowing different coefficients by ethnicity generate an effect of -.044 (s.e.=.019) in the non-AA population, while the AA interaction term adds a further .064 (s.e.=.030) to this reduction. Figure 4 summarizes the effects of b_{3i} on fertility increments separately by ethnicity. The sample used to construct this figure includes both first- and second-borns.

Figure 4 shows that, as with the sex-mix instruments, boy-3 affects fertility over a wider range than do multiple births. In the AA population, in particular, b_{3i} reduces the likelihood of having more than 4 children as well as the likelihood of higher order births, up to 7, beyond which the effect is no longer significant. In the non-AA population, on the other hand, b_{3i} reduces the likelihood of having 4 or more children, with no significant effect at higher order births.

D. Instrument Validity

A possible concern in any IV study is correlation between the instruments and potential outcomes, either because of confounding or violations of the exclusion restriction. As in the Angrist and Evans (1998) study using sex-composition instruments, however, there is no relation between sex-mix and any of the background variables or covariates in our matched data set (detailed results available on request). We also replicated the common finding that twin births are associated with older maternal age. For example, the mothers of first- and second-borns who had twins at second or third birth were .3-.5 years older at first birth than those who had singletons. Twinning is not otherwise associated with subject demographics with one exception: in the 1995 sample of 2+ subjects, twin rates are higher for younger cohorts. Since twins can be identified only when birth records are complete, the fact that the quality of birth records improved over time seems likely to explain this finding. In any case, the 3+ sample does not exhibit this pattern. Because the results are similar in the 2+ and 3+ samples, the change in quality of birth records seems unlikely to have had a major impact on our findings.

It's also worth noting that multiple-birth-enhancing fertility treatments, a possible source of bias when using twins instruments, became available in Israel only in the mid 1970's. The effect of this on twin rates is first evident in vital statistics data starting in the mid-1980's (Blickstein and Baor, 2004). Since fewer than five percent of the third-born siblings in our 3+ sample and fewer than one percent of second-born siblings in our 2+ sample were born after 1984, the spread of fertility treatments is unlikely to be a factor in our analysis.

A further concern with twins instruments, raised by Rosenzweig and Zhang (2009), is the possible violation of exclusion restrictions due to the fact that twins have lower average birth weight than singletons, and perhaps worse health or cognitive achievement later on. Rosenzweig and Zhang (2009) argue that some parents therefore allocate resources away from twins, towards older singleton-birth children. Such parental behavior may offset any quantity-quality effects, making them harder to find using twins instruments to estimate effects on non-twins.

To see if resource reallocation is a problem, we estimated reduced-form twins effects on outcomes in samples where twins have little effect on family size. If the Rosenzweig and Zhang household resource story is true, first and second born children who have younger twin siblings should come out better in these “no-first-stage samples” since they benefit from the resources shifted away from less promising twins, with no off-setting increase in family size. Households with little or no twins first stage include those likely to have large families anyway, such as mothers who gave birth early and/or space births closely. Columns 1-3 of Table A2 therefore reports estimated reduced-form twin effects in 2+ samples with young mothers (first birth before age 21), closely spaced births (less than 2 years), and AA ethnicity. The first-stage effect of twins-2 in the young-mother and closely spaced samples are small and insignificant; the first-stage effect of twins-2 in the AA subsample is much smaller than in the complementary sample, though marginally significant at .177 (s.e.=.09).

The results presented in Table A2 fail to support the claim that parents favor older children following a twin birth. Although there is no first stage among mothers who gave birth before 21, there are no twins effects on the outcomes of the first born child, as can be seen in column 1 of the table. The reduced-form effects of twinning are also zero in samples stratified by birth spacing and ethnicity. [19]

The same sex instruments might also violate the exclusion restriction, a possibility raised by Rosenzweig and Wolpin (2000). Specifically, Rosenzweig and Wolpin (2000) argue for pure sex-composition effects on family size due to household efficiencies in families with samsesex sibships. To check this, columns 4-6 of Table A2 report on an investigation that parallels the no-first-stage investigation for twins instruments. The no first-stage samples for sex composition are again defined as those with young mothers, from families with tight spacing, and AA subjects, looking in all cases at first-

[19] We also implemented a version of the twins/non-twins strategy discussed by Rosenzweig and Zhang (2009). This is discussed at the end our results section. As an additional check, we also followed a referee’s suggestion and regressed outcomes in the 4+ sample on a twins-2 dummy and outcomes in the 5+ sample on a twins-3 dummy. This provides a specification check since fertility in these higher-parity samples is largely unaffected by earlier twinning (though the sample construction conditions on endogenous variables). These results also show no QQ trade-off.

born boys in the 2+ sample. Because sex composition has no effect on family size in these subsamples, effects of confounding factors related to sex composition should therefore surface. Consistent with a causal interpretation of the sex-composition IV estimates, however, there is no reduced-form relation between the two-boy instrument and any outcome variable in any subsample.

IV. OLS and 2SLS Estimates

When estimated using separate 2+ and 3+ samples, the causal effect of interest is the coefficient ρ in the model

$$y_i = W_i'\mu + \rho c_i + \varepsilon_i \tag{5}$$

where y_i is an outcome variable and W_i includes the covariates X_i, as well as instrument- and sample-specific controls (e.g., b_i). As discussed in the previous section, 2SLS estimates of this equation capture siblings' weighted average causal response to the birth of an additional child for those whose parents were induced to have an additional child by the instrument at hand. The outcome variables measure human capital, economic well-being, and social circumstances. In particular, we look at measures of subjects' educational attainment (highest grade completed and indicators of high school completion and college attendance), labor market status (indicators of work last year and hours worked last week) and earnings (monthly earnings and the natural log of earnings for full time workers), marital status (indicators of being married at census day and married by age 21) and fertility.

A. The 2+ Sample

As is typical for regressions of this sort, OLS estimates of the coefficient on family size in equation (5) indicate a negative association between family size and measures of human capital and economic circumstances. Larger families are also associated with earlier marriage and increased fertility. These results can be seen in column 2 of Table 5, which presents OLS estimates for first-borns in the 2+ sample (column 1 reports the means). Not surprisingly, given the sample sizes, all the OLS estimates are very precise. Control for covariates reduces but does not eliminate this negative relationship, as can be seen in column 3 of the table.

In contrast with the negative OLS estimates, 2SLS estimates point to zero or even positive effects. These results appear in columns 4-8 of Table 5, which report 2SLS estimates for different sets of instruments. For example, the effect on schooling estimated using twins instruments with AA interaction terms, reported in column 5, is .105 (s.e.= .131). The corresponding estimates using sex-composition instruments with AA interaction terms, reported in column 7 is .222 (s.e.=.176).

To increase precision, we also estimated specifications that combine twins and sex-composition instruments *within* a given sample (in this case, 2+) to produce a single, more efficient IV estimate. Although each instrument potentially generates its own local average treatment effect, the combination of instruments in this context can be justified by the desire to pin down what appears to be a common effect (of zero) as precisely as possible.

Combining both twins and sex-composition instruments generates an estimate of .16 (s.e.=.106), reported in column 8.[20] The combination of instruments generates a substantial gain in precision relative to the use of each instrument set separately; the schooling effect in the first row of column 8 is significantly different from the corresponding OLS estimate of −.145 reported in column 3. Likewise, the estimated effect on college attendance is small, positive, and reasonably precise.

This discussion highlights the fact that a key concern with the IV analysis is whether the estimates are precise enough to be informative. Of particular interest is the ability to distinguish IV estimates from the corresponding OLS benchmark. As it turns out, the estimates in column 8, constructed by pooling twins and sex-composition instruments with AA interaction terms, meet this standard of precision remarkably often. In particular, 6 out of 7 estimates of effects on non-marriage and fertility outcomes presented in this column are estimated precisely enough that the associated 95% confidence interval exclude the corresponding OLS estimates reported in column 3. Moreover, most estimates of

[20] The combined first-stages are reported in the appendix.

effects on schooling are very close to zero. A few of the estimated effects on college attendance are significant and positive, though given the large number of reported effects, this may be a chance finding.

A second set of noteworthy results are those for marriage and fertility. The IV estimates of effects on marital status suggest that subjects from larger families are more likely to be married and got married sooner. Using both twins and sex-composition instruments, the estimated effects on marital status are significantly different from zero and substantially larger than the corresponding OLS estimates. On the other hand, the marriage effects generated by sex-composition instruments are larger than the twins estimates, a point we return to below.

The marriage effects are paralleled by (and are perhaps the cause of) an increase in fertility: the combination-IV estimate of the effect on the probability of having any children is 0.079, four times larger then the corresponding OLS estimate, 0.019. In addition to the likelihood that increased marriage rates increase fertility, these fertility effects may reflect an intergenerational causal link in preferences over family size, a possibility suggested by Fernandez and Fogli (2005).[21]

B. The 3+ Sample

Estimates in the 3+ sample, reported in Table 6, are broadly similar to those for the 2+ sample, though there are some noteworthy differences. Columns 2-6 in Table 6 parallel columns 4-8 in Table 5 in that they report results from a similar sequence of instrument lists, with the modification that the twins instruments were generated by the event of a multiple 3rd birth and the sex-composition instruments are dummies for same-sex triples. A further change in Table 6 is the addition of a column (7) which reports results combining all instruments (with AA interaction terms) and a dummy for boy-3 (also with an AA interaction term). This addition provides a modest further gain in precision.

[21]To see whether the earnings effects are driven by the fact that many subjects are in their early 20s, we re-estimated the models in Table 5 restricting the sample to those aged at least 30. The results based on this restricted sample, which are not reported here but are available from the authors, are virtually identical to those using the full sample.

The OLS results in Tables 5 and 6 are virtually identical. The 2SLS estimates in the 3+ sample exploit more sources of variation than were used to construct estimates in the 2+ sample, so here we might expect some differences. The first key finding, however, is preserved: 2SLS estimates using both twins and sibling-sex composition generate no evidence of an adverse effect of larger family size on human capital or labor market variables. Moreover, as in Table 5, a few of the estimated effects on schooling outcomes are positive and (marginally) significant, though the significant estimates are fewer and smaller in this case. The marriage effects in the 3+ sample are also smaller and less consistently significant than in the 2+ sample. In particular, the twins instruments generate no significant marriage estimates when used alone, though they are still positive. Likewise, there are no longer any significant fertility effects.

As a check on the exclusion restrictions for sex-composition instruments, we also looked at estimates omitting these instruments but retaining boy-3. These results, reported in column 8, again provide no evidence of any adverse effects of family size. In general, same-sex instruments appear to generate smaller 2SLS estimates (i..e., closer to zero or less likely to be positive) than do twins instruments or the combination of twins with boy-3. This is inconsistent with Rosenzweig and Wolpin's (2000) conjecture regarding possible beneficial effects of having a sibling of the same sex. The boy-3 instrument may also have direct effects, as suggested by Butcher and Case (1994) for girls, but others have found little evidence for this (e.g., Kaestner, 1997).

Interpreting Average Causal Response

The results in Tables 5 and 6 are largely consistent across instruments, samples, and subjects' birth order. This is important because, as shown in the previous section, different instruments shift the fertility distribution very differently for different ethnicities. Moreover, sex-composition instruments shift fertility over a wide range of parities, with substantial shifts in large families, especially for the AA sample. Twins instruments, by contrast, increase completed fertility close to the parity where a multiple birth occurred. The twins and sex-composition IV estimates therefore capture the effects of different fertility increments. A related point is that the fertility shifts induced by both sets of instruments are over very different ranges in the 2+ and 3+ samples. Finally, we might expect different types of instruments to have different omitted variables biases, if any. Over-identification tests generate a formal measure of the equality of a set of IV estimates in models with multiple instruments (see, e.g., Angrist, 1991). Although not reported here in detail, the over-identification tests for the 2SLS estimates in Tables 5 and 6 generate no evidence of significant differences across instrument sets.

Because the effects of a family size on older siblings might differ at different ages (perhaps because parental investments before a fertility shock are unaffected by the shock), it's also noteworthy that multiple birth and sex composition experiments expose children to an increase in family size at a wide range of ages. For example, first-born children in the 2+ sample were about 7 years old on average when a singleton third child was born but only 4 years old upon the arrival of a twin. Similarly, first-born children in the 3+ sample were about 9.5 years old when a singleton fourth child was born but only 7.75 years old when the fourth-born was a twin. On average, first-born children exposed to a parity-six singleton birth were about 12 years old at the time. We also observe significant ethnic variation in age of exposure due to tighter birth-spacing in AA families. Of course, as noted in the introduction, we have no direct evidence on the effects of family size on the last child born (and indeed it is hard to imagine how these effects could ever be credibly identified except perhaps for adoptees). Neverthess, the consistency of our results across widely ranging parities and ages of exposure weighs against substantially heterogeneous effects by birth order.

The Rosenzweig and Zhang (2009) Bounding Strategy

As noted in the discussion of instrument validity, Rosenzweig and Zhang (2009) argue that lower average birth weight may induce some parents to allocate household resources away from twins, towards older singleton-birth children. Such parental behavior could offset any quantity-quality effects, making them harder to find in studies using twins instruments to estimate effects on older non-twin siblings. As discussed above, our direct investigation of the exclusion restriction generated no evidence of this behavior. Nevertheless, we explore the Rosenzweig and Zhang argument further here.

Rosenzweig and Zhang (2009) specifically argue that estimates of the effect of twins-2 on first-borns underestimate the quantity-quality trade-off. To avoid this bias, they suggest that comparisons of twins and non-twins at parity 2 be taken as an upper bound on the magnitude of the quantity-quality effect.[22] Comparisons of twins and non-twins tend to *overestimate* any negative effects of larger family size because twins have lower average birth weight than non-twins and may differ in other ways. Rosenzweig and Zhang (2009) therefore also suggest that when looking at the impact of twins on non-twins, it's useful to control for birth weight as a measure of twin quality, though control for birth weight is problematic because birth weight is an endogenous variable that is itself affected by twinning. We don't have data on birth weight but we can compare twins and non-twins in the spirit of Rosenzweig and Zhang's (2009) suggestion that such comparisons provide an upper bound on quantity-quality trade-offs.

Here, the bounding approach is implemented by comparing twin and non-twin outcomes for second- and third-born individuals using regression models similar to those used to produce our OLS and 2SLS estimates. As before, these regressions control for gender, age, missing month of birth, mother's age, mother's age at first birth, mother's age at immigration, father's and mother's place of birth, and census year. The regression-adjusted twin/non-twin comparisons show no significant differences between twins and singletons for outcome variables related to schooling, earnings, or labor supply.[23] Thus, the sort of contrasts seen by Rosenzweig and Zhang as bounding the size of the causal effect of interest also produce an estimate of zero in our data.[24]

---

[22] "The effect of twinning at the second pregnancy on the outcomes of second- (first-) birth children provides an upper (lower) bound on the average negative effect on child outcomes of increasing family size" (Rosenzweig and Zhang, 2009, page 1157).

[23] Detailed tables showing twin/non-twin contrasts are available upon request.

[24] Twins usually have lower birth weight (LBW) than singletons, but the question of whether this matters for adult outcomes remains controversial (see, e.g Almond et al., 2005; Behrman and Rosenzweig, 2004; Conley, 2006;

C. Combining 2+, 3+, 4+, and 5+ Samples

To further increase precision we also pooled estimates *across* the 2+, 3+, and two higher-parity samples. For example, we constructed a single twins-IV estimate using t_{2i}, t_{3i}, t_{4i}, and t_{5i} as instruments in a data set that implicitly stacks the 2+, 3+, 4+, and 5+ samples, while restricting the IV estimates from the different parity-specific sub-samples to be the same. Because the instrument list and conditioning variables are different in each parity-specific subsample, this procedure requires a modification of conventional 2SLS.

The Parity-Pooled Setup

Our pooled analysis works with the union of subjects from 2+, 3+, 4+, and 5+ sub-samples. The total sample therefore includes individuals who are first-born subjects in the 2+ sample, first- and second-born subjects in the 3+ sample, first-through-third-born subjects in the 4+ sample, and first-through-fourth-born subjects in the 5+ sample. In other words, the sample includes all birth orders up to $p-1$ from families with at least p children, for $p \leq 5$. The p+ sub-samples are not mutually exclusive; for example, a given first-born subject in the 5+ sample must also be a member of the 2+, 3+, and 4+ sub-samples.

The restriction that motivates pooled estimation is that the causal effect of childbearing is a constant, denoted ρ_0 (Tables 5 and 6 suggest $\rho_0=0$). In terms of potential outcomes, we have

$$Y_i(j) = Y_{0i} + \rho_0 \cdot j. \tag{6}$$

In addition, let $Y_{0i} = X_i'\mu_0 + \upsilon_i$ denote the regression of Y_{0i} on X_i in the population from which the parity-pooled sample is drawn. The residual, υ_i, is orthogonal to X_i in this population by construction. The observed outcome, y_i, is linked to this causal model by

$$y_i = X_i'\mu_0 + \rho_0 c_i + \upsilon_i. \tag{7}$$

Note that the residual, υ_i, may be correlated with c_i.

Black et al., 2007 and Royer, 2009). Many of these studies use within-twin comparisons in birthweight to control for omitted factors. At the same time, it is also unclear whether the effects of LBW within twins are very general.

The following Lemma provides the econometric justification for pooled estimation:

LEMMA. Let d_{pi} denote membership in a p+ sample and let Z_{pi} denote an instrumental variable satisfying $Z_{pi} \perp\!\!\!\perp Y_{0i} \mid W_{pi}, d_{pi}=1$, where W_{pi} includes X_i, plus possibly additional instrument-specific controls. Let $Z_{pi}^* = Z_{pi} - W_{pi}'\Gamma$ where Γ is the coefficient vector from a regression of Z_{pi} on W_{pi}' in the p+ population. Assume there is a first stage for Z_{pi}, i.e., $E[Z_{pi}^* c_i| d_{pi}=1] \neq 0$. Then $E[d_{pi}Z_{pi}^* \upsilon_i]=0$ where υ_i is the error term in (7) and the expectation is taken in the population containing subjects of birth order up to p−1 from families with at least p children.

PROOF: $E[d_{pi}Z_{pi}^*\upsilon_i] = E[d_{pi}Z_{pi}^*(y_i - X_i'\mu_0 - \rho_0 c_i)] = E[Z_{pi}^*(y_i - X_i'\mu_0 - \rho_0 c_i)| d_{pi}=1]P[d_{pi}=1]$. Note that $E[Z_{pi}^* X_i| d_{pi}=1]=0$ by construction. Given the constant-effects causal model, (6), and the conditional independence assumption at the beginning of the lemma, $\rho_0 = E[Z_{pi}^* y_i| d_{pi}=1]/ E[Z_{pi}^* c_i| d_{pi}=1]$. Therefore $E[Z_{pi}^*(y_i - \rho_0 c_i)| d_{pi}=1]=0$. ♦

This Lemma shows how a common causal parameter can be estimated in a parity-pooled sample. For example, we can combine t_{2i} in the 2+ sample and t_{3i} in the 3+ sample. The data set required for this is the union of the 2+ and 3+ samples, i.e., first-borns in the 2+ sample, and second-borns in the 3+ sample (first-borns in the 3+ sample are included in the 2+ sample). After partialing out the relevant set of covariates as described in the lemma, $d_{2i}[t_{2i}]^*$ and $d_{3i}[t_{3i}]^*$ are valid instruments for equation (7) in the pooled {2+ U 3+} sample. Similarly, we can combine $d_{2i}[b_{12i}]^*$, $d_{2i}[g_{12i}]^*$, $d_{3i}[b_{123i}]^*$, $d_{3i}[g_{123i}]^*$, and $d_{3i}[(1-s_{12i})b_{3i}]^*$, where the first-step regression-adjustment of each instrument accounts for the fact that sex-composition instruments involve different sets of controls in the 2+ and 3+ sample, in addition to the set of common covariates, X_i.

Before turning to a discussion of parity-pooled empirical results, we briefly discuss first stage relations in higher parity samples, focusing on the sex-composition instruments. Figures 5 and 6 report the effects of sex composition on fertility in the 4+ sample, using a format similar to the one used in Figures 3 and 4. The figures document the fact that sex mix sharply increases family size in this sample. Effects are again larger for all-female than for all-male sibships and for the AA population. In the AA samples, an all-girl sibship increases the likelihood of family sizes as large as nine. A full set of first-stage estimates is given in Appendix Table A3. The largest first-stage effect for sex mix is .365 (=.242+.123) among the AA population as a result of five girls. On the other hand, an all-male sibship still increases fertility in both the 4+ and 5+ samples. The effect of a multiple fourth birth, reported in column 5, is almost one child for non-AA Jews in the 4+ sample. For this group, the twins experiment amounts to a randomized trial with perfect compliance.

Parity-Pooled Results

The empirical strategy using parity-pooled samples leads to a considerable gain in precision, while most of the estimated effects on outcomes other than marriage and fertility remain small and insignificant. This can be seen in Table 7, which reports pooled results using twins instruments in columns 1-3, results pooling sex composition instruments in columns 4-6, and the results of pooling all instruments in columns 7-9. The table shows results from three samples for each instrument set: the union of subjects from the 2+ and 3+ sub-samples, the union of subjects from 2+, 3+ and 4+ sub-samples, and the union of subjects from 2+, 3+, 4+, and 5+ sub-samples. For example, the estimated effect on highest grade completed using all available twins instruments in the union of the 2+, 3+, 4+, and 5+ samples is .031 (s.e.=.055), shown in the first row of Table 5. The corresponding estimate using all available sex composition instruments is .054 (s.e.=.068), in column 6.

The estimates combining both twins and sex composition instruments in the union of 2+, 3+, 4+, and 5+ samples, reported in column 9 of Table 7, are the most precise we have been able to construct. For example, the estimated effect on highest grade completed is .040 (s.e.=.043), in comparison with .072 (s.e.=.076) in Table 6. Similarly, the estimated effect on annual employment is .005 (s.e.=.009), compared to .035 (s.e.=.017) reported in Table 6. All estimates of effects on non-marriage and fertility outcomes in column 9 of Table 7 generate confidence intervals that exclude the corresponding OLS estimates with covariates.

Most of the parity-pooled estimates of effects on marriage and some of the effects on fertility remain at least marginally significantly different from zero. For example, the estimated effect on marriage using twins instrument in the pooled 2+,3+,4+, and 5+ sample is .023 (s.e.=.010), and the corresponding estimate using sex-composition instruments is .045 (s.e.=.012). While sex composition instruments generate larger effects on marriage than do the twins instruments, the fact that this effect turns up in both IV strategies suggests the IV estimates reflect the causal effect of childbearing and not just a propensity for older girls to marry in response to the birth of a younger sister, a point discussed further in the next section.

D. Analyses by Ethnicity and Gender

Large numbers of Sephardic Jews came to Israel from the Arab countries of Asia and North Africa in the 1950s. Although fertility among Sephardic Jews ultimately fell to close to the Israeli average, the AA cohorts in our sample come from much larger families than other Jews. While almost 60 percent of AA Jews in the 2+ sample come from families with 4 or more children, only 26 percent of other Jews in the sample come from families this large.

In addition to having higher fertility, the AA group is less educated and poorer than other Jewish ethnic groups. For example, only 12 percent of AA Jews in our 2+ sample are college graduates, while the overall college graduation rate in the 2+ sample is 20 percent. The gap in living standards by ethnicity is especially big in larger households. Among those born in Israel, the average 1990 income in AA households with 5 or more members was about 60 percent of the income of similarly-sized European-American households, only 15% larger than the income of non-Jews (Central Bureau of Statistics, 1992, Table 11.4). These differences suggest estimates in the AA sub-sample may be especially relevant for poorer populations.

OLS estimates by ethnicity, reported in columns 1 and 3 of Table 8, generally show somewhat larger adverse effects on schooling and labor market outcomes in the non-AA sample than in the AA sample. All of the 2SLS estimates in Table 8 are for the full parity-pooled sample including the union of subjects from 2+, 3+, 4+, and 5+ families and using the full set of instruments. The resulting 2SLS estimates by ethnicity reported in columns 2 and 4 generate no evidence of an effect on human capital or labor market variables for either ethnic group. For example, the estimated effect on highest grade completed in the non-AA sample is .043 (s.e.=.064), while the corresponding estimate for AA is .031 (s.e.=.057). The estimated effects on hours worked is .30 (s.e.=.87) for non-AA subjects and .45 (s.e.=.59) for AA.

As in the sample that does not differentiate by ethnicity, there is again evidence for an effect of family size on marriage rates or timing in both groups. For example, the estimated effects on marriage are .030 (s.e.=.011) for non-AA subjects and .035 (s.e.=.010) for AA subjects. Effects on early marriage are almost identical in the two groups. The effects on fertility are also positive in both samples and are slightly larger for the AA population.

Also of interest are separate estimates for men and women, especially in view of the effects on marital status discussed above. We therefore estimated separate models by sex using the full set of instruments in the largest parity-pooled sample, with results reported in columns 6 and 8 of Table 8. The OLS estimates reported in columns 5 and 7 are similar for men and women. Again, however, 2SLS estimates by sex show no evidence of negative effects on schooling or labor market variables for either group. For example, the estimated effects on log earnings are .009 (s.e.=.031) for men and .015 (s.e.=.026) for women.

2SLS estimates of effects on marriage rates are more pronounced for women than for men, and more precise. For example, the effect on women, reported in column 8, is .04 (s.e.=.009), while the corresponding effect for men, reported in column 6, is .021 (s.e.=.011). Moreover, the estimated effect on early marriage for women is about 6-7 percentage points and significantly different from zero. In contrast, the corresponding estimate for men is negative and insignificant.

The consistency and relative precision of results across instrument sets suggests that early marriage may indeed be a consequence of increased family size, especially for older daughters. The marriage effects seem to generate a small effect on fertility as well (also apparent in Table 7). Stronger marriage effects for women may reflect the fact that marriage is the main route to an independent household for girls in traditional Jewish families. Moreover, older daughters in Israel may be tempted to marry sooner when crowded by younger sisters. This is consistent with traditional Jewish values and can be traced back to the Biblical story of Rachel and Leah's joint betrothal to Jacob. We might therefore expect marriage effects estimated using sex-composition instruments to be larger than effects estimated using twins instruments, as seems to be the case.

V. Possible Explanations

Exogenous increases in family size in a Becker-Lewis-type setup (due, say to a change in contraceptive costs; p. S283) should reduce child quality since an increase in quantity increases the shadow price of quality. Along these lines, Rosenzweig and Wolpin (1980) interpret twin births as a subsidy to the cost of further childbearing (p. 234). They argue that this price change should reduce quality unless quantity and quality are strong complements in parental utility functions. While the quantity-quality tradeoff is less clear-cut in more recent theoretical discussions, the traditional view provides an intellectual foundation for policies that attempt to reduce family size in LDCs.

The most important question our findings raise is what might account for the absence of a causal link between sibship size and later outcomes. A definitive answer to this question must await future empirical research. Here, we briefly review a number of possible explanations. One theoretical possibility is that, as far investment in human capital goes, parents use perfect capital markets to fund investment irrespective of resource constraints. It seems unlikely, however, that capital markets are so nearly perfect, especially in Israel during the period we are studying, when financial markets were not well-developed.[25]

A more relevant possibility is that parents adjust to exogenous increases in family size on margins other than quality inputs. For example, parents may work longer hours or take fewer or less expensive vacations (i.e., consume less leisure). Parents may also substitute away from personal as opposed to family consumption (e.g., by drinking less alcohol). Direct evidence on this point is difficult to obtain since consumption data rarely come in the form needed to replicate our research design.

The Angrist-Evans (1998) results for wives raise the possibility of an explanation linked to female labor supply. Clearly one effect of additional childbearing is to increase the likelihood of at-home child-care for older siblings (an effect also documented by Gelbach, 2002). It may be that home care is better, on average, than commercial or other out-of-home care, at least in the families affected by the fertility shocks we study. On the other hand, estimates of AE-98 type models for samples of Israeli mothers show only modest effects of child-bearing on labor supply (Marmer, 2000).[26]

On the institutional side, the quantity-quality trade-off within households may be partially offset by welfare payments and public schooling. If so, this may limit the external validity of our findings, or at least their applicability to countries with a less developed welfare state. On the other hand, the Israel setting is especially interesting because different cohorts were exposed to different institutions. The Israeli Compulsory Schooling Law enacted soon after the establishment of the Israeli Parliament in 1949 allowed 9 years of free and compulsory education starting from kindergarten until 8th grade. In 1969, the law was changed to provide two additional years of free and compulsory education, until 10th grade. In 1978, a further extension provided free (though not compulsory) schooling for grades 11th and 12th.

---

[25] Another theoretical possibility, outlined by Deaton and Paxson (1998), is that larger households are better off at the same level of per-capita expenditure due to household scale economies. This seems less relevant for our results since we are investigating effects without holding per-capita resources constant.

[26] During the childhood years of the subjects in our study, the labor force participation of married women aged 18-35 increased from 28.7% in 1961 to 52% in 1980. This increase, was accompanied with a sharp increase in preschool attendance of children aged 2-5: from 33% in 1965, to 50% in 1970, reaching 86% in 1980 (Ben Porath and Gronau, 1985)

During the childhood years of the subjects in our study, school enrollment of Jewish children below age 14 was about 95 percent whereas school enrollment for ages 14-17 increased from 67 percent in 1970 to 80 percent in 1980 (Central Bureau of Statistics, 1996). Despite this sharp increase in educational attainment, we find no significant cross-cohort differences in IV estimates of family size effects.

As in many developed and middle-income countries, Israel offers tax concessions to larger families in the form of child allowances. But these payments were low during the period members of our samples were young (Manski and Mayshar, 2003) and therefore seem unlikely to explain the absence of a quantity-quality trade-off. We confirmed this in an analysis introducing interaction terms for changes in eligibility for child allowances and the level of child allowances by cohort.

An additional explanation for the absence of a causal link between sibship size and the outcomes studied here might be called "marginally irrelevant inputs." Using research designs similar to ours, Caceres (2006) finds some evidence for a decreased likelihood of private school enrollment. However, private school attendance and early marriage (at least for girls) may matter little for human capital and earnings. A final explanation that is consistent with our findings is that the presence of siblings directly enhances child welfare, perhaps because children with siblings benefit socially or take on more responsibility sooner. This conjecture is consistent with Qian's (2004) IV estimates for China, which show that the presence of a younger sibling increases older children's school enrollment.

VI. Summary and Directions for Further Work

We study the causal link running from sibship size to human capital, economic well-being, and family structure using a unique sample combining population registry and census data. Our research design exploits variation in fertility due to multiple births and preferences for a mixed sibling-sex composition, along with ethnicity interactions and preferences for male children. The natural experiments embodied in these IV strategies capture a wide range of fertility variation.

The evidence reported here is remarkably consistent across research designs and samples: while all instruments exhibit a strong first-stage relation, and OLS estimates are substantial and negative, IV estimation generates no evidence for negative consequences of increased sibship size on outcomes. The estimates do suggest, however, that girls from larger families marry sooner. This marriage effect may have a modest effect on fertility, but it does not appear to reduce schooling, employment, or earnings. In future work, we hope to shed light on possible explanations by generating new evidence on the effect of family size on resource allocation across generations.

DATA APPENDIX

The Israeli population registry, our source of information on families of origin, contains updated administrative records for Israeli citizens and residents, whether currently living or dead, including most Israelis who have moved abroad. This data base also includes the Israeli ID numbers held by citizens and temporary residents. ID numbers are issued at birth for the native-born and upon arrival for immigrants. In addition to basic demographic information on individuals (date of birth, sex, country of birth, year of immigration, marital status, religion and nationality), the registry records parents' names and registrants' parents' ID numbers.

The construction of an analysis file proceeded by first using subjects' ID numbers to link to non-public-use versions of census long-form files that include ID numbers with registry records for as many subjects as we could find. In a second step, we used the registry to find subjects' mothers. Finally, once mothers were linked to census respondents, we then located all the mothers' children in the registry, whether or not these children appear in the census. In this manner we were able to observe the sex and birth dates of most adult census respondents' siblings.

The likelihood of successful matches at each stage of our linkage effort is determined primarily by the inherent coverage limitations of the registry. Israel's population registry was first developed in 1948, not long after the creation of the state of Israel. Census enumerators went from house to house, simultaneously collecting information for the first census and for the administrative system that became the registry. Later, the registry was updated using vital statistics data. Thus, in principle, the sample of respondents available for a census interview in 1983 and 1995 should appear in the registry, along with their mothers' ID numbers, if they were resident in 1948, born in Israel after 1948, or immigrated to Israel after 1948.

Assessing the Quality of the Registry Match

To assess whether the limitations in the matching process outlined above introduce a bias that might affect our estimates, we constructed a reverse match starting with the registry and going forward to the censuses. The reverse match includes Jews in the registry alive on the 1983 or 1995 census date. We applied the same sample restrictions related to mother's year of birth and year of immigration, mother's age at 1st birth, and subject's year of birth used to construct the main extract. Table A1 reports the effect of family size and the instruments on the probability of appearing in either the 1983 or 1995 census data using the reverse match. Each cell reports estimates of a separate regression. Estimates reported in even columns come from regression that control for an indicator for age at census date, missing month of birth, mother's age at census data, mother's age at first birth, mother's age immigration, and father's and mother's place of birth. Columns 1-4 report estimates for 1st borns in 2+ families; columns 5-8 and 9-12 reports estimates from a sample of 3+ families, 1st born and 2nd born, respectively. This table suggests match rates are unrelated to the instruments. As noted in the text, there is a small decrease in the 1983 match rates with family size, but this is unlikely to affect our empirical strategy. We also reproduced the first-stage regressions using the entire registry population alive on census date and the reverse match sample. Both samples produced virtually identical first-stage estimates to those reported in the paper.[27]

[27] A table with first stage estimates based on these two samples is available upon request.

REFERENCES

Almond, D. K. Y. Chay, and D. S. Lee. "The Costs of Low Birth Weight," *Quarterly Journal of Economics*, 120(3), 2005, pp. 1031–83.

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

Angrist, J. and G. Imbens, "Average Causal Response with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90 (430), 1995, 431-442.

Becker, G. and H. G. Lewis, "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy* 81(2) part 2, 1973, pp. S279-S288.

Becker, G. and N. Tomes, "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy* 84(4) part 2, 1976, pp. S143-S162.

Behrman, J., R. Pollak and P. Taubman, "Family Resources, Family Size, and Access to Financing for College Education," *Journal of Political Economy* 97(2), 1989, pp. 398-419.

Behrman, J. R., and M. R. Rosenzweig, "Returns to Birthweight," *Review of Economics and Statistics*, 86(2), 2004, pp. 586–601.

Ben-Porath, Y. and R. Gronau, "Jewish Mother Goes to Work: Trends in the Labor Force Participation of Women in Israel: 1955-1980", Journal of Labor Economics, Vol. 3 No. 1 Part 2, pp. S310-S327, 1985.

Black, S., P. J. Devereux and K. G. Salvanes, "The More the Merrier? The Effect of Family Composition on Children's Education," *Quarterly Journal of Economics*, 120(2), 2005, pp.669-700.

Black, S., P. J. Devereux, and K. G. Salvanes. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics*, 122(1), 2007, pp.409–39.

Blickstein, I. and Baor L., "Trends in Multiple Births in Israel," *Harefuah,* 143(11), 2004, pp. 794-798.

Bongaarts, J., "Population Policy Options in the Developing World," *Science* 263 (February), 1994, pp. 771-776.

Butcher, Kristin F. and Anne Case, "The Effect of Sibling Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, 109, 1994, 531-563.

Caceres, J., "The Impacts of Family Size on Investment in Child Quality," Journal of Human Resources, 41(4), 2006, pp. 722-737.

Central Bureau of Statistics, *Demographic Characteristics of the Population: National Data from the Complete Enumeration*, 1983 Census of Population and Housing Publications No. 7, Jerusalem, 1985.

Central Bureau of Statistics, *Statistical Abstract of Israel 1992*, Jerusalem: Keter Press, 1992.

Central Bureau of Statistics, *Statistical Abstract of Israel 1996*, Jerusalem: Keter Press, 1996.

Central Bureau of Statistics, *Understanding the Census, a Dictionary of Terms, Methods and Main Findings of the 1995 Census of Population and Housing*, 1995 Census of Population and Housing Publications, No. 16, Jerusalem, 2001.

Cherlin, Andrew, "The Consequences of Welfare reform for Child Well-Being: What Have We Learned So Far and What are the Policy Implications," Paper presented at the American Sociological Association meetings, August, 2004.

Conley, D. and R. Glauber, "Sibling Similarity and Difference in Socioeconomic Status: Life Course and Family Resource Effects," *National Bureau of Economic Research*, Working paper 11320, 2005.

Conley, D. K. W. Strully, and N. G. Bennett, "Twin Differences in Birth Weight: The Effects of Genotype and Prenatal Environment on Neonatal and Post-Neonatal Mortality," *Economics & Human Biology*, 4(2), 2006, pp. 151–183.

Deaton, A. and C. Paxson, "Economies of Scale, Household Size, and the Demand for Food," *Journal of Political Economy*, 106(5), 1998, pp. 897-930.

Duflo, E., "Evaluating the Effect of Birth-Spacing on Child Mortality," mimeo, Department of Economics, Massachusetts Institute of Technology, 1998.

Fernandez, R. and A. Fogli, "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," NBER Working Paper 11268, April 2005.
Galor, O. and D. Weil, "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and beyond," *The American Economic Review* 90(4), 2000, pp. 806-828.
Gelbach, J., "Public Schooling for Young Children and Maternal Labor Supply," *American Economic Review* 92(1), 2002, 307-322.
Hanushek, E., "The Trade-off between Child Quantity and Quality," *Journal of Political Economy* 100(1), 1992, pp. 84-117.
Hazan, M. and B. Berdugo, "Child Labor, Fertility and Economic Growth," *The Economic Journal*, 112 (October), 2002, pp. 810-828.
Heston, Alan, Robert Summers and Bettina Aten, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002
Imbens, G.W. and J. Angrist, "Identification and Estimation of Local Average Treatment Effects," Econometrica 62(2), 1994, 467-75.
Johnson, D. G., "Population and Economic Development," *China Economic Review* 10, 1999, pp. 1-16.
Kaestner, Robert, "Are Brothers Really Better? Sibling Sex Composition and Educational Attainment Revisited," *Journal of Human Resources* 32, 1997, 250-284.
Lutz, W. and V. Skirbekk, "Policies Addressing the Tempo Effect in Low-Fertility Countries," *Population and Development Review* 31, 699-720, 2005.
Malthus, T., "An Essay on the Principle of Population," London, Printed for J. Johnson, In St. Paul's Church-Yard, 1798.
Manski, C. and Y. Mayshar, "Private Incentives and Social Interactions: Fertility Puzzles in Israel", Journal *of the European Economic Association*, 1 No. 1, March 2003, pp. 181-211.
Marmer, V. "The Relationship Between Family Size, Labor Supply and Labor Income in Israel," The Maurice Falk Institute for Economic Research, Supervised Papers Series No. 18, 2000.
Moav, O., "Cheap Children and the Persistence of Poverty," *The Economic Journal*, 115 (January), 2005, pp. 88-110.
Moffit, Robert, "Remarks on the Analysis of Causal Relationships in Population Research," *Demography* 42(1), February 2005, pp. 91-108.
Oreopoulos, Philip, "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws really Matter," *American Economic Review* 96(1), March 2006, pp. 152-175.
Qian, N., "Quantity-Quality and the One Child Policy: The Positive Effect of Family Size on School Enrollment in China," Thesis chapter, Department of Economics, Massachusetts Institute of Technology, 2004.
Rosenzweig, M. and K. Wolpin, "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 1980, pp. 227-240.
Rosenzweig, M. and K. Wolpin, "Natural 'Natural Experiments' in Economics", *Journal of Economic Literature,* 38(4), 2000, pp. 827-74.
Rosenzweig, M. and J.. Zhang, "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's 'One Child' Policy," Forthcoming, *The Review of Economic Studies*, 2009.
Royer H., "Separated at Girth: US Twin Estimates of the Effects of Birth Weight," American Economic Journal: Applied Economics, 1(1), 2009, pp. 49-85.
Schultz, T. P., "Effects of Fertility Decline on Family Well-Being: Evaluation of Population Programs," Draft for MacArthur Foundation Consultation meeting, Philadelphia, PA, 2005.
Weil, David, *Economic Growth*, Boston: Addison-Wesley, 2005.

Table 1

Match Rates and Sample Selection

	Israeli-born Father				Foreign-born Father							
					Foreign-born subject				Israeli-born subject			
	Subject born<1955		Subject born ≥ 1955		Subject born<1955		Subject born ≥ 1955		Subject born<1955		Subject born ≥ 1955	
	(1)		(2)		(3)		(4)		(5)		(6)	
A. 1995 census												
All subjects	9,453		56,534		118,633		72,340		58,767		161,331	
Matched to registry (N,%)	9,057	95.8%	54,073	95.6%	115,123	97.0%	68,788	95.0%	57,098	97.2%	156,096	96.8%
Matched mother + siblings (N,%)	1,573	16.6%	50,597	89.5%	7,600	6.4%	32,472	44.9%	11,351	19.3%	139,783	86.6%
Selected sample												
Mothers born ≥ 1930 whose age at 1st birth ε [15,45]	494		48,683		1,166		26,217		2,556		119,928	
of which: Israeli born mothers or immigrants who arrived since 1948 and before the age of 45	419		47,022		1,127		22,704		2,211		115,783	
of which: first and second borns of families with 2 or more births	349		34,778		1,008		15,443		1,937		67,952	
Estimated fertility coverage: 86%												
B. 1983 census												
All subjects	11,049		12,665		160,459		25,025		66,761		70,662	
Matched to registry (N,%)	9,704	87.8%	10,867	85.8%	140,932	87.8%	20,691	82.7%	60,105	90.0%	62,141	87.9%
Matched mother + siblings (N,%)	1,289	11.7%	9,258	73.1%	7,380	4.6%	14,557	58.2%	10,767	16.1%	50,785	71.9%
Selected sample												
Mothers born ≥ 1930 whose age at 1st birth ε [15,45]	421		7,854		1,065		9,197		2,438		34,560	
of which: Israeli born mothers or immigrants who arrived since 1948 and before the age of 45	318		6,952		1,045		8,913		2,138		32,368	
of which: First and second borns of families with 2 or more births	232		3,657		730		3,425		1,519		12,095	
Estimated fertility coverage: 79%												

Note.- The table reports sample sizes and match rates at each step of the link from census data to the population registry. The target population consists of Jewish Census respondents in 1995 and 1983 aged 18-60. The table also shows the impact of sample selection criteria on sample sizes.

Table 2
Analysis Samples

	Full			Asia-Africa		
	2+	3+		2+	3+	
	1st borns	1st borns	2nd borns	1st borns	1st borns	2nd borns
	(1)	(2)	(3)	(4)	(5)	(6)
1995 Census	0.758	0.753	0.775	0.706	0.705	0.732
Mother married or widowed in 2003	0.910	0.926	0.932	0.921	0.932	0.937
Endogenous variables						
# of children	3.63	4.22	4.32	4.31	4.67	4.76
More than 2 kids	0.74	1.00	1.00	0.87	1.00	1.00
More than 3 kids	0.400	0.545	0.573	0.593	0.686	0.704
Family composition						
Twins at second birth	0.009	0.006	-	0.008	0.006	-
Twins at third birth	-	0.010	0.010	0.008	0.009	0.009
Boy at first birth	0.517	0.518	0.527	-	0.518	0.528
Boy at second birth	0.514	0.515	0.507	0.516	0.514	0.504
Boy at third birth	-	0.515	0.517	-	0.509	0.516
Girl12=1	0.233	0.239	0.237	0.232	0.236	0.234
Boy12=1	0.265	0.272	0.272	0.265	0.267	0.267
Girl123=1	-	0.115	0.114	-	0.117	0.113
Boy123=1	-	0.140	0.140	-	0.138	0.138
Control Variables						
Age on census day	26.2	26.4	25.5	27.4	27.5	26.4
Year of birth	1966	1965	1967	1964	1964	1965
Mother's age on census day	49.1	48.8	50.4	49.7	49.5	50.7
Mother's year of birth	1943	1943	1942	1942	1942	1941
Mother's age at 1st birth	22.7	22.2	22.1	22.0	21.7	21.7
Mother's age at immigration (for non-Israeli mothers)	17.4	15.7	15.9	15.6	15.4	15.7
Mother's ethnicity						
Israel	0.344	0.354	0.315	0.167	0.161	0.138
Asia-Africa	0.397	0.468	0.507	0.792	0.805	0.830
Former USSR	0.115	0.068	0.064	0.011	0.009	0.007
Europe-America	0.144	0.111	0.113	0.030	0.025	0.025
Father's ethnicity						
Israel	0.274	0.282	0.248	-	-	-
Asia-Africa	0.426	0.501	0.535	1	1	1
Former USSR	0.114	0.068	0.068	-	-	-
Europe-America	0.186	0.149	0.148	-	-	-

Table 2
(cont.)

	Full			Asia-Africa		
	2+	3+		2+	3+	
	1st borns	1st borns	2nd borns	1st borns	1st borns	2nd borns
	(1)	(2)	(3)	(4)	(5)	(6)
Subject ethnicity						
Israel	0.836	0.869	0.887	0.856	0.852	0.878
Asia-Africa	0.061	0.074	0.065	0.144	0.148	0.122
Former USSR	0.066	0.029	0.024	0.000	0.000	0.000
Europe-America	0.037	0.027	0.025	0.000	0.000	0.000
Education Outcomes						
Highest grade completed	12.6	12.5	12.3	12.2	12.1	12.0
Schooling ≥ 12	0.824	0.813	0.802	0.759	0.754	0.752
Some College (age ≥ 24)	0.291	0.262	0.224	0.177	0.169	0.143
College graduate (age ≥ 24)	0.202	0.180	0.153	0.117	0.111	0.093
Labor Market Outcomes (age ≥ 22)						
Worked during the year	0.827	0.820	0.809	0.812	0.809	0.798
Hours worked last week	32.6	32.4	31.7	32.5	32.4	31.7
Monthly earnings (in 1995 Shekels)	2,997	2,920	2,721	2,847	2,820	2,621
Ln(earnings) for full time workers	8.24	8.23	8.18	8.20	8.19	8.15
Marriage and fertility						
Married on census day	0.446	0.465	0.418	0.519	0.530	0.479
Married by age 21 (age ≥ 21)	0.172	0.183	0.171	0.198	0.205	0.194
Number of own children (women only)	1.00	1.08	0.98	1.28	1.32	1.20
Number of observations	89,445	65,673	52,964	38,063	32,875	28,357

Note.- The table reports descriptive statistics for the main 3 analysis samples used in the paper. The 2+ sample consists of first-born census subjects from families with two or more births including the subject. The 3+ sample consists of first- and second-born census subjects from families with three or more births including the subject. The Asia-Africa subsample consists of census subjects whose fathers' ethnicity is identified as Asia-Africa in the census.

Table 3
Twins First Stage

	2+		3+			
	1st borns		1st borns		1st and 2nd borns	
	(1)	(2)	(3)	(4)	(5)	(6)
Twins-2	0.437	0.625	-	-	-	-
	(0.050)	(0.057)				
Twins-2 x Asia-Africa	-	-0.484	-	-	-	-
		(0.105)				
Twins-3	-	-	0.522	0.583	0.585	0.692
			(0.045)	(0.045)	(0.043)	(0.049)
Twins-3 x Asia-Africa	-	-	-	-0.132	-	-0.226
				(0.094)		(0.086)
Male	-0.018	0.000	0.016	0.018	0.014	0.006
	(0.010)	(0.012)	(0.018)	(0.023)	(0.011)	(0.015)
Male x Asia-Africa	-	-0.041	-	-0.005	-	0.015
		(0.022)		(0.035)		(0.022)
Asia-Africa	0.242	0.267	0.166	0.161	0.083	0.069
	(0.015)	(0.019)	(0.016)	(0.027)	(0.014)	(0.021)

Note.- The table reports first-stage effects of twins-2 and twins-3 on number of children. The sample includes non-twins aged 18-60 in the 1983 and 1995 censuses as described in Table 1. In addition to the effects reported, the regressions include indicators for age, missing month of birth, mother's age, mother's age at first birth, mother's age at immigration (where relevant), father's and mother's place of birth, and census year. Regressions for columns 3-6 include also controls for girl12, boy12 and twins at second birth. Regressions for columns 5-6 include also indicators for second born and birth spacing between first and second birth. Robust standard errors are reported in parenthesis. Standard errors in columns 5-6 are clustered by mother's ID.

Table 4

Sex-Composition First Stages

	2+				3+							
	First Borns				First Borns				First and Second Borns			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Samesex	0.073	0.071	-	-	0.133	0.095	-	-	0.120	0.070	-	-
	(0.010)	(0.012)			(0.017)	(0.022)			(0.015)	(0.019)		
Samesex x Asia-Africa	-	0.005	-	-	-	0.076	-	-	-	0.097	-	-
		(0.022)				(0.034)				(0.029)		
All girls	-	-	0.110	0.088	-	-	0.180	0.095	-	-	0.183	0.072
			(0.015)	(0.017)			(0.025)	(0.032)			(0.022)	(0.027)
All girls x Asia-Africa	-	-	-	0.051	-	-	-	0.169	-	-	-	0.217
				(0.028)				(0.051)				(0.043)
All boys	-	-	0.039	0.055	-	-	0.092	0.095	-	-	0.065	0.070
			(0.015)	(0.016)			(0.023)	(0.029)			(0.021)	(0.025)
All boys x Asia-Africa	-	-	-	-0.038	-	-	-	-0.005	-	-	-	-0.009
				(0.026)				(0.047)				(0.041)
Boy3 x (1-samesex12)	-	-	-	-	-	-	-0.080	-0.053	-	-	-0.077	-0.044
							(0.018)	(0.023)			(0.015)	(0.019)
Boy3 x (1-samesex12) x Asia-Africa	-	-	-	-	-	-	-	-0.054	-	-	-	-0.064
								(0.035)				(0.030)

Note.- The table reports first-stage effects of sex-composition and boy-3 on number of children. The sample for columns 1-4 includes first born non-twins from families with 2 or more births. The sample for columns 5-8 includes first born non-twins from families with 3 or more births. The sample for columns 9-12 includes first and second born non-twins from families with 3 or more births. Regression estimates are from models that include the control variables specified in the notes to Table 3. Standard errors in columns 9-12 are clustered by mother's ID.

Table 5
Estimates for First Borns in 2+ Sample

Outcome	Means (1)	OLS: basic covs. (2)	OLS: all covs. (3)	2SLS -- Instrument list: twins (4)	twins, twinsAA (5)	girl12, boy12 (6)	girl12, boy12, girl12AA, boy12AA (7)	all (8)
Schooling								
Highest grade completed	12.6	-0.252	-0.145	0.174	0.105	0.294	0.222	0.160
		(0.005)	(0.005)	(0.166)	(0.131)	(0.184)	(0.176)	(0.106)
Years of schooling ≥ 12	0.824	-0.037	-0.029	0.030	0.024	-0.009	-0.015	0.007
		(0.001)	(0.001)	(0.028)	(0.021)	(0.028)	(0.028)	(0.017)
Some College (age ≥ 24)	0.291	-0.049	-0.023	0.017	0.026	0.089	0.089	0.057
		(0.001)	(0.001)	(0.052)	(0.046)	(0.048)	(0.046)	(0.032)
College graduate (age ≥ 24)	0.202	-0.036	-0.015	-0.021	-0.006	0.115	0.115	0.054
		(0.001)	(0.001)	(0.045)	(0.041)	(0.046)	(0.044)	(0.028)
Labor Market Outcomes (age ≥ 22)								
Worked during the year	0.827	-0.025	-0.024	-0.005	0.002	0.062	0.072	0.034
		(0.001)	(0.001)	(0.038)	(0.033)	(0.043)	(0.043)	(0.026)
Hours worked last week	32.6	-1.06	-1.20	-0.97	0.00	1.46	1.06	0.51
		(0.05)	(0.06)	(2.58)	(2.18)	(2.06)	(1.98)	(1.45)
Monthly earnings (in 1995 Shekels)	2997	-217.0	-179.1	-7.7	73.0	266.7	429.1	264.1
		(7.4)	(8.0)	(394.1)	(324.5)	(283.6)	(292.1)	(214.2)
Ln(earnings) for full time workers	8.24	-0.045	-0.028	0.082	0.125	0.120	0.180	0.435
		(0.002)	(0.002)	(0.116)	(0.100)	(0.124)	(0.215)	(3.852)
Marriage and fertility								
Married on census day	0.446	0.023	0.020	0.043	0.060	0.118	0.101	0.078
		(0.001)	(0.001)	(0.029)	(0.025)	(0.034)	(0.032)	(0.020)
Married by age 21 (age ≥ 21)	0.172	0.027	0.022	-0.006	0.024	0.197	0.192	0.110
		(0.001)	(0.001)	(0.037)	(0.032)	(0.047)	(0.046)	(0.026)
Any children	0.448	0.029	0.019	0.090	0.013	0.135	0.134	0.079
		(0.001)	(0.001)	(0.056)	(0.036)	(0.041)	(0.041)	(0.026)

Note.- The table reports means of dependent variables in column 1 and OLS estimates of the coefficient on family size in columns 2-3. 2SLS estimates using different sets of instruments appear in columns 4-8. Instruments with an ‘aa’ suffix are interaction terms with an AA dummy. The sample includes first borns from families with 2 or more births as described in Table 1. OLS estimates for column 2 include indicators for age and sex. Estimates for columns 3-8 are from models that include the control variables specified in Table 3. Robust standard errors are reported in parenthesis.

Table 6

Estimates for First and Second Borns in 3+ Sample

		2SLS -- Instrument list						
Outcome	OLS all covs. (1)	twins (2)	twins, twinsAA (3)	girl123, boy123 (4)	girl123, boy123, girl123AA, boy123AA (5)	all (6)	all, boy3, boy3AA (7)	twins, twinsAA, boy3, boy3AA (8)
Schooling								
Highest grade completed	-0.143	0.167	0.187	-0.116	-0.067	0.069	0.072	0.171
	(0.005)	(0.117)	(0.110)	(0.134)	(0.120)	(0.080)	(0.076)	(0.101)
Years of schooling ≥ 12	-0.031	0.024	0.025	0.000	-0.009	0.009	0.006	0.016
	(0.001)	(0.019)	(0.018)	(0.023)	(0.022)	(0.014)	(0.013)	(0.017)
Some College (age ≥ 24)	-0.021	0.059	0.060	-0.051	-0.025	0.011	0.003	0.029
	(0.001)	(0.036)	(0.036)	(0.031)	(0.025)	(0.021)	(0.018)	(0.027)
College graduate (age ≥ 24)	-0.014	0.052	0.055	-0.060	-0.032	0.004	0.001	0.030
	(0.001)	(0.032)	(0.032)	(0.028)	(0.022)	(0.018)	(0.016)	(0.024)
Labor Market Outcomes (age ≥ 22)								
Worked during the year	-0.027	0.029	0.034	0.033	0.031	0.032	0.035	0.038
	(0.001)	(0.025)	(0.024)	(0.029)	(0.027)	(0.018)	(0.017)	(0.021)
Hours worked last week	-1.40	2.35	2.51	1.31	1.44	1.94	1.79	2.08
	(0.05)	(1.45)	(1.43)	(1.36)	(1.28)	(0.94)	(0.88)	(1.23)
Monthly earnings (in 1995 Shekels)	-184.5	47.2	63.9	118.8	109.2	90.0	93.1	76.3
	(6.8)	(204.1)	(203.8)	(176.5)	(162.2)	(128.7)	(120.8)	(175.7)
Ln(earnings) for full time workers	-0.030	-0.010	-0.016	0.021	0.053	0.014	-0.002	-0.021
	(0.002)	(0.047)	(0.047)	(0.069)	(0.061)	(0.039)	(0.036)	(0.071)
Marriage and fertility								
Married on census day	0.020	0.021	0.015	0.024	0.045	0.029	0.022	0.006
	(0.001)	(0.018)	(0.018)	(0.024)	(0.022)	(0.014)	(0.013)	(0.016)
Married by age 21 (age ≥ 21)	0.023	0.029	0.028	0.046	0.052	0.039	0.036	0.024
	(0.001)	(0.022)	(0.021)	(0.029)	(0.028)	(0.017)	(0.016)	(0.019)
Any children	0.022	0.021	0.015	-0.002	0.016	0.015	0.013	0.010
	(0.001)	(0.028)	(0.026)	(0.025)	(0.023)	(0.017)	(0.016)	(0.023)

Note.- The table reports OLS estimates of the coefficient on family size in column 1. 2SLS estimates using different sets of instruments appear in columns 2-8. Instruments with an 'aa' suffix are interaction terms with an AA dummy. The sample includes first and second borns from families with 3 or more births as described in Table 2. Regression estimates are from models that include the control variables specified in Table 3. Standard errors are clustered by mother's ID.

Table 7
Estimates for Parity-Pooled Samples

	Twins			Sex composition			All		
	2+, 3+ (1)	2+, 3+, 4+ (2)	2+, 3+, 4+, 5+ (3)	2+, 3+ (4)	2+, 3+, 4+ (5)	2+, 3+, 4+, 5+ (6)	2+, 3+ (7)	2+, 3+, 4+ (8)	2+, 3+, 4+, 5+ (9)
Schooling									
Highest grade completed	0.141	0.060	0.031	0.036	0.036	0.054	0.092	0.050	0.040
	(0.079)	(0.059)	(0.055)	(0.093)	(0.074)	(0.068)	(0.060)	(0.046)	(0.043)
Years of schooling ≥ 12	0.022	0.002	-0.006	-0.013	-0.015	-0.010	0.006	-0.005	-0.008
	(0.013)	(0.011)	(0.010)	(0.016)	(0.014)	(0.013)	(0.010)	(0.008)	(0.008)
Some College (age ≥ 24)	0.039	0.010	0.004	0.002	0.009	0.002	0.017	0.010	0.003
	(0.026)	(0.016)	(0.014)	(0.019)	(0.015)	(0.014)	(0.016)	(0.011)	(0.010)
College graduate (age ≥ 24)	0.024	0.012	0.011	0.007	0.011	0.007	0.014	0.012	0.010
	(0.023)	(0.014)	(0.012)	(0.017)	(0.013)	(0.012)	(0.014)	(0.010)	(0.009)
Labor Market Outcomes (age ≥ 22)									
Worked during the year	0.016	-0.007	-0.001	0.044	0.004	0.011	0.031	-0.002	0.005
	(0.018)	(0.014)	(0.013)	(0.020)	(0.016)	(0.014)	(0.013)	(0.010)	(0.009)
Hours worked last week	1.31	0.37	0.66	1.25	-0.41	0.11	1.27	0.00	0.40
	(1.09)	(0.76)	(0.69)	(0.98)	(0.77)	(0.71)	(0.72)	(0.54)	(0.49)
Monthly earnings (in 1995 Shekels)	31.8	-37.4	36.6	187.5	48.8	57.2	118.7	5.3	47.0
	(158.2)	(121.8)	(107.1)	(132.5)	(97.5)	(88.8)	(101.6)	(78.6)	(70.2)
Ln(earnings) for full time workers	-0.022	-0.007	-0.003	0.024	0.012	0.014	0.015	0.004	0.006
	(0.230)	(0.032)	(0.029)	(0.057)	(0.033)	(0.031)	(0.063)	(0.023)	(0.021)
Marriage and fertility									
Married on census day	0.030	0.029	0.023	0.046	0.053	0.045	0.037	0.039	0.033
	(0.014)	(0.010)	(0.010)	(0.017)	(0.013)	(0.012)	(0.010)	(0.008)	(0.007)
Married by age 21 (age ≥ 21)	0.024	0.023	0.020	0.078	0.058	0.052	0.052	0.039	0.035
	(0.016)	(0.012)	(0.012)	(0.021)	(0.017)	(0.015)	(0.013)	(0.010)	(0.009)
Any children	0.019	0.017	0.021	0.040	0.049	0.042	0.031	0.033	0.032
	(0.020)	(0.014)	(0.013)	(0.018)	(0.015)	(0.013)	(0.013)	(0.010)	(0.009)

Note.- The table reports 2SLS estimates of parity-pooled models using different sets of instruments and sub-samples. For example, the columns headed "2+, 3+" report results from the union of 2+ and 3+ samples. Standard errors are clustered by mother's ID.

Table 8
Full Specification By Ethnicity and Sex

	By Ethnicity				By Sex			
	Israel/Europe		Asia/Africa		Males		Females	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)	OLS (7)	2SLS (8)
Schooling								
Highest grade completed	-0.117	0.043	-0.117	0.031	-0.150	0.078	-0.096	0.020
	(0.006)	(0.064)	(0.005)	(0.057)	(0.005)	(0.073)	(0.005)	(0.050)
Years of schooling ≥ 12	-0.029	0.000	-0.023	-0.014	-0.033	0.002	-0.019	-0.011
	(0.001)	(0.011)	(0.001)	(0.011)	(0.001)	(0.014)	(0.001)	(0.009)
Some College (age ≥ 24)	-0.033	0.010	-0.014	-0.001	-0.017	0.010	-0.019	0.000
	(0.001)	(0.019)	(0.001)	(0.011)	(0.001)	(0.016)	(0.001)	(0.012)
College graduate (age ≥ 24)	-0.021	0.019	-0.010	0.003	-0.011	0.021	-0.013	0.006
	(0.001)	(0.017)	(0.001)	(0.009)	(0.001)	(0.014)	(0.001)	(0.011)
Labor Market Outcomes (age ≥ 22)								
Worked during the year	-0.039	-0.007	-0.014	0.011	-0.024	0.013	-0.021	0.000
	(0.001)	(0.016)	(0.001)	(0.012)	(0.001)	(0.014)	(0.001)	(0.012)
Hours worked last week	-2.23	0.30	-0.67	0.45	-1.31	0.56	-1.01	0.23
	(0.07)	(0.87)	(0.05)	(0.59)	(0.06)	(0.91)	(0.05)	(0.55)
Monthly earnings (in 1995 Shekels)	-239.1	119.6	-120.5	7.0	-219.6	65.5	-109.1	33.2
	(8.6)	(147.0)	(6.5)	(74.2)	(8.7)	(162.1)	(5.4)	(59.6)
Ln(earnings) for full time workers	-0.030	0.072	-0.025	-0.032	-0.027	0.009	-0.026	0.015
	(0.004)	(0.037)	(0.002)	(0.025)	(0.002)	(0.031)	(0.003)	(0.026)
Marriage and fertility								
Married on census day	0.034	0.030	0.010	0.035	0.016	0.021	0.022	0.040
	(0.001)	(0.011)	(0.001)	(0.010)	(0.001)	(0.011)	(0.001)	(0.009)
Married by age 21 (age ≥ 21)	0.039	0.034	0.014	0.038	0.013	-0.012	0.027	0.065
	(0.001)	(0.014)	(0.001)	(0.012)	(0.001)	(0.010)	(0.001)	(0.013)
Any children	0.034	0.024	0.015	0.038	-	-	0.020	0.032
	(0.002)	(0.016)	(0.001)	(0.012)			(0.001)	(0.009)

Note.- The table reports OLS and 2SLS results from models estimated separately by ethnicity and by sex. The 2SLS estimates are from models estimated for the full parity-pooled samples using the full set of instruments (i.e. corresponding to column 9 in table 7). Standard errors are clustered by mother's ID.

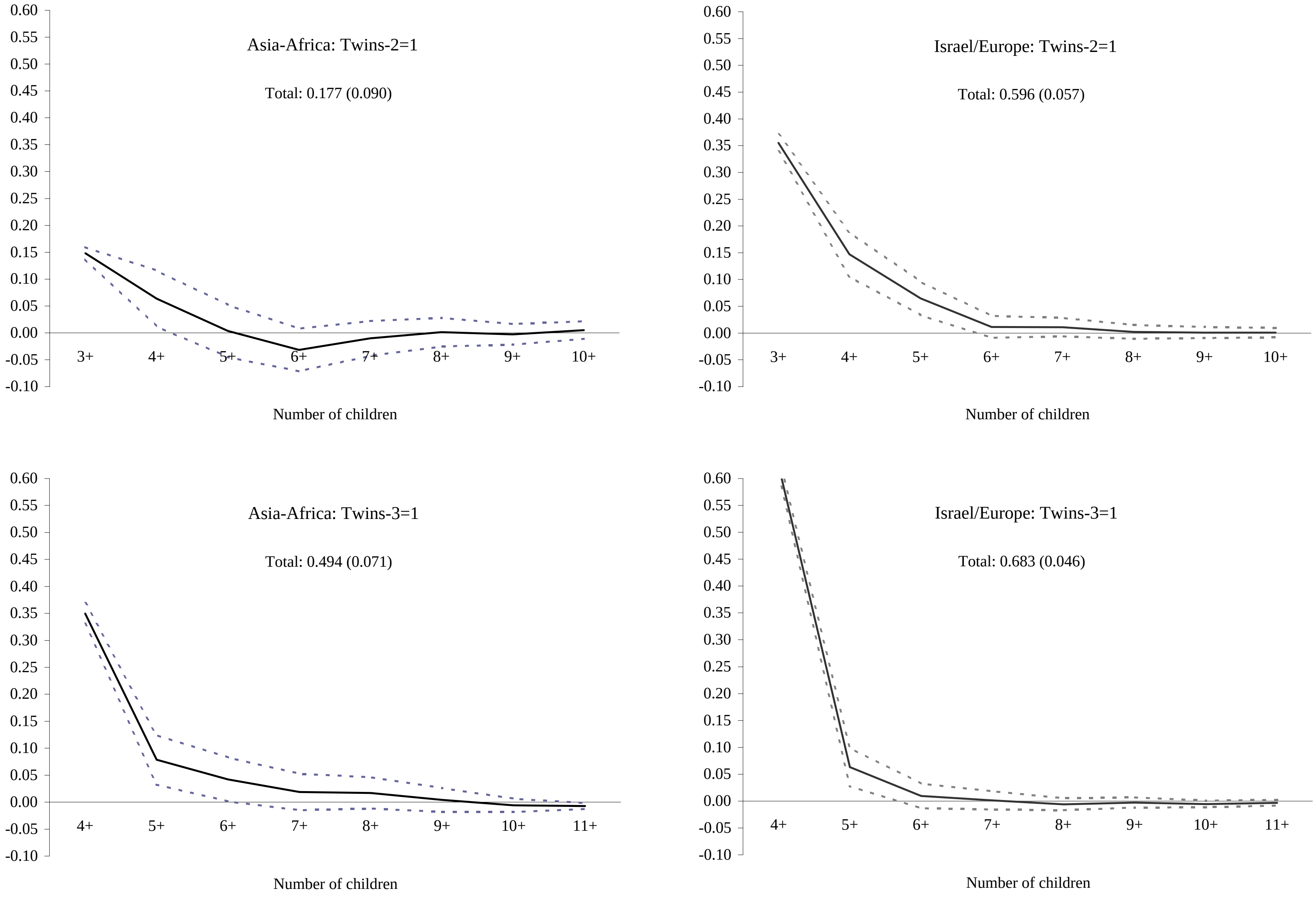


Figure 1: First borns in the 2+ sample, first stage effects of twins-2 (top panel). First and second borns in the 3+ sample, first stage effects of twins-3 (bottom panel).

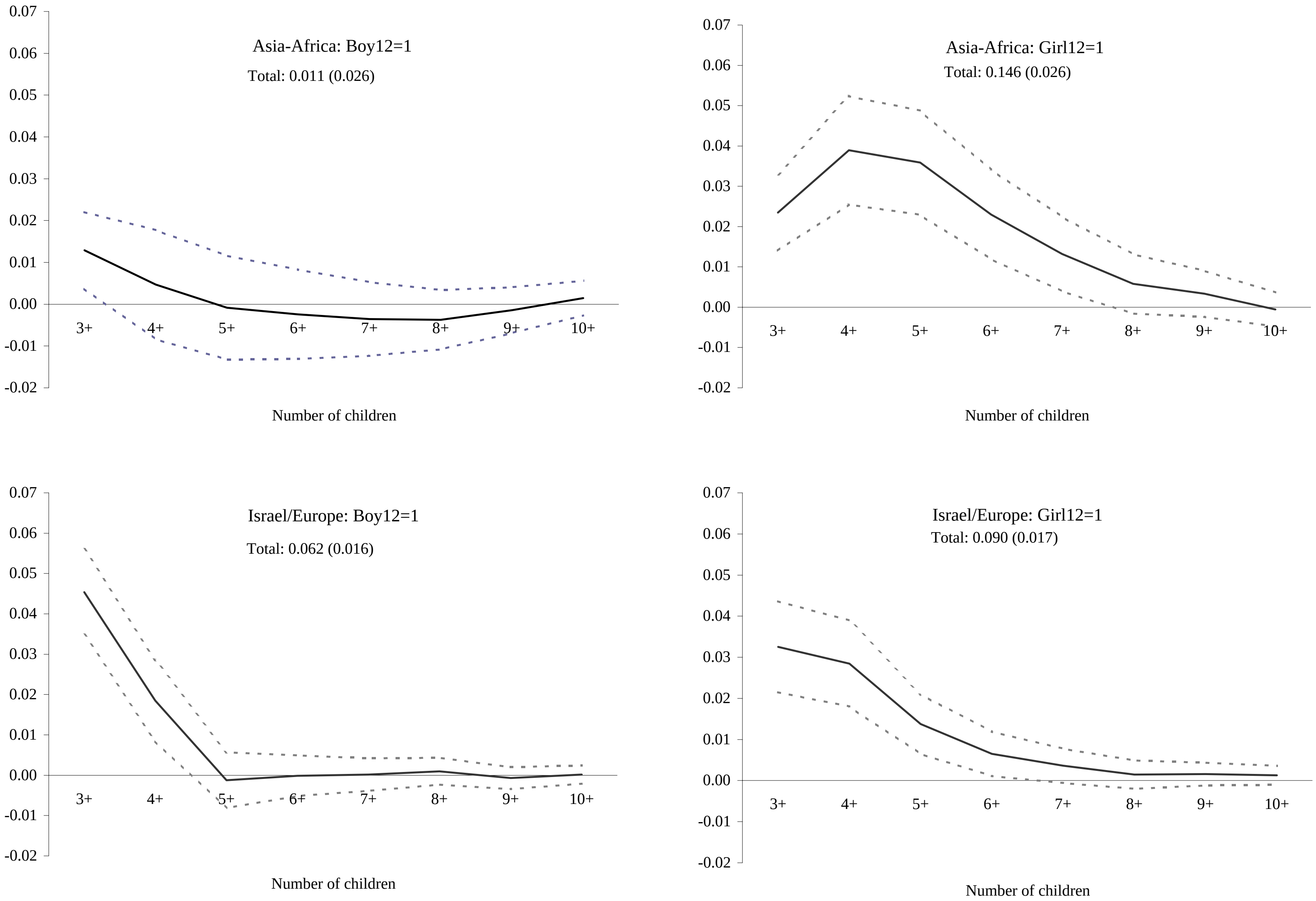


Figure 2: First-borns 2+ sample. First stage effects by ethnicity and type of sex-mix.

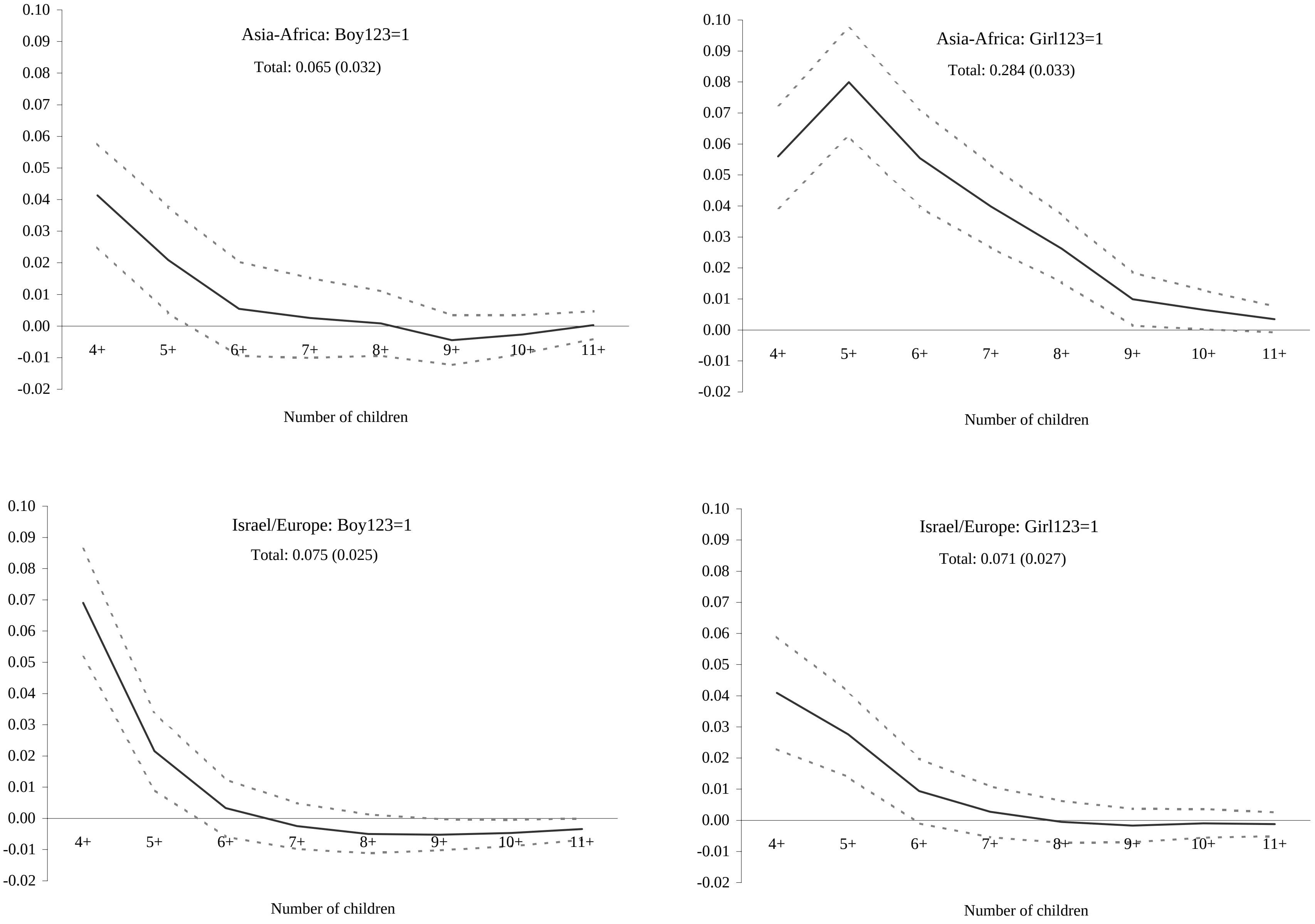
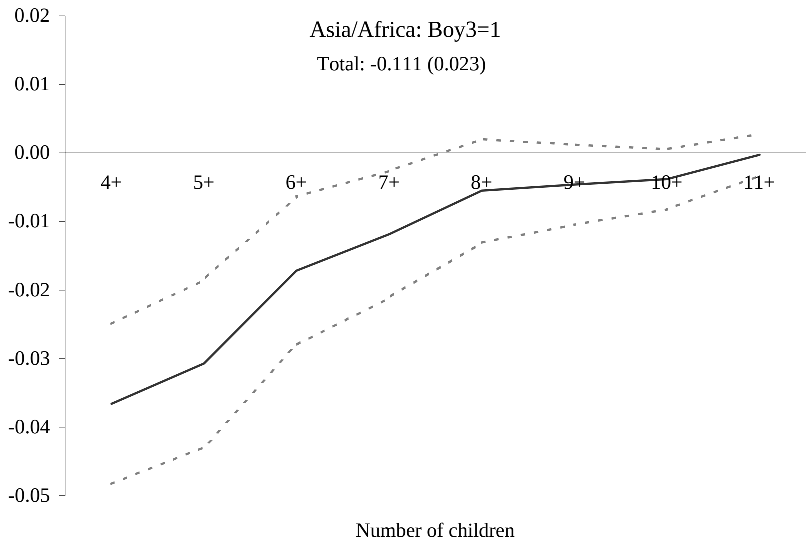


Figure 3: First and second borns 3+ sample. First stage effects by ethnicity and type of sex-mix.



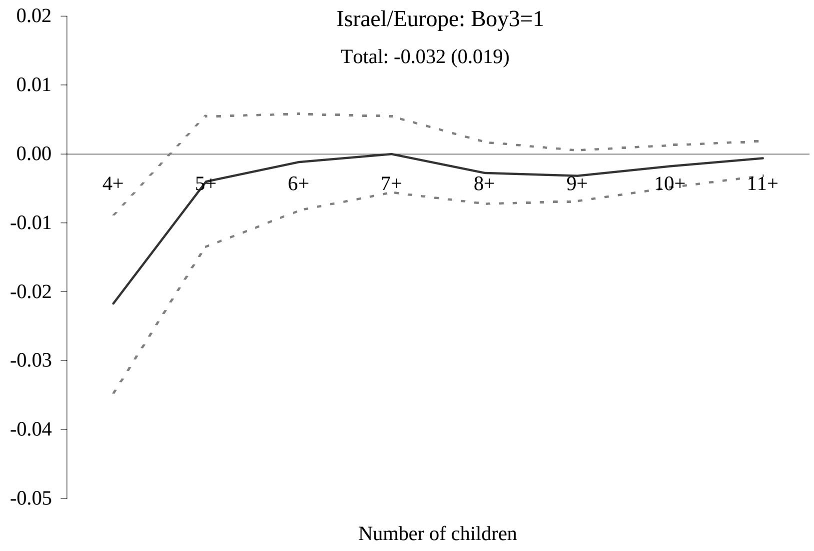


Figure 4: First and second borns 3+ sample. First stage effects of Boy3.

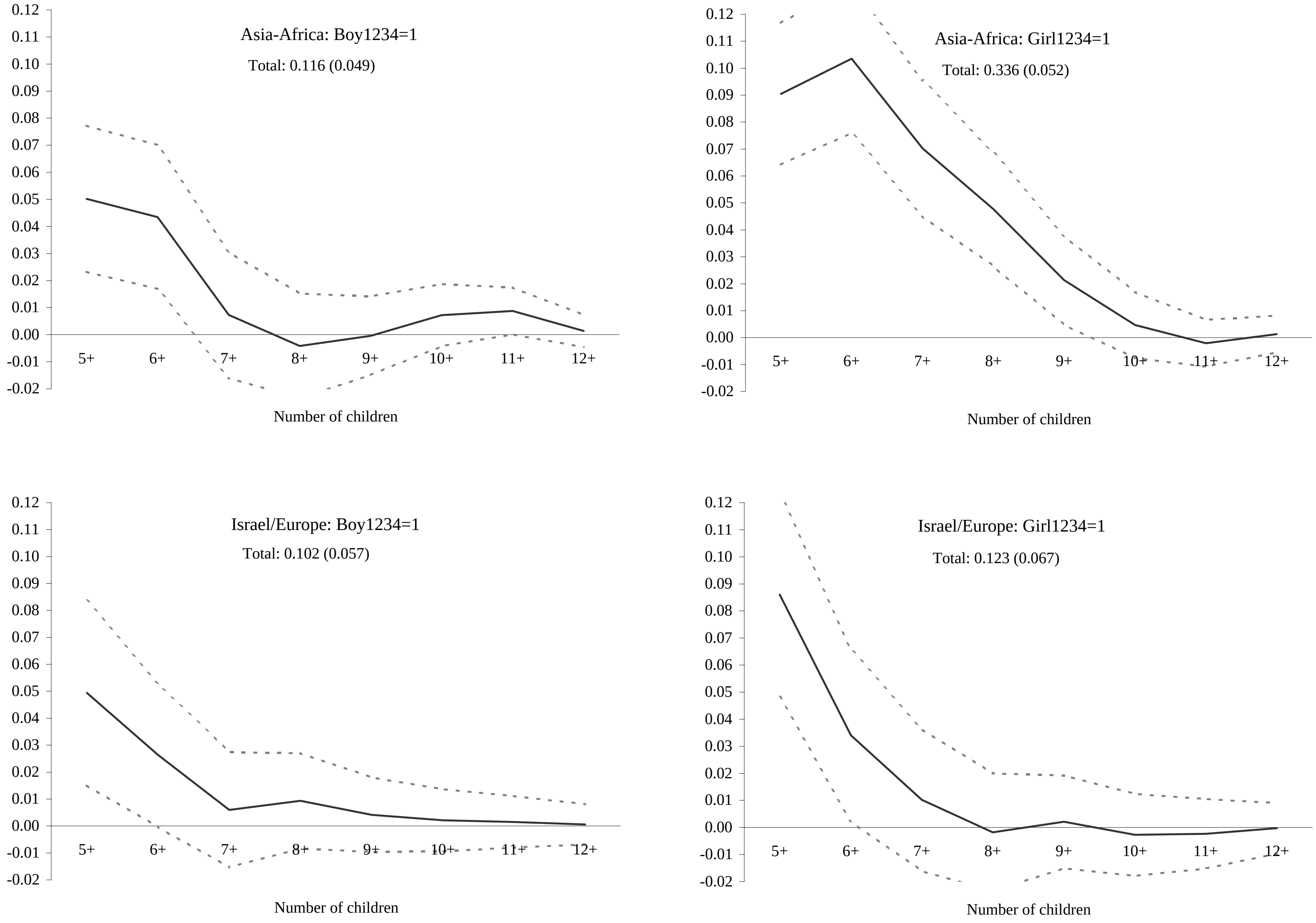


Figure 5: First, second, and third borns 4+ sample. First stage effects by ethnicity and type of sex-mix.

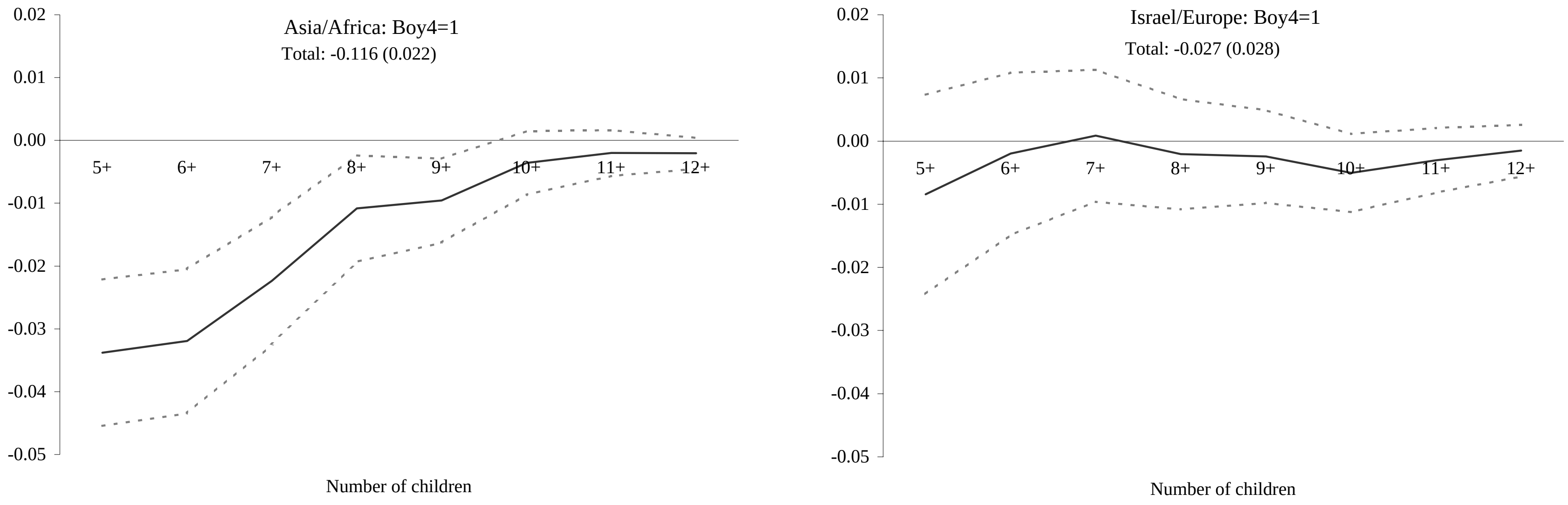


Figure 6: First, second, and third borns 4+ sample. First stage effects of Boy4.

Table A1

Relationship Between Family Structure and Selection into Census Data

	2+				3+							
	1st borns				1st borns				2nd borns			
	census83		census95		census83		census95		census83		census95	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of children	-0.0008	-0.0025	0.0008	0.0005	-0.0012	-0.0032	-0.0004	-0.0006	-0.0015	-0.0034	0.0002	0.0005
	(0.0005)	(0.0006)	(0.0003)	(0.0004)	(0.0006)	(0.0006)	(0.0004)	(0.0004)	(0.0006)	(0.0007)	(0.0004)	(0.0005)
Subject = boy	0.0003	-0.0002	-0.0034	-0.0034	0.0004	-0.0001	-0.0032	-0.0031	-0.0001	-0.0002	-0.0052	-0.0052
	(0.0020)	(0.0020)	(0.0012)	(0.0012)	(0.0023)	(0.0022)	(0.0014)	(0.0014)	(0.0026)	(0.0026)	(0.0016)	(0.0016)
Twins	0.0108	0.0110	-0.0037	-0.0037	-0.0062	-0.0054	-0.0080	-0.0083	-0.0138	-0.0131	-0.0002	0.0001
	(0.0118)	(0.0117)	(0.0062)	(0.0062)	(0.0112)	(0.0112)	(0.0068)	(0.0068)	(0.0134)	(0.0133)	(0.0076)	(0.0076)
Samesex	-0.0013	-0.0014	-0.0007	-0.0008	-0.0011	-0.0010	0.0014	0.0014	0.0000	0.0003	-0.0003	-0.0002
	(0.0020)	(0.0019)	(0.0012)	(0.0012)	(0.0026)	(0.0026)	(0.0016)	(0.0016)	(0.0030)	(0.0030)	(0.0018)	(0.0018)
All boys	0.0009	0.0002	-0.0030	-0.0031	0.0016	0.0015	-0.0007	-0.0007	0.0009	0.0007	-0.0018	-0.0019
	(0.0022)	(0.0022)	(0.0014)	(0.0014)	(0.0032)	(0.0032)	(0.0020)	(0.0020)	(0.0037)	(0.0037)	(0.0022)	(0.0022)
All girls	-0.0028	-0.0023	0.0022	0.0022	-0.0039	-0.0037	0.0034	0.0034	-0.0011	-0.0002	0.0016	0.0018
	(0.0023)	(0.0023)	(0.0014)	(0.0014)	(0.0035)	(0.0035)	(0.0022)	(0.0022)	(0.0041)	(0.0041)	(0.0024)	(0.0024)
Boy at last birth	-	-	-	-	0.0030	0.0030	-0.0015	-0.0016	0.0024	0.0019	0.0010	0.0010
(for mixed sex in n-1 births)					(0.0026)	(0.0026)	(0.0016)	(0.0016)	(0.0030)	(0.0030)	(0.0018)	(0.0018)
Full controls	-	√	-	√	-	√	-	√	-	√	-	√
Number of observations	135,568		392,504		102,349		283,573		75,272		235,421	

Note.- The table reports the effects of family size and the instruments on the probability of appearing in the census data. Each cell reports estimates of a separate regression. The samples include all Jewish individuals registered at the population registry who were alive at the census date. Estimates reported in even columns come from regressions that control for indicators for age at census date, missing month of birth, mother's age at census date, mother's age at first birth, mother's age at immigration (where relevant), and father's and mother's place of birth.

Table A2
Reduced Form Effects in No-First-Stage Samples

	Twins-2			Boy-12		
	Mother's age at 1st birth < 21	Spacing between 1st and 2nd birth<2	Asia/Africa Ethnicity	Mother's age at 1st birth < 21	Spacing between 1st and 2nd birth<2	Asia/Africa Ethnicity
	(1)	(2)	(3)	(4)	(5)	(6)
			A. Effect on Family Size			
Number of Children	0.067	0.132	0.177	0.017	-0.033	0.012
	(0.126)	(0.156)	(0.090)	(0.033)	(0.047)	(0.026)
			B. Effect on Outcomes of First Borns			
Schooling						
Highest grade completed	0.094	-0.162	0.091	0.049	0.027	0.023
	(0.136)	(0.147)	(0.131)	(0.041)	(0.053)	(0.033)
Years of schooling ≥ 12	0.023	-0.012	0.009	0.008	0.008	0.007
	(0.031)	(0.038)	(0.022)	(0.008)	(0.010)	(0.006)
Some College (age ≥ 24)	0.042	-0.092	0.000	0.011	0.000	-0.006
	(0.045)	(0.045)	(0.028)	(0.008)	(0.012)	(0.007)
College graduate (age ≥ 24)	-0.029	-0.084	-0.015	0.003	0.002	-0.003
	(0.032)	(0.031)	(0.023)	(0.007)	(0.010)	(0.005)
Labor Market Outcomes (age ≥ 22)						
Worked during the year	0.011	-0.004	-0.011	-0.002	-0.003	-0.005
	(0.035)	(0.049)	(0.026)	(0.007)	(0.010)	(0.005)
Hours worked last week	0.370	-0.451	-1.055	0.059	0.119	0.200
	(2.013)	(2.973)	(1.418)	(0.456)	(0.652)	(0.353)
Monthly earnings (in 1995 Shekels)	-420.5	-283.1	-106.4	-9.0	-61.0	-74.6
	(341.7)	(284.8)	(213.8)	(83.2)	(95.3)	(62.5)
Ln(earnings) for full time workers	-0.020	-0.059	-0.028	0.003	-0.007	-0.012
	(0.098)	(0.115)	(0.060)	(0.016)	(0.022)	(0.012)
Marriage and fertility						
Married on census day	0.025	0.020	-0.018	0.006	0.009	0.008
	(0.027)	(0.035)	(0.021)	(0.006)	(0.008)	(0.005)
Married by age 21 (age ≥ 21)	-0.034	-0.061	-0.034	0.001	0.010	0.000
	(0.034)	(0.041)	(0.022)	(0.006)	(0.008)	(0.004)
Any children	0.008	0.027	0.049	---	---	---
	(0.038)	(0.059)	(0.030)			

Note.- Panel A reports first stage effects of Twins-2 and Boy-12 on number of children for different subsamples. Panel B reports reduced form effects of Twins-2 and Boy-12 on the different outcomes. The samples include first borns from families with 2 or more births. Regression estimates are from models that include the control variables specified in Table 3. Robust standard errors are reported in parenthesis.

Table A3
Pooled First Stage

	2+	3+		4+		5+	
	1st borns	1st borns	1st+2nd	1st borns	1st+2nd+3rd	1st borns	1st+2nd +3rd+4th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Twins	0.637	0.594	0.698	0.879	0.998	0.629	0.714
	(0.057)	(0.045)	(0.049)	(0.114)	(0.110)	(0.232)	(0.162)
Twins x Asia-Africa	-0.482	-0.138	-0.228	-0.135	-0.291	-0.052	-0.159
	(0.105)	(0.094)	(0.086)	(0.148)	(0.137)	(0.268)	(0.196)
All Girls	0.090	0.102	0.079	0.158	0.127	0.147	0.242
	(0.017)	(0.032)	(0.027)	(0.083)	(0.068)	(0.212)	(0.161)
All Girls x Asia-Africa	0.052	0.165	0.212	0.202	0.212	0.173	0.123
	(0.032)	(0.051)	(0.043)	(0.105)	(0.085)	(0.245)	(0.180)
All Boys	0.061	0.103	0.077	0.096	0.102	0.489	0.307
	(0.017)	(0.029)	(0.025)	(0.073)	(0.058)	(0.218)	(0.155)
All Boys x Asia-Africa	-0.045	-0.011	-0.015	0.067	0.007	-0.535	-0.336
	(0.031)	(0.047)	(0.041)	(0.095)	(0.076)	(0.247)	(0.176)
Boy at last birth	-	-0.048	-0.038	-0.024	-0.024	0.039	0.031
(for mixed sex in n-1 births)		(0.023)	(0.019)	(0.036)	(0.029)	(0.067)	(0.048)
Boy at last birth x Asia-Africa	-	-0.057	-0.067	-0.085	-0.087	-0.112	-0.126
(for mixed sex in n-1 births)		(0.035)	(0.030)	(0.045)	(0.036)	(0.075)	(0.054)
Subject = boy	0.013	0.018	0.006	0.021	0.025	-0.023	0.002
	(0.017)	(0.023)	(0.015)	(0.045)	(0.024)	(0.090)	(0.039)
(Subject = boy) x Asia-Africa	0.007	-0.005	0.016	0.005	0.010	0.054	0.018
	(0.031)	(0.035)	(0.022)	(0.055)	(0.028)	(0.100)	(0.043)
Asia-Africa	0.242	0.189	0.103	-0.001	-0.054	-0.276	-0.253
	(0.024)	(0.032)	(0.026)	(0.056)	(0.041)	(0.111)	(0.072)

Note.- The table reports first-stage effects on number of children using the full set of instruments. The regression estimates are from models that include the control variables specified in Table 3. Regressions for columns 4-7 also include controls for twins and sex-composition at lower parity births. Regressions for columns 3,5, and 7 also include controls for birth order and spacing. Robust standard errors are reported in parenthesis. Standard errors in columns 3, 5, and 7 are clustered by mother's ID.