Investigating the Structure of Son Bias in Armenia with Novel Measures of Individual Preferences

Matthias Schief[1], Sonja Vogt[2,3], and Charles Efferson[4]

[1]Department of Economics, Brown University, U.S.A.
[2]Department of Social Sciences, University of Bern, Switzerland
[3]Centre for Development and Environment, University of Bern, Switzerland
[4]Faculty of Business and Economics, University of Lausanne, Switzerland

Last updated: March, 2021

A Appendix

A.1 Sampling Protocol and Data Collection

Target population The study focuses on three areas: Yerevan, Gegharkunik, and Syunik. Yerevan is the capital region, while Gegharkunik and Syunik are the regions with the highest and lowest sex ratio among children below the age of 16 in the 2011 national census. Within each region, we randomly sampled married couples with at least one child under 16 living at home according to the two-stage sampling approach described below. For a sampled couple, we collected data with both the husband, the wife, and the husband's mother during the same visit. If a couple was sampled, but either the wife or the husband had deceased or was living in a distant region or abroad for the duration of the study, we still collected data with the other spouse (and the husband's mother if available). In terms of sampling, we call such couples "incomplete couples." The rationale behind collecting data from incomplete coupes was to avoid selection bias that may have resulted if complete couples and incomplete couples were structurally different in terms of any variable of interest. If a couple was sampled, and both spouses were available at some time during the study, participation of both souses in the data collection was a pre-condition for participation of the given household in our study. We call such couples "complete couples." Finally, the husband's mother was included in the study whenever she was available on the day of data collection and she agreed to participate.

The study took place between May 2017 and March 2019. In each of the three regions, we planned to collect data from 900 participants leading to a total sample size of 2700 participants[1]. We sampled households based on a two-stage sampling approach that oversampled communities with especially high or low sex ratios at birth to maximize variation in the outcomes of interest, while at the same generating a subsample that is representative of the target population in the three regions. The details of the sampling protocol are described below. Our final sample consists of 1,212 households, in which

[1]Due to a facilitator's mistake in entering the subject identification number we lost one observation, which leads to a final sample size of 2699 participants.

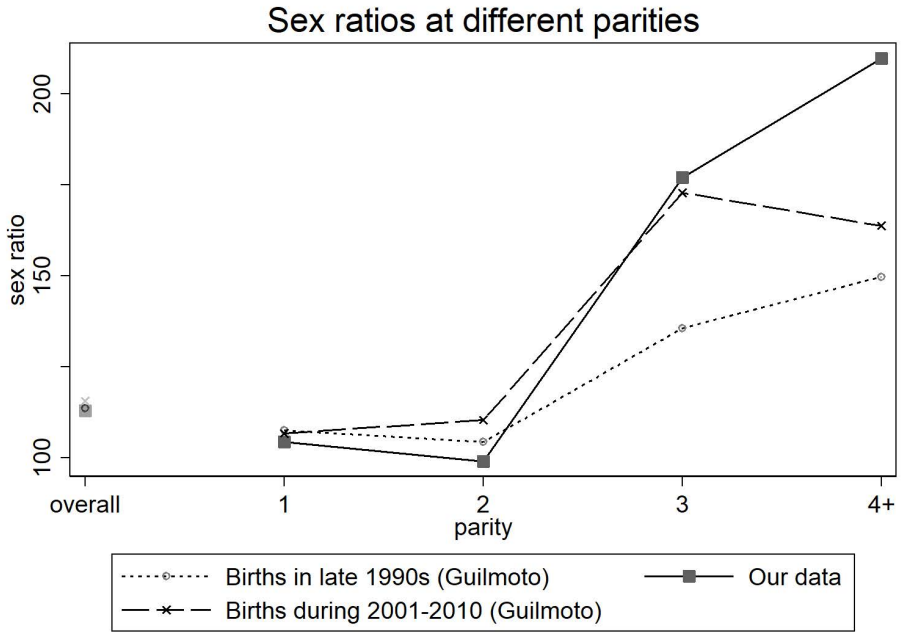

Figure A.1: Comparison of sex ratios in our data and sex ratios obtained from national birth records.

we collected data from 2.23 household members on average. Overall, 44% of our respondents are wives, 33% are husbands, and 23% are mothers-in-law. In 73% of all households we collected data from both the husband and the wife and in 34% we have data from all three eligible household members.

Representativeness While our data come from an interesting set of regions made up by the capital as well as two region at the opposite extreme ends of the regional variation in the sex ratio at birth, no claim is made about the representativeness of our sample for the country of Armenia. We explore the extent to which households in our sample are nationally representative in terms of their propensity to engage in sex selection by comparing sex ratios at various parities in our data to national figures. In particular, we focus on one section in our questionnaire in which wives anonymously provided their complete reproductive histories and from this we calculate the sex ratios at birth that we then compare to those reported from a nationally representative sample in Guilmoto (2013). Figure A.1 shows the results. Guilmoto (2013) reports sex ratios for births in the late 1990s and for births during the years 2001-2010. The latter category is closest to our sampling frame. The comparison provides a very close match for parities up to three and suggests that in terms of the prevalence of sex-selective abortions our sample may indeed be representative of Armenia at large.

Two-stage Sampling We implemented a two-stage sampling approach based on complete records of the sex of all children born in the three regions of interest. The data was provided by the *Statistical Committee of the Republic of Armenia* and is taken from the *2011 National Census*. In the first stage of our sampling strategy, we selected communities from within each region. In the second stage, we then sampled addresses from within a given sampled community. In each region, we created three distinct samples. First, we created a sample that is representative for our target population at the

regional level. We will call this sample the *representative sample* and two third of our participants are part of this sample. Next, we specifically sampled addresses from "extreme" communities that are characterized by unusually high or unusually low observed sex ratios at birth. We will call these samples *lower tail sample* and *upper tail sample* respectively to indicate that for these sample we have specifically chosen communities from the tails of the distribution of community-level sex ratios at birth. Those samples comprise one sixth of our participants each.

Determining the Sample Size We committed to collecting data from 900 participants per region. Assuming that we would on average be able to collect data with 2.25 participants per household, we concluded that we would need to work in approximately 400 households in a given region. Moreover, we assumed that only every third address that we sampled would eventually lead to a successful recruitment of the household living at that address. Finally, we wanted to avoid creating too high a concentration of sampled household in any given location and we therefore restricted ourselves to sample only 1 out of every 5 addresses that met our sampling requirements. Given these restrictions, we concluded that our *representative sample* should consist of 800 households while the *lower tail sample* and *upper tail sample* should consist of 200 households each.

Constructing the Representative Sample We constructed the *representative sample* by consecutively picking communities from within a given Marz and summing up the number of eligible households in the sampled communities. Communities were picked at random and we stopped picking additional communities once we had reached the target of 800 sampled households. In the second stage we then randomly sampled 20% of the households in those communities.

Constructing the Upper Tail Sample The task is to identify the communities whose inhabitants are most prone to selectively aborting female fetuses. However, we do not observe sex-selective abortions directly and must rely instead on data about sex ratios at birth, which are imperfect proxies for the prevalence of sex-selective abortions. Moreover, we did not want to simply select the communities with the highest sex ratios, because for many communities we observe only relatively few births and distorted sex ratios may then be the outcome of chance rather than a true reflection of widespread son bias. Hence, our aim was to identify the communities whose sex ratios at birth suggest the highest prevalence of sex selection provided that there is adequate evidence that the observed sex ratios truly reflect sex-selective abortions.

We think of the observed sex ratio in a given community as one realization of a data generating process. The data generating process specifies the probability that a child is born as a boy and therefore captures the underlying prevalence of sex selection. Let p denote this probability, which is assumed to vary across communities. With this interpretation at hand, we can formulate a null hypothesis specifying the probability that a child is born as a boy in a given community and we can perform a one sided binomial test to decide whether the null hypothesis can be rejected in favor of the alternative hypothesis that assumes the data generating process assigns a higher probability to the birth of a boy.

To select communities for the *upper tail sample*, we first excluded all the communities that had already been sampled as part of the *representative sample* and the searched for the communities for which we could reject the most 'extreme' null hypothesis. More precisely, we set the required significance level to 10% (i.e. $\alpha = 0.1$) and looked for the highest probability p^{null} subject to the constraint that the communities for which we can reject the null in favor of the one sided alternative

hypothesis $p^{alternative} > p^{null}$ together contain at least 200 eligible households.

Constructing the Lower Tail Sample To construct the *lower tail sample* we employed a similar strategy. First, we assume that in the absence of sex-selective abortions, 105 boys are born for every 100 girls. Consequently our null hypothesis is ($p^{null} = \frac{105}{105+100} \approx 0.512$). As an alternative hypothesis, we assume that sex-selective abortions are practiced in a given community, and we should therefore expect $x > 105$ boys to be born for every 100 girls ($p^{alternative} = \frac{x}{x+100}$). The lower tail sample consist of all communities for which we cannot reject the null hypothesis in favor of a given alternative hypothesis provided that we have adequate statistical power to actually reject the null hypothesis if the alternative hypothesis was true. The task was then to find the lowest alternative hypothesis $p^{alternative}$ such that the number of eligible households in the selected communities is at least 200.

Logistics of data collection To actually implement the implicit association test and the computerized questionnaire, we hired and trained approximately 20 facilitators and ran the software packages Inquisit (www.millisecond.com) and Limesurvey on laptops. The facilitators were recruited by the national statistical agency of Armenia, and most facilitators had prior experience with face-to-face data collection. We spent several days training these facilitators, and we practiced and refined our methods by running several pilot sessions with a sample of households that are not in our main study. After development and pre-testing, facilitators were divided into teams of three, and each team member was assigned a computer. When data was collected in one of the households in our study, the team would set up the three computers in different corners of the apartment and use folding dividers made out of cardboard to create a kind of isolation booth around each computer and participant.

To begin with a participant, a facilitator would sit down with the participant and explain the abbreviated implicit association test. The two of them would go through the test together, and the facilitator would evaluate the participant's understanding. They repeated this exercise as necessary. When the participant was prepared to proceed to the full implicit association test, the facilitator would help the participant put on headphones and then start the test. After verifying that the participant could hear the audio recordings, the facilitator would step away immediately, so that the participant could complete the test in privacy.

Most participants were assigned a facilitator of the same sex. While we would have liked to randomize participant-facilitator combinations within a given household, cultural norms around gender implied that it would have been difficult in some households for male facilitators to interview wives and for female facilitators to interview husbands. Faced with this trade-off, we allowed facilitators to exercise their own judgment regarding the best assignment of participants to facilitators within a given household, and we made sure that all data collection teams included both male and female facilitators.

Quality control. The data collection was monitored by independent employees of the *Women's Resource Center of Armenia*, who made both announced and unannounced visits to the teams of facilitators in the field. At least once a week the data was transferred to us so that we could check it for consistency. Thanks to log files stored on the computers, we were able to confirm, among other things, that i) the husband and the wife in a given household were interviewed at the same time by different facilitators ii) participants were allowed to proceed to the main implicit association test only once they had reached a threshold level of proficiency in the trial categorization task, or iii) enough time was spent on each questionnaire item for the participants to understand the question and think

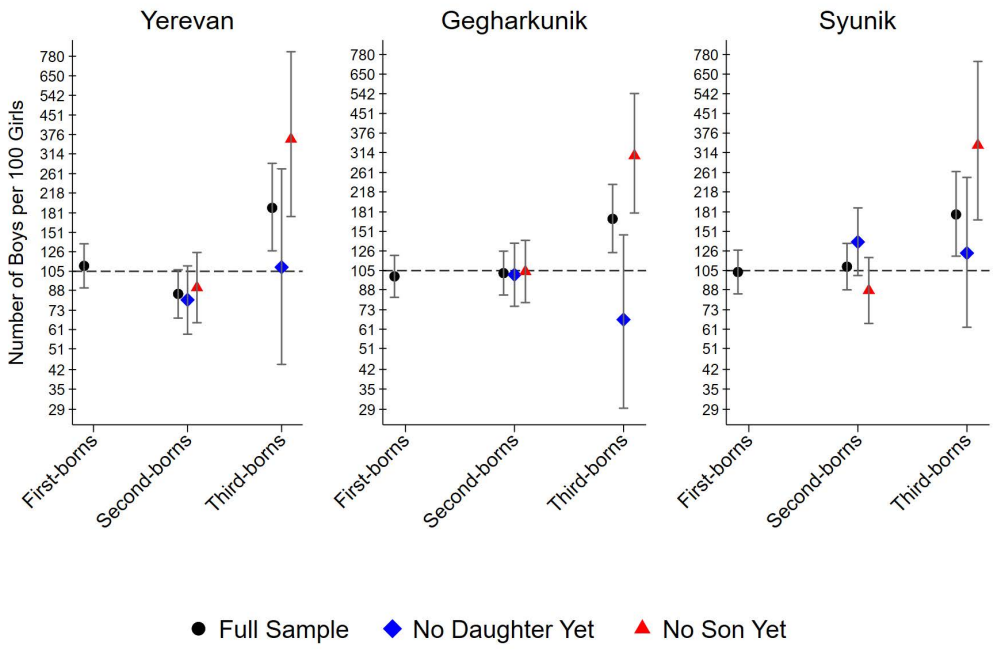

Figure A.2: Conditional sex ratios by region

about how to answer it.

A.2 Conditional sex ratios

In figure A.2, we replicate figure 1 from the main text separately for the regions of Yerevan, Gegharkunik, and Syunik. We find extremely similar patterns in all three regions.

In figure A.3, we further dis-aggregate the sex ratio among third-born children. In particular, we compute the sex ratio at third parity separately for all possible gender compositions among the first two children. We find that in families with a mixed gender composition the point estimates for the sex ratio among third-born children lie above the natural rate of 105 boys for every 100 girls, suggesting that at least for some parents son bias goes beyond wanting to have at least one son. The degree of distortion in the sex ratio is much less severe, however, compared to families that were still lacking a son and not statistically significant at standard significance levels.

A.3 Implicit Association Tests

Implicit association tests measure associations between target stimuli presented in neutral terms and valued stimuli. In our case, the target stimuli were drawings of families with either sons or daughters, while the valued stimuli were audio recordings of value-laden words.

Implicit association tests have been successfully used to study associations related to race, sexual orientation, religion, and other sensitive topics in contemporary societies (Nosek et al., 2007). Implicit association tests are much less prone to producing socially desirable responses than traditional survey methods. We implemented our implicit association test with all of our sampled participants. Our implicit association test followed the structure presented in Nosek et al. (2007). The test required

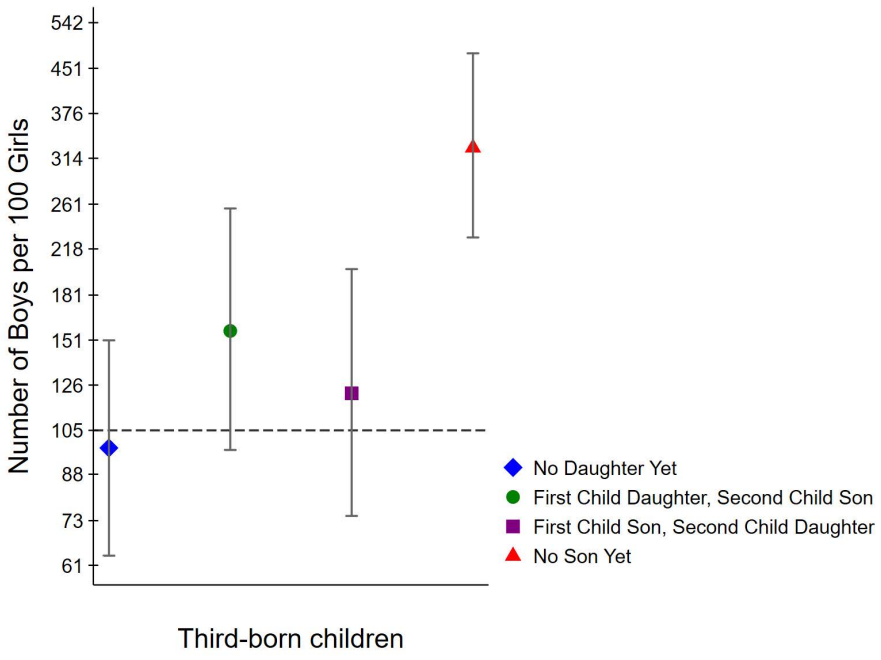

Figure A.3: The ratio of boys to girls among third born-children. 95% confidence intervals are computed using the Clopper-Pearson method for calculating binomial proportion confidence intervals.

participants to categorize all stimuli correctly. If a participant made a mistake, a large red "X" appeared on the screen, and the participant had to try again before continuing.

Stimuli. For our target stimuli, we presented drawings showing two different Armenian families engaged in daily activities like eating, watching television, or doing homework (see figures A.4 and A.5). In both families there are two children, but in one family the children are girls, and in the other family they are boys. A given drawing showed one of the two families engaged in one of the various activities.

The valued stimuli differed across the two versions of the implicit association test that we implemented. In the *Valence IAT*, participants were presented with positive and negative words. To positive category included the words "to cherish", "spectacular", "joyous", "excellent", and "glad". The negative category included the words "yucky", "horrible", "to fail", "harmful", and "sad." In the *Stereotype IAT*, participants were asked to sort words into two different explicitly labeled categories. The first category was labeled "to flourish" and included the words "offspring", "descendant", "to multiply", "standing", and "to immortalize". The second category was labeled "to wither" and included the words "childless", "infertility", "to interrupt", "to fade", and "extinction."

Introduction of the families and counterbalancing scheme. At the beginning of the implicit association test, the two families were shown together on the screen, and participants listened to a short audio recording over headphones that introduced the families. All family members were introduced by their names, and the participants learned about their respective ages. We chose ages for the parents and the children that made it look very unlikely that another child would be born into this family in

Figure A.4: Sample drawings for family with two sons

Figure A.5: Sample drawings for family with two daughters

the future, and the participants were told explicitly that the parents would not have any more children.

To ensure that our stimuli are focused on the gender of the children rather than any attributes of the parents, we used a counterbalancing scheme in which half of our participants took a version of the implicit association test in which the family with the two daughters (sons) was headed by the first (second) set of parents, while the other half took a version in which the family with the two daughters (sons) was headed by the second (first) set of parents. Furthermore, the left and right positions of the two families on the introductory screen were also counterbalanced across participants. For the case in which the family with the two girls is on the left, the introductory recording translates as follows:

> Look at the two families on the screen.
>
> Look at the family on the left. This is the Hovhannisyan family. The husband, Tigran, is 45, and the wife, Hasmik, is 43. They planned to have two children. They have two daughters, and they will not have any more children. The oldest daughter, Nare, is 18 years old, and the youngest daughter, Marie, is 15 years old.
>
> Look at the family on the right. This is the Gasparyan family. The husband, Hayk, is 45, and the wife, Gayane, is 43. They planned to have two children. They have two sons, and they will not have any more children. The oldest son, Davit, is 18 years old, and the youngest son, Narek, is 15 years old.
>
> In the coming task you will see several images of the two families. Look at the two families closely and try to remember them. This exercise will take about 15 minutes. Press the long key with the white sticker at the bottom of the keyboard when you are ready to continue.

Following Greenwald et al. (2003), our implicit association test consisted of seven basic blocks of trials. We counterbalanced the order of categorization rules across participants. Specifically, some participants initially faced a categorization scheme that paired negative words with the family with daughters, and they later faced a scheme that paired positive words with this family. Counterbalanced participants went in the opposite order. Altogether, our complete counterbalancing scheme produced multiple versions of the test. Counterbalancing allowed us to distribute the versions evenly and unsystematically across participants, and this eliminated the potential for any artifacts associated with the parents, the spatial locations of stimuli, or the ordering of categorization rules.

Scoring algorithm. D scores quantify a participant's relative response times under the two categorization schemes in the implicit association test. To the extent that relative response times vary systematically with some underlying attributes of the participants that are of interest to the researcher – in our case the degree of son bias in our participants' fertility preferences – D scores can be interpreted as a measure of implicit associations and preferences. Several extraneous factors have been shown to contaminate the measurement of implicit associations by exerting an independent influence on D scores. These nuisance factors include the relative order in which the categorization tasks are presented and individual differences in average response latency (**?**Nosek et al., 2005; Greenwald et al., 2003; **?**).

We addressed these challenges in several ways. First, we randomized across participants the order in which the target stimuli were presented. Second, we familiarized all participants with the general protocol of the implicit association test by running a trial version that involved unrelated stimuli. Participants were allowed to proceed to the main test only once they had reached a threshold level of

proficiency in the trial categorization task. Finally, we implemented the scoring algorithm developed in (Greenwald et al., 2003) that has been shown to minimize the influence of individual differences in average response latency.

This scoring algorithm produces D scores that are distributed on the interval [-2; 2] where a value of zero implies the absence of the implicit association in question. Positive scores in the Valence IAT imply an association of families with sons and words with positive valence. Similarly, positive scores in the Stereotype IAT imply an association of families with sons and words with connotations of flourishing.

Households were randomly assigned either the Valence IAT or the Stereotype IAT and all families members within a household completed the same version of the implicit association test. Because both versions produce similar results (cf. figure 3 in the main text), we combine the D scores from the Valence and Stereotype IAT for the main analyses in the paper. In all our analyses, we account for the versions of the IAT and the initial pairing of the target stimuli.

A.4 A Measure of Explicit Son-Biased Fertility Preferences

Husbands and wives in our sample are asked to envision the optimal future family of their youngest child, and we elicit preferences regarding the sex composition of this child's future children. Similarly, we elicit the preferences of the husband's mother regarding the same child's future children by adjusting the wording of the question to refer to this particular grandchild of hers. A first question asks for the optimal family size without giving the respondents the chance to specify the sex of the children. Then, at later points in the questionnaire, the respondents are asked to assume that the focal child would have exactly one, two, three, or four children respectively. In each of these four cases, the respondents are asked to specify the optimal number of boys and girls as well as the exact birth ordering while taking the overall number of children as given. To avoid anchoring effects, these four questions are each placed in different sections of the questionnaire, and the order in which the questions are presented is randomized across participants. Finally, in a last question, our facilitators present the respondents with four different scenarios regarding the future children of their youngest child and ask the respondents to pick their favorite scenario. These four scenarios describe families ranging in size from one to four children, and the questionnaire is programmed such that the gender distribution among the children corresponds exactly to the preferred gender composition of the participants as elicited in the earlier questions. Comparing the number of children in a participant's favorite scenario to the preferred number of children when the gender of the children could not be specified allows us to disentangle preferences over the number of children from preferences over the sex distribution.

Figure A.6 shows the relative frequency of reported optimal sex compositions for families of varying sizes. Two interesting features of the data stand out. First, there is a remarkable level of similarity in the relative frequency of preferred sex compositions across husbands, wives, and mothers-in-law. Second, the data suggest that there is not only a preference for having at least one son, but also a clear preference for a balanced sex composition of children. In fact, for the questions where the overall number of children is set to be two or four and the participants can therefore in principle achieve a balanced gender composition, more than 90% percent of all reported preferred sex compositions are completely balanced. However, despite the overall similarity in the distribution of preferred sex compositions across husbands, wives, and mothers-in-law, it is still possible to detect systematic gender differences. For example, the share of husbands that favor a family composition with more male than female children is higher than the share of women who favor such a family composition. Similarly, a

higher share of husbands than of wives want the first child to be a son or want all children to be boys.

To summarize the degree of son-bias in the stated preferences in a single statistic, we collapse the information contained in figure A.6 into a single index of son bias. More specifically, we implement a scoring algorithm that assigns one point whenever the respondent (a) wants the first child to be a boy, (b) wants more boys than girls, or (c) wants all children to be boys. Importantly however, the index does not double count in cases where two scoring criteria are equivalent. In the case of one child, we have that (a), (b), and (c) are all equivalent, and hence participants can score at most one point, while in the case of two children we have (b) and (c) are equivalent and therefore participants can score at most two points. Summing over the four questions with varying fixed family sizes, the resulting index can take any integer between zero and nine[2]. Figure A.7 shows a histogram of our index of son bias by region and subject type.

A.5 Internal consistency of D scores and Index values

In this section, we show that the correlation between D scores and index values depicted in figure 4 remains positive and significant after accounting for potential confounding factors.

As is shown section IV.1 in the main text, the largest differences in D scores occur between male and female participants, and between participants living in different regions. A possible concern is therefore that i) our measures of son bias are directly affected by a participant's gender even if latent son bias is held constant, or that ii) there are unobserved regional variables that do not reflect son bias but affect our measures of son bias. To the extent that this is the case, the correlation documented in figure 4 would be spurious. Figure A.8 shows binned scatter plots of the associations between D scores and index values by gender of the respondent and by region. We find that the positive association between D scores and index values is also obtained *within* regions and *within* participants of the same gender.

This result can be put to a more rigorous test. Table 1 reports estimation results from a regression of index values on D scores. Column 1 documents the raw association between D scores and index values and establishes that higher D scores indeed predict higher values of the index. In column 2 we control for subject type and find that the association between explicit and implicit measures of son bias is robust to exploiting only variation in son bias within wives, husbands, and mothers-in-law. One may be concerned that both the D scores and the index are affected by the gender composition of the children in a participant's family. For example, similarity between one's own family and the families depicted in the implicit association test may affect response times. This will affect D scores. Similarly, respondents may take into account the gender of their youngest child when deciding on the ideal gender composition in this child's future family. This will in turn affect the index. This can lead to spurious correlations between the implicit and the explicit measures of son bias across participants with different family compositions. Column 3 therefore includes a full set of family composition fixed effects and shows that this potential confounding factor is not driving the results. In column 4, we additionally account for 45 community fixed effects. The coefficient retains its significance at the 5% level.

Given that D scores as well as index values are measuring true underlying son bias with some amount of noise, the estimated coefficient will suffer attenuation bias and we should expect the coefficient to become smaller as we move from column (1) to column (4). Controlling for region, subject

[2]This index was preregistered by Efferson et al. (2018)

(a) Preferred sex of only child

(b) Preferred gender composition for family with two children

(c) Preferred gender composition for family with three children

(d) Preferred gender composition for family with four children

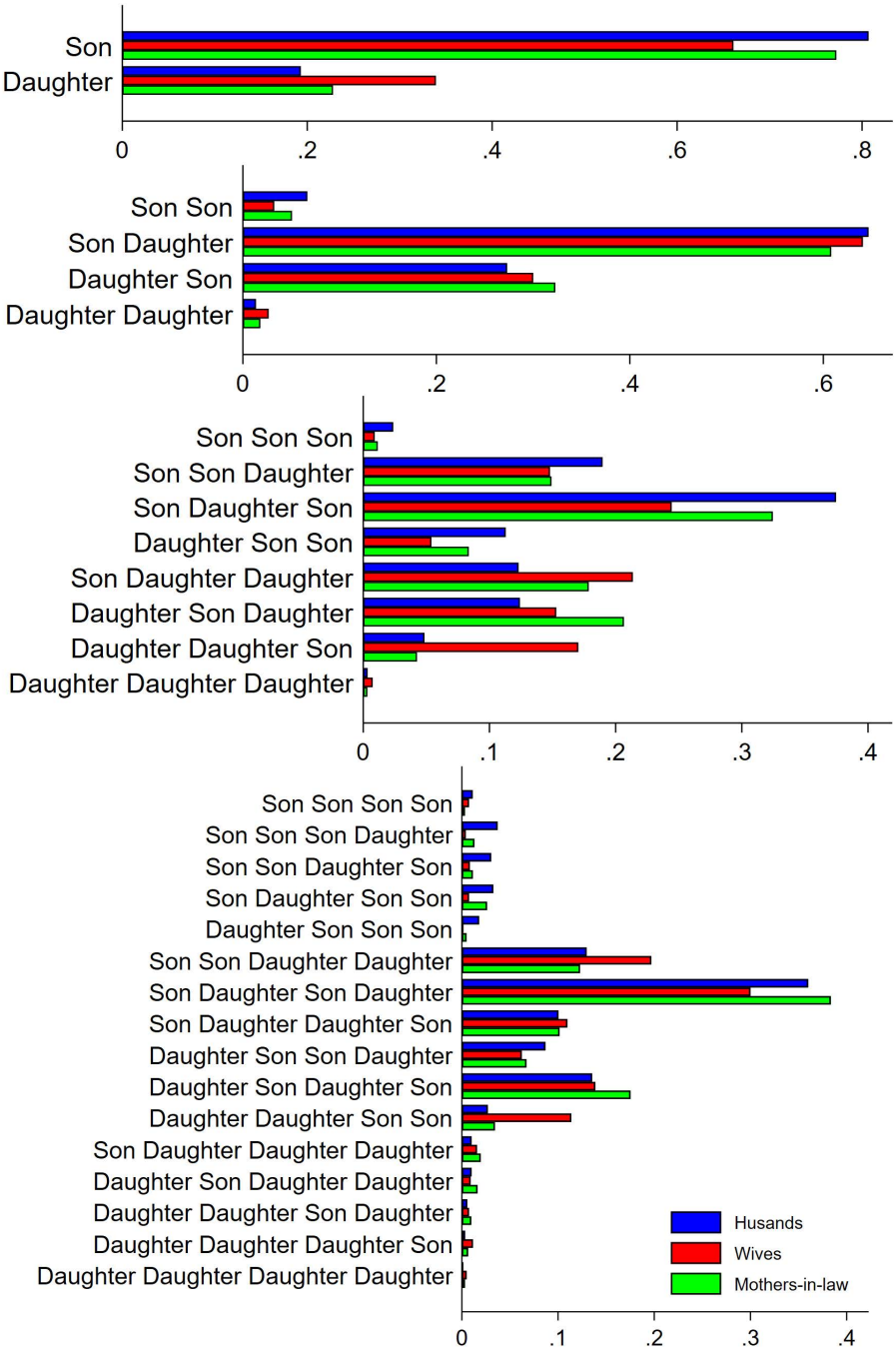

Figure A.6: Frequencies of reported optimal gender composition for the future families of own offspring.

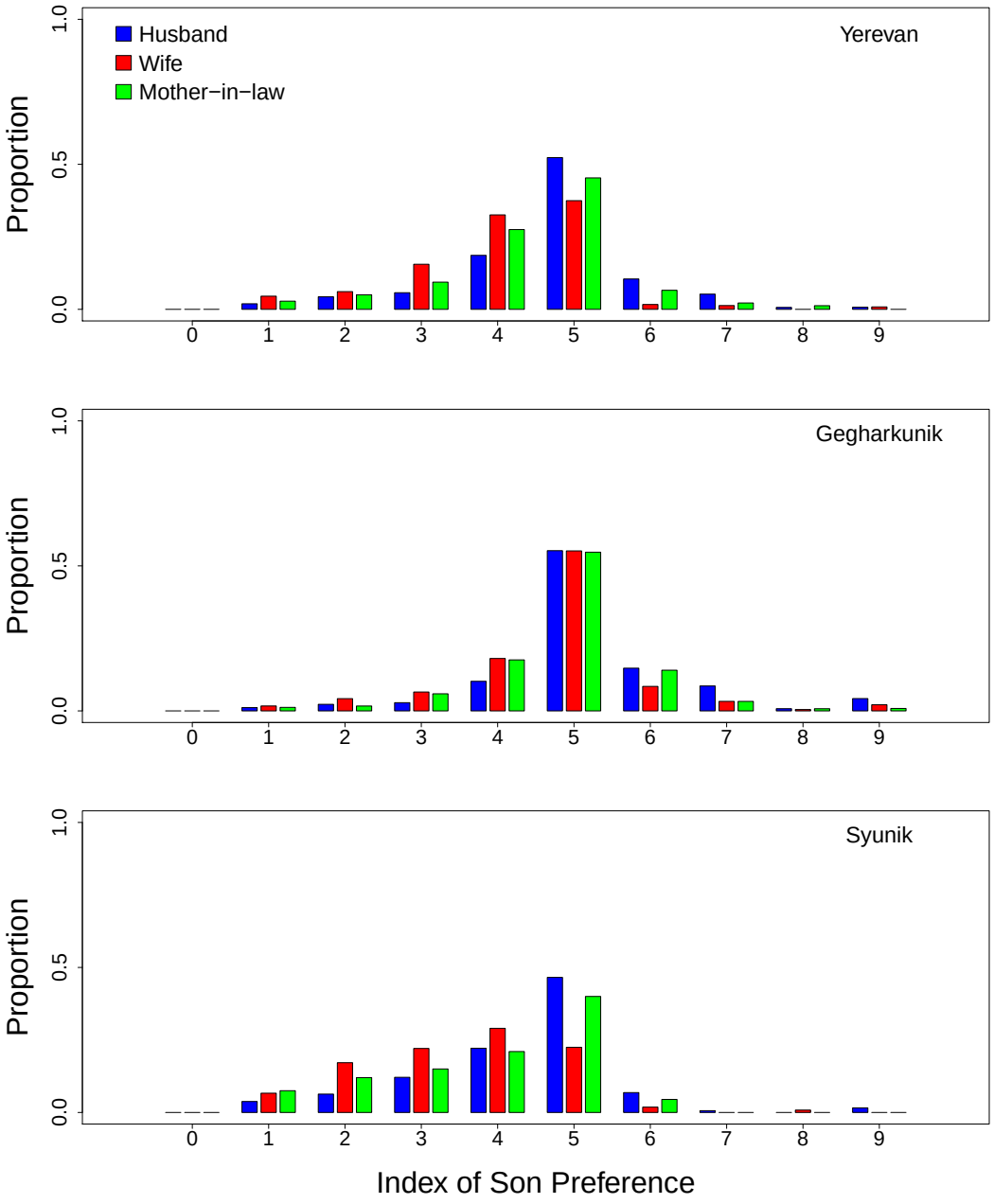

Figure A.7: Histogram of our index of son bias by region and subject type.

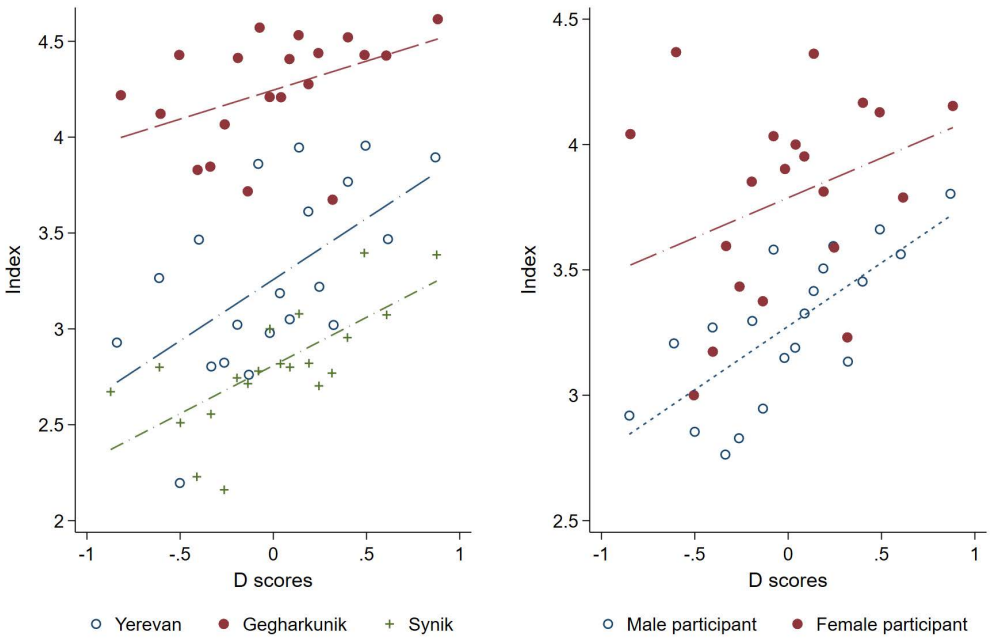

Figure A.8: Binned scatter plots

type, or family composition absorbs some of the variation in son bias, as can be seen by the fact that the adjusted R^2 increases from 0.016 to 0.192. The control variables cannot, however, absorb any measurement error. Hence, the signal-to-noise ratio worsens and attenuation bias becomes more severe.

Finally, in column 5 we include a full set of 1,212 household fixed effects. We find that even if we effectively control for any observed or unobserved variable that varies at the household level, the two measures of son bias are still strongly positively correlated with each other, and the association remains statistically significant at the 1% level. The coefficient in column (5) is identified from variation in D scores and index values across subject types and attests to the fact that measured son bias is systematically larger for men in our sample relative to women.

Comparison to standard measure of son bias A complementary exercise is to compare our measures of son bias to the more standard measure of son bias in the *Demographic and Health Surveys* (DHS). While we cannot replicate the DHS question using our questionnaire items, we can approximate it. For each participant, we can focus on the preferred scenario in terms of family size and compute the share of sons in this scenario. The resulting variable can take values between zero and one, with intermediate steps that depend on the family size in the preferred scenario.

The correlation between our index of son bias and this variable is 0.549. Because our index is based on the same underlying data, we should expect a positive correlation. Finding a correlation well below unity suggests that the index contains a considerable amount of information that is not contained in this simpler DHS-style variable. The raw correlation between our D scores and the DHS-style variable is 0.119. Interestingly, if we use the index and the DHS-style variable to predict D scores in a multivariate regression while accounting for subject-type fixed effects, region fixed effects, and

	Son Pref. Index				
OLS estimates	(1)	(2)	(3)	(4)	(5)
D score	0.578***	0.426***	0.381***	0.205**	0.426***
	(0.0807)	(0.0845)	(0.0849)	(0.0795)	(0.129)
Subject type fixed effects	No	Yes	Yes	Yes	No
Family composition fixed effects	No	No	Yes	Yes	No
Community fixed effects	No	No	No	Yes	No
Household fixed effects	No	No	No	No	Yes
Observations	2695	2695	2695	2695	2695
Adjusted R^2	0.016	0.033	0.076	0.192	0.251

Standard errors are clustered at the household level.

* p<0.1, ** p<0.05, *** p<0.01

Table 1: Internal consistency of our implicit and explicit measures of son bias.

own children fixed effects, only the index comes in significant. This suggests that our index contains more information or is less biased than the variable approximating the DHS-style question.

A.6 Construction of income, education, and occupation variables

The variables on income, education, and occupation used in table 2 in the main text are constructed as follows.

Income level Income levels are measured on an index scale from one to six, based on answers to the question (cf.(**?**)):

> "Which of the following statements best describes the financial situation of your household?"

Answer categories are as follows (percentages of answer in parentheses).

1. The income is hardly enough to buy food (35.8%)
2. The income is enough to buy food, but we can't afford new clothes (23.7%)
3. The income is enough to buy new clothes, but we can't buy technical equipment (27.4%)
4. The income is enough to buy technical equipment, but we can't buy a new car (8.8%)
5. The income is enough to buy anything but a new apartment (3.7%)
6. There are no financial difficulties, and we could buy an apartment if needed (0.6%)

We also asked participants about the monthly household income in Armenian Dram. The index defined above is strongly correlated with self-reported household income ($\rho = 0.44$). However, the correlation is stronger for men ($\rho = 0.48$) than for women ($\rho = 0.39$), which is expected if men in our sample are more likely than women to be the primary breadwinners and to have more accurate information about the exact level of household income. We prefer the index as our main measure of household income because it is likely to be relatively comparable across subject type.

Education level Education levels are measured on an index scale from one to six, based on the question:

> "What is the highest level of education that you have completed?"

Answer categories are (percentages of answer in parentheses):

1. Primary school or no schooling completed (0.6%)
2. Middle school (3.6%)
3. High school (42.2%)
4. Vocational education (27.2%)
5. Higher education (25.3%)
6. Post graduate education (1%)

Occupation We asked husbands and wives about their current occupations. Since we expected a large fraction of mothers-in-law to have reached retirement age, we instead asked them about their main occupation in their thirties and forties.

A.7 Fertility outcomes and measured son bias of husbands and wives

Whether the fertility preferences of father or mothers bear more responsibility for the skewed sex ratio in Armenia is an important question. Table2 shows that fertility outcomes at the household level are correlated with measured son bias of both mothers and fathers, suggesting that the fertility preferences of both parents matter. Note, however, that these individual-level correlations reflect not only the effect of parental son bias on fertility outcomes, but are likely also partly driven by the causal effect of realized fertility outcomes on measured son bias (cf. III.3 in the main text).

	Share of sons			Last child son		
Mother's degree of son bias	0.096***		0.078***	0.128***		0.110***
	(0.016)		(0.016)	(0.025)		(0.025)
Father's degree of son bias		0.100***	0.084***		0.105***	0.083***
		(0.015)	(0.015)		(0.024)	(0.025)
R-squared	0.046	0.054	0.082	0.032	0.022	0.045
N	766	766	766	766	766	766

Table 2: OLS regresssions of son-biased fertility outcomes on a measure of parental son bias. The measure of son bias used in these regression is the composite measure described in section IV.2. We include in these regressions all households with more than one child in which we were able to collect data from both the husband and the wife.