Bias in parental investments and human capital gaps in India

Romaine Loubes*

Paris School of Economics - Université Paris 1

2023

Abstract

This papers shows evidence and produces policy recommendations as to the impact of gender on children's human capital production and parental investments. I use a dynamic model of human capital development, allow for gender-specific cognition and health factors, and estimate production and investment functions separately by genders. I find that, in India, although children produce skills very similarly across genders, parental investments are gender-biased in their elasticity to these skills. While parental investments are elastic to higher skills for boys, they do not react significantly to girls' skills. At an early age, differences in parental investments come from better returns on characteristics for boys, creating scarring effects and eventual skills gender gaps at later stages, thus highlighting the role of gender-biased parental investments in determining educational and occupational gaps.

*romaine.loubes@psemail.eu. I am extremely grateful to David Margolis for his support and feedback. Emily Nix's enlighting explanations of her codes and databases made the project possible. I thank Thomas Breda and seminar participants at PSE for their useful comments.

1 Introduction

South Asia is the region where the educational attainment gender gap is the largest (Barro and Lee, 2013). Simultaneously, as children grow older, boys' and girls' timetables diverge: at 15, girls spend two times more hours per day caring for household members and 41% more time completing chores than their male siblings (Young Lives India). Is this time use difference causing the attainment gap, or is it optimal due to pre-existing educational lags? What of these two gaps precedes the other has important policy implications, because how to achieve gender educational equality depends on how children accumulate human capital: if boys and girls do not produce human capital at the same critical periods in their development, policies that aim at reinforcing the educational system could fail in decreasing the gender gap or even deepen it. Similarly, if the key issue lies in gender-biased parental investments rather than with the schooling system, educational policies taking place outside of the households will be inefficient. Disentangling gender gaps caused by parental investments from those inherent to different production patterns is crucial to achieve higher human capital stocks at national scales, with high development stakes. It is, however, as important as it is complex, especially given that investments and skills are deeply intertwined and persistent making early adverse events scarring (Attanasio et al., 2020a). In this paper, I adapt a dynamic model of non-linear skills production and parental investments, and answer these questions: how does gender factor into the production of cognition and health among children? Does gender lead to different investment patterns on the parental side, and are these differences justified by gender-dependent skills production functions?

In order to disentangle gendered production of skill from gendered reaction of parental investment to these skills, I adapt the Attanasio et al. (2020a) model of dynamic human capital production to allow for sex-specific health, cognition and parental investment factors, and estimate models separately by gender on Young Lives India panel data. First, I estimate the joint distribution of latent cognition, health and parental investment factors for two gender-specific samples. Secondly, I estimate separate skill production and parental investment functions for these two samples. Finally, I use the results from the separate estimations to provide policy recommendations through an Oaxaca-Blinder decomposition of gaps in parental investment and simulations of counterfactual parental investment scenarios on gender gaps outcomes.

Splitting the estimation results in production and investment patterns that using gender as a control would not have allowed to uncover. First, I find that boys and girls' skill creation patterns are very similar: for both sexes, skills feed into themselves from age to age to a great extent, and they are also hugely dependent on parental investments, similarly to Attanasio et al. (2020a), to Cunha et al. (2013) and a number of other models. The persistence of skill and dependence on investments signal both the importance the latter and the potential scarring effect of its scarcity at early ages. Secondly, I show that parental investments, however, do not react to the same signals across genders: they are significantly elastic to boys' skills, while they are not to girls' and seem to be more elastic to children goods prices for the latter - highlighting, as notes by Almond and Currie (2009), the possibility that the same income shocks might have different effect on children across genders. The investment results contrast with Attanasio et al. (2020a), who found no big effect of gender on investments, differently from empirical works such as Barcellos et al. (2014). Importantly, this suggests that allowing for factors to be gender-specific and estimating production functions separately uncovers overlooked gaps. Finally, in my decomposition, I show some evidence that boys benefit from higher valuation of their skills at early ages, resulting in higher investments and, eventually, higher later ages skills.

I hope to bring value to several strands of the human capital literature. First, empirical works interested in quantifying the impact of parental inputs (investment) gaps in terms of outputs (anthropometric measures, test scores, skill factors) have yet to reach a consensus, even in settings like India where the existence of son preference has been documented at length (Jayachandran and Kuziemko, 2011; Barcellos et al., 2014; Jayachandran and Pande, 2017; Bhalotra and Cochrane, 2010). Barcellos et al. (2014) find evidence of gaps in parental investments once allowing for endogenous household size because of gender-biased stopping behaviors. However, even with this striking result, they do not find evidence of children's outcomes being affected, making it difficult to attach policy implications to their finding. Similarly, determining the cause of gender parental input gaps is also not straightforward, for three reasons. First, parental investments could be linked to potentially different timings in terms of human capital production across genders, making differences in investments optimal: seminal works like Sommerfelt and Arnold (1998) show that girls do better in terms of anthropomorphic measures up to the age of 2, but not afterwards, an example of lagged process of health acquisition. Second, different investments could also be optimal given prior skill endowments, which are themselves related to prior investments. Finally, analyses of skill factors based on real-life measurements could be biased by the fact that gender-dependent norms make children's sex more than a control in the production-investment nexus: parents could potentially not care about the same dimensions of their children's skills, modifying not only the level but also the determinants and the composition of the inputs they put into their children. Children of different sexes showing similar cognitive or health-related measurements are not necessarily going to be assigned the same skill score by those observing them, like their parents - for example, because they are raised to exhibit different levels of competitiveness or self-efficacy impeding that their skills be perceived in the same way, as shown by recent works Berhman et al. (2021). I contribute to this literature through a dynamic model that can precisely disentangle and solve for the intertwining of production and investments. Having a dynamic model, estimating production and investment functions at several ages allows me to account for potentially different timings of skill production across genders.

My paper is situated in the flourishing literature of dynamic models of children's human capital accumulation, (Cunha et al., 2010, 2013; Attanasio et al., 2020a,b). Recent advances incorporated by the Attanasio et al. (2020a) model that I use include parents' ignorance of their children's skill production function, which is justified by empirical evidence from Cunha et al. (2013), the integration of the over-restriction risk discussed in Agostinelli and Wiswall (2016) through normalizing the skills factor only at baseline, and integration of measurement error in skill factor components (Cunha et al., 2010), and allowing for endogeneity of parental investments to skill production shocks -"safety nets"-, which they later find to be significant at all ages. My paper's contribution to that literature is assessing the consequences of allowing for gender-specific human capital factors, and especially, showing that dropping the assumption that parents different genders' skills similarly leads to very different results in terms of parental investments.

My results on gender-biased parental investments and their persistence also add to a rich strand of studies of human capital development (reviewed in Almond and Currie (2009)). In particular, gender-dependent patterns in human capital production have been the object of developed (Buchmann et al., 2007) and developing countries literatures, be it individual patterns - Conti et al. (2010), for example, study the interlinks between education and healthy behaviors, finding a feedback effect of the former onto the latter - or intergenerational, asi in Jayachandran and Lleras-Muney (2009) through mothers' influence on investments and educational in the Philippines. In particular, having a dynamic model highlights that early gender-dependent gaps in parental valuation of skills can set children on different skill accumulation paths through strong dynamic effects of differential investments, without gender-dependent skill production patterns - suggesting persistence at the generational scale as well. Finally, I wish to speak to the intrinsically important issue of gender inequality in parental investments. In the most drastic cases, gaps in parental inputs leads to disproportionately higher mortality rates for girls (Sen, 1990), a 14% share of which can be explained by investments like breastfeeding in India (Jayachandran and Kuziemko, 2011). I contribute by offering policy-relevant insights through an Oaxaca-Blinder decomposition of the causes gender gaps in parental investments and simulations of different input changes on skill gaps. As Barcellos et al. (2014) state, the right strategy depends on the reasons why girls are subject to less parental input, and gender policies could be useless if not well informed. Similarly, interventions aiming at closing educational gaps in India might fail if unaware of mechanisms behind these gaps – whether they be investments, different timings for skill formation, or gender-biased aspirations - making it important to produce evidence and policy recommendations that are comprehensive of gender's role in this nexus.

The rest of this paper is structured as follows: section 2 expands on the empirical strategy, and section 3 presents the Young Lives database. In section 4, I will estimate the model on separate samples, and after recovering the estimated coefficients, use it in policy-making and normative exercises. Discussion and conclusion follow.

2 Empirical Strategy

My empirical strategy consists of the recovery of the joint distribution of health, cognition, and investment factors in the form of a mixture of multivariate normal laws. This is done through maximum likelihood estimation of the mixture with an E-M algorithm followed by a minimum distance estimation. When these parameters are obtained, 10.000 synthetic individuals are drawn from each of the two latent factor joint distributions, on which parameters for the skill production functions and parental investment functions are estimated. I first detail how the latent factor are jointly estimated, before turning to the functional forms taken by production and investment functions.

2.1 Factor estimation

The model departs from a measurement system which encompasses tests scores, anthropometric measures and controls such as household caste and religion and order them into a measurement system that shares common parameters with the distribution of the latent skills θ - cognition, health and parental investments at ages 1, 5, 8 and 12. Importantly, Attanasio et al. (2020a) allows measurement of latent skills measures to be error-ridden :

$$m_{jkt} = a_{jkt} + \lambda_{jkt} ln(\theta_{kt}) + \epsilon_{jkt} \tag{1}$$

m being the measure j of skill k at time t (for example, PPVT[1] for cognition at age 12), and θ_{kt} skill k at time t. The measurement is dedicated: that each measurement depends on one factor only. In matrix form, this gives

$$M = A + \Lambda ln(\theta) + \Sigma\epsilon \tag{2}$$

With Σ the diagonal matrix of standard deviations for the errors (as the control measurements are assumed to have no measurement error, their standard errors are set to 0) and ϵ a vector of normally distributed errors.

To achieve comparability of factors across ages and across genders, factors must have a common scale. For this, one measurement per latent factor serves as an anchor (Agostinelli and Wiswall, 2016), which means that their loading λ is set to one. This is the PPVT for cognition, health for age Z-score for health and clothing expenditures for parental investments. Some other measurements are left free and others, assumed

[1] Peabody Picture Vocabulary Test

to not be related to the latent factors or being measurements for controls, have their coefficient set to zero in Λ.[2]

We are trying to estimate the distribution of θ, the underlying skill vector, which is assumed to be a mixture of two multivariate normal distributions. This means that the joint distribution of θ can be written:

$$F_\theta = \tau\Phi(\mu_A, \Omega_A) + (1-\tau)\Phi(\mu_B, \Omega_B) \tag{3}$$

With ϕ the distribution of multi-variate normal laws of vector of means μ and variance-covariance matrixes Ω, mixed with weights τ. The first step in this estimation is to obtain the distribution of the M matrix in equation 2, since it is the one we depart from. When we assume the errors to be normally distributed, we obtain for M :

$$F_M = \tau\Phi(\Pi_A, \Psi_A) + (1-\tau)\Phi(\Pi_B, \Psi_B) \tag{4}$$

$$\begin{aligned} \text{With } \Psi_A &= \Lambda^T\Omega_A\Lambda + \Sigma \\ \Pi_A &= A_t + \Lambda\mu_A \end{aligned} \tag{5}$$

$$\begin{aligned} \Psi_B &= \Lambda^T\Omega_B\Lambda + \Sigma \\ \Pi_B &= A_t + \Lambda\mu_B \end{aligned} \tag{6}$$

Factors are normalized at baseline only, to allow for factor growth as children get older (Agostinelli and Wiswall (2016)). This normalization as well as the imposed factor loadings allow us to identify the means, mixture weights and covariance-variance matrices in 3 steps:

1. MLE is used to recover the means, covariance matrices and mixture weights of the matrix form of the measurement system in equation 2 separately for each gender. it is implemented using an expectation-maximisation algorithm drawn from Dempster et al. (1977) and further developed in Arcidiacono and Jones (2003) (Attanasio et al., 2020a). The E steps estimates the probability for one individual from the data to be drawn out of the dataset with particular parameters for joint normals, and then the M steps maximise this probability, iterating over the parameters until convergence is reached.

2. Then, a minimum distance estimator is implemented : taking into account the restrictions on the loadings matrix (zero loadings, normalisations, etc.), we recover Λ, the constant as well as the missing parameters for the normal distributions. These steps are performed separately for boys and girls, as

[2] For example, for the two first loadings with respectively 3 and 2 measurements, we would have $\begin{pmatrix} 1 & 0 \\ \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & 1 \\ 0 & \lambda_3 \end{pmatrix}$

there is no obvious reason in our gendered analysis to assume that latent health and cognition as well as parental investment are distributed the same. In subsection A.2, I show that these factors depart from the same scale and location, making them comparable across ages and genders.

3. From these distributions, 10.000 synthetic individuals are then drawn randomly for the investment and production function coefficients to be estimated on.

2.2 Functional form

In the model, skills (cognition and health) at stage $t+1$ are a CES function of their and the other skill's lagged values as well as on parental cognition and health, investment and a variety of controls.

θ represents latent factors, with indexes determining what element of the latent factor vectors it refers to. For cognition (c) and health (h), and for each gender g, we estimate

$$\begin{aligned} \theta_{(c,g)t+1} &= [\delta_{(c,g),t}\theta_{(c,g)t}^{\rho_t} + \delta_{(h,g),t}\theta_{(h,g)t}^{\rho_t} + \delta_{(cp,g),t}\theta_{(cp,g)}^{\rho_t} + \delta_{(hp,g),t}\theta_{(hp,g)}^{\rho_t} + \delta_{(I,g),t}\theta_{(I,g)t}^{\rho_t}]^{1/\rho_t} A_{(c,g)t} \\ \theta_{(h,g)t+1} &= [\alpha_{(c,g),t}\theta_{(c,g)t}^{\zeta_t} + \alpha_{(h,g),t}\theta_{(h,g)t}^{\zeta_t} + \alpha_{(cp,g),t}\theta_{(pc,g)}^{\zeta_t} + \alpha_{(hp,g),t}\theta_{(hp,g)}^{\zeta_t} + \alpha_{(I,g),t}\theta_{(I,g)t}^{\zeta_t}]^{1/\zeta_t} A_{(h,g)t} \end{aligned} \tag{7}$$

cp, hp and I are, respectively, parental cognition, health and investment. A represents efficiency terms of the form $A_{(c,g)t} = e^{d_{0t}+d_{X_t}X_t+u_{(c,g)t}}$ and $A_{(h,g)t} = e^{a_{0t}+a_{X_t}X_t+u_{(h,g)t}}$, ρ and ζ represent elasticity terms, and $\Sigma\delta = 1$ and $\Sigma\alpha = 1$.

Human capital also depends on parental investments, which in turn are influenced by skill production shocks: investment is thus instrumented by children-related prices and income, and errors from the production and investment functions are tied together through a control function κ. With parental investments I for gender g at time t

$$\begin{aligned} \log(\theta_{(I,g)t}) &= \gamma_0 + \gamma_{(c,g)t}\log(\theta_{(c,g)t}) + \gamma_{(h,g)t}\log(\theta_{(h,g),t})\log(\theta_{(cp,g)}) + \gamma_{(hp,g)t}\log(\theta_{(hp,g)}) + \gamma_{X_t}^{'}\log(X_t) \\ &+ \gamma_{(cp,g)t}\log(\theta_{(cp,g)}) + \gamma_{pt}^{'}\log(p_{I_t}) + \gamma_Y\log(\theta_{Y_t}) + v_t \end{aligned} \tag{8}$$

With income Y and p a price vector for children-related goods, we thus have errors of this form:

$$\begin{aligned} E(u_{(c,g),t}|Q_t, Z_t) &= \kappa_{(c,g),t}v_t \\ E(u_{(h,g),t}|Q_t, Z_t) &= \kappa_{(c,g),t}v_t \end{aligned} \tag{9}$$

Q_t and Z_t are respectively the controls in the production equation and the excluded instruments from the investment equation, such as income.

In this specification, income only affects skills through its correlation with parental investment, because there could be reverse causality issues was income to be included in the skills production functions: for example, through less need for tutoring if a child has high cognition or less need for medical care if they are healthy. Robustness tests show that the estimates do not change when income is included in the production function or when it is excluded from both the production and the investment functions.

3 Data

Skills are latent factors that cannot be measured through one test only. For that reason and in the context of developing countries in particular, lack of data has been an issue in conducing quantitative research on human capital (Attanasio et al. (2020b), Almond and Currie (2009)). This has been partly addressed by the creation of the Young Lives Database (also used in Berhman et al. (2021) and Attanasio et al. (2020a)) which covers 4 developing countries. For its India section, it is a panel study going from 2002 to 2017, comprised of 5 rounds and conducted in Hyderabad and Andra Pradesh. The younger cohort, which is the one being examined here, was sampled at ages 1, 5, 8, 12 and 15. A variety of scales (community, household and child-level) are explored, and living standards, school and test results, anthropometric measures as well as health behaviours are measured.It assesses children's cognitive (PPVT, maths, language tests) and health (height for age, BMI, wasting, etc.) capital stocks. As tests are conducted at home, this has the additional value of avoiding selection (and most likely a gendered one) on children being present at school (Berhman et al. (2021)). The panel has very low attrition rates: 5,9% for the 5 rounds, including a mortality rate of 2,2% (4.8% for the four rounds). There are 1910 children present in four rounds (884 girls), and 1890 (872 girls) present in five.

I use the Attanasio et al. (2020a) datasetderived from Young Lives for purposes of comparability, although some adjustments, detailed and justified in the appendix have also been made. I will first explain the variables used for the computation of health and cognitive skills before presenting summary statistics.

In the health and cognition factors' computations, one variable is defined as an anchor through time, which means that it is present in all rounds (Agostinelli and Wiswall (2016) as cited in Attanasio et al. (2020a)) and that its weight is always set to one in the factor loadings. In our case, it is the z-score of height for age for health and the percentage of correct answers on the PPVT for cognition. As shown in Borghans et al. (2016), grades and achievement test measures such as the PPVT are better in predicting later outcomes than "pure" intelligence measures such as the IQ test, because they are sensitive to personality variables such as self-confidence and self-efficacy that influence these outcomes positively. When one cannot measure gender gaps in non-cognitive skills, having these measures at hand and constructing factors separately for each genders thus has the advantage of encompassing how well a child functions in a broader sense than pure intelligence. Others variables used for these cognition factors include language (English and Telugu), mathematics and EGRA[3] tests. The PPVT is a test that requires associating an image among four with the word given by the instructor. I chose to use the percentage of correct answers and not the score because, as the number of questions asked varied across rounds (without a difference in overall difficulty), the same scores correspond to different levels of cognitive skills across rounds. [4]

Health for age is the anchor for the health factor. Weight and parental opinion of their children's health comprise the rest of the measures. Investment is anchored on a measure of parental expenses on clothing for the child. Other expenses include shoes, school uniform, once again , the number of meals per day and the number of food groups present in the child's diet.

Parental cognition and age are measured at the first wave of the sample through their education and literacy levels as well as the mother's weight and height, and are fixed through time. All of these measures are standardized over the pooled sample with both genders (of course, the z-scores of height for age are computed separately for each gender as boys and girls grow differently), before splitting for factor estimation.

	Girls	Boys	P-value
Region			
Telangana	0.31	0.38	0.003***
Costal Andhra	0.37	0.34	0.157
Rayalaseema	0.32	0.29	0.107
Urban	0.22	0.25	0.191
Religion and Socio-economic status			
Muslim	0.05	0.08	0.022**
Hindu	0.89	0.87	0.141
SC	0.20	0.17	0.214
ST	0.15	0.15	0.933
BC	0.48	0.46	0.382
Wealth index	0.40	0.41	0.304
Family characteristics			
Mother's level of education	2.98	2.94	0.829
Mother's age	23.64	23.66	0.890
Nb older siblings	0.65	0.73	0.087
Any older brother	0.24	0.25	0.755
Child early health			
Height-for-age z-score	-1.21	-1.44	0.002***
Stunted	0.28	0.35	0.003***

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

Table 1: Baseline

[3] Early Grade Reading Assessment

[4] In rounds 2 and 3, the questionnaire begins age-dependent set and goes backwards until the child gives two correct answers. All of the questions are then asked until a whole set (8 answers) is wrong, and the score is then computed assuming that all of the answers after this ceiling would have been wrong, and all above the base set right. In rounds 4 and 5, all of the questions are asked. Less than 5% of children get 8 or more answers wrong in a row before answering one question correctly, which suggests that our results are comparable accross rounds. The 57 questions in rounds 4 and 5 are drawn uniformly from the 204 question set of rounds 2 and 3, so overall difficulty is unchanged.

3.1 Summary statistics

As we can see in the baseline statistics (when Young Lives children are 1 year old) presented in table 1, most households pertain to BCs (backwards castes), or to a lesser extent, scheduled castes or tribes[5]. The sample is also characterized by low rates of education for mothers (less than third grade) and a mostly rural setting. Girls seem to be under-represented in Telangana and in Muslim households: as region and religion have been cited to be determinants of household eldest son preference in Jayachandran and Pande (2017), this justifies investigating these differences in endowments in our decomposition and simulations.

Girls also seem to be consistently healthier than boys (with respect to both measures) and less stunted, with z-scores for height being significantly different up until 15 years old, as it can be seen in tables 3 and 4 - although one can notice very high rates of stunting for both genders. At baseline, girls do not belong to significantly larger families and are as likely as boys to have at least one older brother, which is a variable we have introduced to account for possible eldest son preference. However, over time, they become significantly (at the 1% level for age 12 and at the 5% level for the rest) more likely to have a younger brother, and consequently they tend to pertain to larger families (significantly below the 10% threshold for ages 12 and 15).

[5]Scheduled caste or tribes do not pertain to the four main Indian castes and were historically violently discriminated against. Scheduled tribes are also isolated geographically speaking. Backwards castes are other communities that are also discriminated against, but without pertaining to SC/ST.

	age5			age8			age12			age15		
	Girls	Boys	P-value	Girls	Boys	P-value	Girls	Boys	P-value	Girls	Boys	P-value
Test scores												
Percentage correct in PPVT	13.31	13.58	0.578	27.20	29.97	0.000***	75.03	76.04	0.112	82.18	83.89	0.008***
Raw score in Math Test				10.29	10.34	0.829	12.77	12.78	0.966	9.80	10.72	0.000***
Reading (EGRA)				2.36	2.54	0.067*						
Health variables												
Height-for-age z-score	-1.57	-1.71	0.01***	-1.37	-1.48	0.038**	-1.48	-1.45	0.526	-3.16	-1.46	0.278
Weight-for-age z-score	-1.84	-1.89	0.232	-1.77	-1.96	0.000***						
Stunted	0.34	0.38	0.095 *	0.28	0.31	0.078 *						
Younger Brother	0.24	0.21	0.120	0.28	0.24	0.041**	0.31	0.25	0.005***	0.30	0.25	0.022**
Number of siblings	1.46	1.42	0.411	1.57	1.51	0.180	1.63	1.53	0.054*	1.61	1.53	0.078*
Parental income and expenses												
Wealth index	0.46	0.46	0.604	0.51	0.52	0.065*	0.58	0.59	0.079*	0.63	0.64	0.138
Medical Expenses	765.94	809.78	0.640	520.13	510.60	0.926	493.39	833.37	0.119			
once again expenses	153.87	157.29	0.774	431.80	446.96	0.611	710.49	883.76	0.000***			
Time Use												
Hours/day spent in caring for hh members	0.22	0.14	0.003***	0.25	0.18	0.002***	0.19	0.09	0.000***	0.25	0.11	0.000***
Hours/day spent in hh chores	0.07	0.04	0.025**	0.45	0.24	0.000***	1.04	0.71	0.000***	1.41	1.00	0.000***
Hours/day spent at school	5.63	5.84	0.044**	7.61	7.73	0.019**	7.93	8.05	0.115	7.67	7.97	0.034**
Hours/day spent studying outside school	1.02	1.06	0.490	1.89	1.78	0.026**	1.94	1.90	0.456	2.18	2.06	0.062*
Hours/day spent in leisure activities	5.64	5.80	0.262	4.67	4.86	0.016**	3.76	4.06	0.000***	3.51	3.63	0.164

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

Table 2: Summary statistics

Panel data also allows us to look at the formation of test scores gaps. We can observe a sudden drop in girls' relative maths scores at age 15 (with a gap of nearly one point) and a p-value below 10% for differences in EGRA scores. As for PPVT, we can see that girls fare consistently worse than boys (although statistically significantly so at ages 8 and 15 only).

Only at age 12 do parental monetary investments seem to differ. Time-use, on the contrary, is different from an early age and increasingly so: girls spend relatively more and more time caring for household members, performing chores or studying outside of school, and boys get relatively more school and leisure time for every round. At age 15, girls spend 1.41 hour each day performing household chores an 7.67 hours at school while boys spend respectively 1 and 7.97 hours on these activities: this increasing gap justifies searching for the determinants of test scores gaps, as inferior ones for the girls might lead to less investment in school (and, in turn, to an increasing educational gap).

4 Results

I will now present results on investment and skills functions, before turning to the consequences of observed differences in coefficients through an Oaxaca-Blinder type decomposition. Note that we are comparing coefficients for parental investments, cognitive and health skills that are computed in a gender-specific manner: the factors do not follow the same distributions, and neither do the loadings on the Λ matrix, between genders. However, they have the same baseline scale and location, and are scaled on the same variable at each age, making them comparable in decompositions, as is shown in the Appendix. Also, through our exploration of the differences in determinants of production between boys and girls, we will be focused on significance, and not magnitude, of coefficients: significance can always be compared and these factors are conceptually equivalent.

4.1 Skills production functions

Table 3 and Table 4 present health and cognition production functions for girls and boys, from the estimation of equation 7 CES functions. Coefficients are to be interpreted as elasticities: at age 8, a 10 % increase in t cognition is associated with an increase of 4.9% of t cognition for both genders at age 8, and with an elasticity of respectively 0.8 and 0.91 for girls and boys at age 12. The first result from these estimations is that each skills reproduce itself to the same extent for all children at all ages - their bootstrapped confidence intervals always include each others' estimates - and skills also depend on parental investments to the same extent: a lot for cognition at all ages and less for health passed age 5, probably because of the superior persistence of health itself. This is in line with Attanasio et al. (2020a), my seminal model. It is reassuring that

	Girls			*Boys*		
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	0.49 [0.26,0.63]	0.8 [0.65,1]	NA [NA,NA]	0.49 [0.29,0.58]	0.91 [0.79,0.95]
Health	0.03 [0,0.06]	0.03 [0,0.09]	0.01 [−0.07,0.03]	0.02 [−0.01,0.04]	0.05 [0.02,0.09]	0 [−0.01,0.06]
Parental Cognition	0.01 [−0.07,0.1]	−0.05 [−0.08,0.04]	0.04 [−0.07,0.07]	0.02 [−0.03,0.07]	0.02 [−0.02,0.06]	0.02 [0,0.08]
Parental Health	−0.01 [−0.07,0.06]	−0.05 [−0.11,0.01]	0.01 [−0.01,0.09]	−0.03 [−0.1,0.02]	0 [−0.06,0.04]	0.01 [−0.02,0.05]
Investment	0.97 [0.56,1.25]	1.37 [0.86,1.46]	1.09 [0.43,1.18]	1.23 [0.64,1.36]	0.84 [0.47,1.09]	1.51 [0.66,1.69]
Complementarity	0.03 [−0.27,0.43]	−0.3 [−0.43,0.13]	−0.15 [−0.22,0.52]	−0.21 [−0.34,0.31]	0.09 [−0.13,0.41]	−0.54 [−0.72,0.2]
Elasticity of Subst.	1.03 [0.79,1.72]	0.77 [0.7,1.15]	0.87 [0.82,2]	0.82 [0.75,1.45]	1.1 [0.89,1.7]	0.65 [0.58,1.24]
Log TFP	−0.01 [−0.06,0.03]	0.36 [0.23,0.42]	1.61 [1.38,1.71]	−0.02 [−0.04,0.02]	0.47 [0.39,0.51]	1.56 [1.47,1.64]
Investment Residual	−0.84 [−1.01,−0.53]	−0.5 [−0.82,−0.32]	−0.15 [−0.37,−0.07]	−0.97 [−1.14,−0.79]	−0.47 [−0.71,−0.32]	−0.08 [−0.19,0.03]
Number of Children	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.03,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.02,0]
Older Siblings	0.01 [0,0.02]	−0.01 [−0.01,0.01]	0 [−0.01,0.02]	0 [−0.01,0.01]	−0.01 [−0.01,0]	0 [−0.01,0.01]
Urban	0 [−0.01,0.01]	−0.01 [−0.01,0]	−0.01 [−0.02,0]	0.01 [0,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.01]
Hindu	−0.01 [−0.03,0]	−0.01 [−0.02,0]	0.01 [−0.01,0.02]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.02 [0.01,0.03]
Muslim	0 [−0.01,0.01]	0 [0,0.01]	0 [−0.03,0]	0 [0,0]	0 [0,0]	0 [−0.01,0]
Mother's Age	−0.01 [−0.02,0.01]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.02,0]
Scheduled Caste	−0.01 [−0.02,0]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0]	0.01 [0,0.01]	0.01 [0,0.02]
Scheduled Tribe	0.02 [0.02,0.04]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.02 [0.01,0.03]	0 [−0.01,0]	0 [0,0.01]
BC Caste	0 [−0.01,0.01]	0 [−0.01,0.02]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.01 [0,0.02]

p-value=0: p-value<0.0001
Confidence intervals obtained through 100 bootstraps samples

Table 3: Cognition - girls and boys

	Girls			*Boys*		
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	−0.05 [−0.21,0.05]	−0.04 [−0.18,0.16]	NA [NA,NA]	−0.12 [−0.23,0.02]	0.05 [−0.03,0.13]
Health	0.48 [0.39,0.58]	0.91 [0.8,1.06]	0.9 [0.79,1.05]	0.48 [0.42,0.53]	0.86 [0.75,0.91]	0.87 [0.79,0.93]
Parental Cognition	−0.01 [−0.06,0.08]	0.06 [0.03,0.13]	−0.01 [−0.06,0.06]	−0.01 [−0.07,0.06]	0.06 [0.04,0.11]	−0.01 [−0.05,0.04]
Parental Health	0.18 [0.08,0.37]	0.01 [−0.05,0.1]	0.14 [0.05,0.22]	0.31 [0.21,0.42]	0.1 [0.06,0.19]	0.01 [−0.05,0.07]
Investment	0.86 [0.62,0.96]	−0.21 [−0.34,0.11]	−0.11 [−0.29,0.13]	0.69 [0.56,0.82]	−0.22 [−0.36,0.02]	0.21 [−0.1,0.27]
Complementarity	−0.03 [−0.07,0.06]	0.23 [−0.07,0.35]	0.08 [−0.12,0.2]	0.01 [−0.07,0.08]	0.2 [0,0.28]	−0.07 [−0.16,0.24]
Elasticity of Subst.	0.97 [0.93,1.07]	1.3 [0.93,1.53]	1.09 [0.89,1.25]	1.01 [0.93,1.08]	1.25 [1,1.38]	0.93 [0.86,1.32]
Log TFP	−0.37 [−0.44,−0.25]	0.1 [−0.01,0.17]	−0.06 [−0.16,0.14]	−0.33 [−0.4,−0.26]	0.12 [0.04,0.2]	−0.06 [−0.14,0.02]
Investment Residual	−0.75 [−1.47,−0.42]	−0.03 [−0.29,0.15]	−0.01 [−0.12,0.28]	−0.48 [−0.85,−0.1]	−0.12 [−0.3,0.11]	−0.07 [−0.17,0.03]
Number of Children	0.01 [−0.02,0.03]	−0.01 [−0.02,0]	−0.01 [−0.03,0.01]	0.02 [0,0.04]	0 [−0.01,0.02]	0 [−0.01,0.02]
Older Siblings	−0.03 [−0.06,0]	0.02 [0,0.03]	0.01 [−0.01,0.03]	−0.02 [−0.04,0]	−0.01 [−0.02,0]	0.01 [−0.01,0.02]
Urban	0.01 [−0.01,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.01,0]
Hindu	0 [−0.02,0.02]	0 [−0.01,0.01]	0 [−0.03,0.02]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]
Muslim	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]
Mother's Age	0 [−0.01,0.02]	0 [−0.02,0.01]	−0.02 [−0.03,0]	0 [−0.02,0.01]	0 [−0.01,0.02]	−0.02 [−0.03,0]
Scheduled Caste	0 [−0.01,0.01]	−0.01 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.02,0.01]	0 [−0.02,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0 [−0.01,0.02]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.02 [0.01,0.03]	−0.01 [−0.02,0]	0 [0,0.01]
BC Caste	−0.02 [−0.03,0]	0.01 [0,0.02]	−0.01 [−0.02,0]	−0.01 [−0.03,0]	0.01 [0,0.02]	0 [−0.01,0.01]

p-value=0: p-value<0.0001
Confidence intervals obtained through 100 bootstraps samples

Table 4: Health - girls and boys

much of the results are similar with their conclusions, and means that any diverging result on investment can be compared directly rather than accounting for other conflicting results in mind. One coefficient is significantly higher than in the pooled model of Attanasio et al. (2020a): TFP, which is A in the equations and is an efficiency factor, is positive and increases a lot for both genders' cognition production functions. One explication could be that, as children get older, they are more likely to absorb any input, an hypothesis that the fact that the TFP also increases with age in the pooled sample analysis in the Appendix supports. Parental cognition and health only seem to play a role in children's health levels, and a slightly more persistant one for boys as the positive effect of parental health on children's health lasts until age 8 for boys, whereas it stops having an effect at age 5 for girls. Parental investments are associated positively and significantly to cognition at all ages, and only at age 5 for their effects on health for both genders.
ρ or ζ, the coefficients for complementarity, always have bootstrapped confidence intervals that include 0, and we thus cannot reject the fact that the CES function is equivalent to a Cobb-Douglass one. Finally, investment residuals are an important part of our analysis for two reasons: they constitute a test for exogeneity of investment with regards to skills production - a test that is failed for all instances of cognition and for health at age 8 for both genders - and they can be interpreted as safety nets: the fact that investment rises to face negative production shocks in the children's skills development. Here, we can see that the existence of such safety nets is made possible by most of the coefficients being significant (especially for cognition). These nets seem to be of comparable scales between genders, with bootstrapped confidence intervals overlapping systematically. Other factors do not seem to have an effect no matter the gender, skill, or age.

Armed with the results that children's skills production patterns are similar across genders, I now turn to the analysis of the determinants of parental investments.

4.2 Parental Investments

Table 5 presents the determinants of parental investment between genders from the estimation of Equation 8. Similarly to production functions, coefficients are to be read as elasticities. The main finding from this estimation is that parental investment seem to be more elastic to health and cognition when they are investing in boys: we do not reject the hypothesis that girls' cognition does not affect investments at each age, while it has a significant positive coefficient at each age for boys. Health also has a positive and significant effect on investment for boys at age 12 that it does not have for girls: an increase of 10% of health units at age 8 leads to a rise of 1.80% of parental investments units at age 12 for boys.
Parental cognition does not seem to play a role in parental investments no matter the gender and age, like parental health (except negatively for investment in boys at age 12). In terms of prices, we can see

	Girls			*Boys*		
	Age 5	Age 8	Age 12	Age 5	Age 8	Age 12
Cognition	NA $[NA,NA]$	0.138 [−0.08,0.3]	0.449 [−0.22,1.08]	NA $[NA,NA]$	0.293 [0.03,0.45]	0.726 [0.15,1.23]
Health	0.032 [0,0.06]	0.002 [−0.06,0.07]	0.038 [−0.09,0.12]	0.02 [0,0.04]	−0.026 [−0.07,0.02]	0.181 [0.02,0.38]
Parental Cognition	0.028 [−0.05,0.09]	0.068 [−0.02,0.18]	−0.085 [−0.17,0.09]	0.02 [−0.01,0.06]	−0.081 [−0.11,0.03]	−0.13 [−0.19,0.12]
Parental Health	0.006 [−0.05,0.06]	0.031 [−0.04,0.15]	0.02 [−0.15,0.15]	0.012 [−0.02,0.06]	−0.055 [−0.11,0.04]	−0.24 [−0.42,−0.03]
Price Clothes	0.005 [−0.02,0.04]	−0.057 [−0.11,−0.01]	−0.028 [−0.11,0.05]	0.043 [0.03,0.06]	0.008 [−0.04,0.04]	−0.051 [−0.14,0.06]
Price Notebook	0.008 [−0.02,0.03]	−0.027 [−0.08,0]	0.144 [0.08,0.21]	−0.001 [−0.02,0.02]	−0.047 [−0.08,0]	0.048 [−0.04,0.14]
Price Mebendazol	0.023 [−0.01,0.04]	−0.029 [−0.08,0.03]	−0.106 [−0.19,−0.03]	0.005 [−0.02,0.01]	−0.043 [−0.07,0.01]	−0.026 [−0.08,0.05]
Price Food	0.008 [−0.02,0.03]	0.04 [−0.02,0.09]	−0.003 [−0.05,0.06]	0.006 [−0.03,0.02]	0.038 [0.01,0.1]	−0.054 [−0.14,0.01]
Older Siblings	0.004 [−0.04,0.04]	−0.003 [−0.09,0.07]	0.021 [−0.12,0.11]	0.011 [−0.02,0.04]	0.028 [−0.04,0.09]	0.131 [−0.03,0.29]
Number of Children	−0.038 [−0.07,0]	−0.029 [−0.1,0.05]	−0.181 [−0.27,−0.03]	−0.039 [−0.08,0]	−0.048 [−0.1,0.02]	−0.326 [−0.45,−0.13]
Income	0.317 [0.1,0.36]	0.511 [0.26,0.61]	0.683 [0.2,0.9]	0.25 [0.14,0.32]	0.541 [0.3,0.6]	1.387 [0.8,1.49]
Urban	−0.065 [−0.09,0.03]	−0.045 [−0.11,0.02]	0.016 [−0.1,0.17]	−0.065 [−0.09,−0.02]	−0.015 [−0.07,0.02]	−0.134 [−0.28,−0.01]
Hindu	0.014 [−0.01,0.05]	−0.016 [−0.1,0.05]	0.028 [−0.03,0.11]	0.017 [−0.01,0.04]	−0.009 [−0.08,0.04]	0.066 [−0.02,0.18]
Muslim	0.008 [−0.04,0.05]	−0.036 [−0.14,0.02]	0.041 [−0.09,0.12]	0.003 [−0.02,0.03]	−0.029 [−0.09,0.03]	0.101 [0.02,0.24]
Mother's Age	0.019 [−0.01,0.04]	0.059 [0.02,0.12]	−0.014 [−0.09,0.12]	0.017 [0,0.04]	−0.012 [−0.04,0.03]	0.074 [−0.01,0.14]
Scheduled Caste	0.005 [−0.05,0.03]	−0.042 [−0.1,0.03]	−0.106 [−0.19,0.02]	−0.001 [−0.03,0.02]	0.018 [−0.03,0.06]	−0.111 [−0.22,0.03]
Scheduled Tribe	0 [−0.07,0.03]	−0.066 [−0.11,0.03]	−0.146 [−0.25,−0.03]	0.004 [−0.03,0.03]	−0.006 [−0.06,0.03]	−0.045 [−0.15,0.07]
BC Caste	−0.014 [−0.08,0.01]	−0.034 [−0.09,0.06]	−0.129 [−0.29,0.01]	−0.016 [−0.05,0.01]	0.042 [−0.01,0.07]	−0.084 [−0.23,0.08]
Prices and Income (F Statistic)	18.08	36.42	45.66	25.89	72.41	54.35
Prices and Income (P-values)	0.002	0	0	0	0	0
Prices (F Statistic)	4.11	7.27	16.89	17.48	9.27	3.91
Prices (P-values)	0.39	0.12	0.002	0.001	0.054	0.41

p-value=0: p-value<0.0001

Confidence intervals obtained through 100 bootstraps samples

Table 5: Parental Investments

that clothes play a role in investment in girls age 8 with an elasticity of -0.057, but not for boys. Price of Mebendazol[6] also only plays negatively for girls, and we could say that investment in boys seem to be less likely to be sensitive to prices overall.

We control by older siblings as well as by overall number of children : older siblings seem to have no effect for both sexes but for a given number of older siblings, more children overall - more younger siblings - have a more negative influence on boys age 12 (with an elasticity of -0.326) than it has on girls (-0.181), although both seem to get less parental investment when their family gets bigger. This result that was to be expected as the household's income per capita gets lower as household size grows, but does not fit eldest son preference theories, although one might argue that there might be some endogeneity of number of younger siblings with respect to gender, undermining the negative impact for girls.

Income plays a significantly and increasingly positive role on investment for both genders, although at age 12 parental investments are much more dependent on income for boys than for girls: a household that gets richer by 10% will spend 6.83% more in their girl child's latent investment good, and 13.87% more on their boys's one: this more than one elasticity can mean that there is a reallocation of the household budget towards investment in boys when there is a positive income shock, similarly to a superior good, but not towards investment in girls.

Finally, we can see that being an urban household has no effect on investment in girls but a negative effect on investment in boys, and that mother's age has a positive effect on investment in girls age 8, and a zero effect on investment in boys. This is consistent with the idea that urban zones have less son preference (Jayachandran and Pande 2017), and that older women tend to be more empowered and to have more agency over the household's budget allocation, although this is just one possible explanation for this results. As to tribe or caste issues, the only striking result is that belonging to a Scheduled Tribe has a very strong negative impact on investment in girls at age 12. Our instruments' p-values show they are very strong when one includes income as well as prices, and much less when only prices are included, with only half of the p-values over the 5% threshold. This justifies using this specification, although as we can see in subsection A.3, the results are quite the same when taking income off investment both for the investment function itself and the skills production functions.

[6] Medecine against worms.

4.3 Simulations

4.3.1 Decomposition : disentangling the causes for differences in human capital between genders

Accounting for the differences I observe in parental investments, I now seek to explain these gap. To do so, I apply an Oaxaca (1973) and Blinder (1973) decomposition to the investment measures, resulting in this equation:

$$ln(\overline{I_m}) - ln(\overline{I_f}) = (\overline{ln(X_m)} - \overline{ln(X_f)})\beta_m + \overline{ln(X_f)}(\beta_m - \beta_f) \tag{10}$$

βs are the coefficients from the estimations of Equation 8. This decomposition reflects how investment can be different between genders either because of differentials in parents' valuation of their children's characteristics (cognition, health or region, number of brothers, etc.), or because of differences in children's endowments of said characteristics. Table 6 presents the coefficients we obtain:

Age	$ln(\overline{I_m}) - ln(\overline{I_f})$	$(\overline{ln(X_m)} - \overline{ln(X_f)})\beta_m$	$\overline{ln(X_f)}(\beta_m - \beta_f)$
age 5	-0.076%	0.019%	-0.095%
age 8	3.649%	-0.165%	3.814%
age 12	7.356%	8.74%	-1.384%

Table 6: Decompositions

The first column illustrates the fact that the investment factor grows faster for boys than for girls. The baseline normalization and scaling on clothing expenditure means that our age 5 factors have the same scale and location (as is apparent with the near-zero gaps), and from this, what we can observe is due to different growth paths in factors across genders. At age 8, boys receive 3.649% of their investment factors relative to girls, and they receive 7.36% more of it at 12. The first gap can be attributed to boys getting more returns β on their characteristics $\overline{ln(X)}$ (third column); while their actual characteristics actually grew a bit slowlier than girls'. The second gap is due to an actual gap in characteristics, that could be attributed to early differences in investments at a time where parental investments are very important for skills production. This is an important result as it shows the way that early gaps in investments can set children on very separate skills growth paths.

4.3.2 Simulation of changes in determinants of parental investments

Wishing to explore further the consequences of these investment gaps for cognition and health, which in turn determine future investments, I also use these insights to construct two counterfactuals and study these dynamics more thoroughly. Namely, I want to:

1. Simulate that parents react to their girls' characteristics as if they were boys, setting β_f to the value of β_m.

2. Simulate that girls share baseline characteristics with boys but that parents react in a gendered manner, on the opposite, which is equivalent to having $\overline{ln(X_m)} = \overline{ln(X_f)}$.

For the first simulation, I first replace the coefficients from girls' investment functions with the boys' coefficients. At age 5, I thus obtain the simulated investment levels that they would have obtained were their cognition and health and controls valued the same as boys - from Table 6, they would have obtained 0.997 investment rather than 0.999, for example. I then plug this simulated value into their health and cognition production functions to obtain the factors that they would have had if they had obtained that level of investment - so here, at age 5, they would have 1.195 of cognition rather than 1.199 and 0. 859 of health rather than 0. 861. I plug this value into the investment function for age 8, and so on, using boys' rather than girls' estimated investment function coefficients, and so on.

The second set of results, boys and girls' coefficients if they had the same baseline characteristics, are obtained using the girls' investment functions and skill production function coefficients on the boys draws for age 5, as it is equivalent to having girls with boys characteristics at baseline. As with the first simulation, I first determine investment at age 5, which I plug into into skills at 5 to finally determine investment at age 8, and so on.

Before turning to the results, I want to stress some caveats that make this back-of the enveloppe calculation suggestive insights and not my main results. First, I have chosen to use the estimates for the alternative specification without control function (cf. subsection A.3) because given factors after age 5 are predicted and not estimated in our simulation, making errors from the investment function not replicable and not useable to estimate investiment residuals as in Table 3 or Table 4. This version is the one that makes for the best prediction of original values of the set from other specifications that I tested for (giving the control function the value of the difference between the initial and the predicted investment values, for example). Finally, giving each individual the value of its parental investment error in the original set would have been an assumption that girls are subjected to the same parental investment and skill production shocks than boys which is an assumption we did not want to make with regards to the rest of the paper's result. To keep the results consistent, I compare the prediction that I made not with the actual factors, but with a set of factors that I predict from the specification without control function as well. Also, these simulations do not take into account the bootstrapped confidence intervals, as we just use the estimated coefficients: this is a worry in case of big, but imprecisely estimated such as the impact of girls' cognition on parental investments. As we are applying this method to the two sets we compare, this is not a problem for the direct comparison in item 4.3.2, but rather something to be taken into account for comparison with the other results that we obtain with other specifications.

The first set of simulation results, where we assume that parents react to their girls' skills as if they were boys, shows that girls would have 5,2% more investment and 0,6% more cognition at age 8 (12% more investment and 0,4% more cognition at age 12) a result of these increased inputs, if parents treated them as boys. The second results from the simulation of equal baseline characteristics yields an estimation that is very similar to the true characteristics, suggesting that baseline skills gaps are not driving the results. One can also say that although our results are small, they could be a lower bound for the true mechanisms for two reasons: we did not quantify time investment which, in the literature, is said to be gender-biased, and also because gender could change characteristics in other ways not accounted for here, such as through income.

	cog5	cog8	cog12	health5	health8	health12	invest5	invest8	invest12
1st simulation - mean									
simulation	1.195	1.210	1.222	0.859	3.460	3.108	0.997	1.418	2.784
original data	1.199	1.202	1.217	0.861	3.476	3.121	0.999	1.340	2.467
1st simulation - median									
simulation	1.151	1.184	1.179	0.697	3.267	2.898	0.996	1.418	2.787
original data	1.152	1.164	1.162	0.697	3.204	2.854	0.997	1.329	2.460
2nd simulation - means									
simulation	1.202	1.203	1.22	0.878	3.504	3.141	0.999	1.34	2.467
original data	1.199	1.202	1.217	0.861	3.476	3.121	0.999	1.340	2.467
2nd simulation - medians									
simulation	1.136	1.146	1.141	0.695	3.147	2.82	0.995	1.32	2.459
original data	1.152	1.164	1.162	0.697	3.204	2.854	0.997	1.329	2.460

Table 7: Simulations

5 Discussion and conclusion

In this paper, I first estimate a dynamic model of cognition and health production, allowing for gender-dependent skills and investment factors. My results point to similar skill production patterns across genders, contrasting with start differences in patterns of parental investments that are elastic to boys' skills but do not react to girls' skills. Decomposing parental investmnents into differences caused by skills and by valuation of said skills, I find that early gender gaps in parents' valuation of children's characteristics tend to fuel later gaps in actual skills and investment levels. My results could be summarized in three points: first, children's human capital is produced the same across genders: different investment timings across genders, or different needs, cannot in this context be a justification for differential investments. Second, parental investment is crucial, especially at early ages but throughout time for cognition, which highlights potential scarring effect from its scarcity or from is non-reaction to children's skills, as can be seen with girls in my results. Finally and in relation with other dynamic models, my results also constitute rejection of the assumption that gender is a mere control, or that cognition and health can be computed in a gender neutral way: works interested in gender in human capital formation should consider allowing for gender-specific dynamics, both while computing latent factors and from estimating function coefficients. As my decomposition and naive simulations show, early efforts in making parental investment more gender neutral could yield sizeable gains, both for households and at the national levels.

Avenues forward include the inclusion of latter rounds or older cohorts from the Young Lives database: investment is biased at boys' advantages, but is it because of families' rational calculus with respect to the labor market ? For this, we must link individuals' inputs with their occupational outputs. Another potentially interesting avenue could be the replication of the analyses for other countries in the Young Lives database, to check for the persistence of these dynamics across countries. Finally, integrating time investments into this gender-split dynamic model of human capital accumulation could be a fruitful avenue for future research, given the even bigger gaps documented - for example by Barcellos et al. (2014) - for this kind of investments.

References

Agostinelli, F. and M. Wiswall, "Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development," *NBER Working Paper 22441*, 2016.

Almond, D. and J. Currie, "Human Capital Development Before Age Five," *NBER*, 2009.

Attanasio, O., C. Meghir, and E. Nix, "Human Capital Development and Parental Investment in India," *Review of Economic Studies*, 2020, (87), 2511–2541.

_ , R. Bernal, M. Giannola, and M. Nores, "Child development in the early years : Parental investment and the changing dynamics of different dimensions," *NBER Working Paper 27812*, 2020.

Barcellos, S.H., L.S. Carvalho, and A. Lleras-Muney, "Child Gender and Parental Investments in India : Are Boys and Girls Treated Differently ?," *American Economic Journal: Applied Economics*, 2014, (6), 157–189.

Barro, R.J. and J.W. Lee, "A new data set of educational attainment in the world," *Journal of Development Economcis*, 2013.

Berhman, J., J. Hervé, R. Laxminarayan, S. Mani, A. Nandi, and A. Sankhil Lamkang, "Gender Gaps in Cognitive and Noncognitive Skills : Roles of SES and Gender Attitudes," *PIER Working Paper 21-007*, 2021.

Bhalotra, S. and T. Cochrane, "Where Have all the Young Girls Gone? Identification of Sex Selection in India," *IZA Discussion Paper No. 5381*, 2010.

Blinder, A., "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 1973, *8*, 436–465.

Borghans, L., B.H.H. Golsteyn, J. Heckman, and J.E. Humphries, "What grades and achievement tests measure," *PNAS*, 2016, *113* (47).

Buchmann, C., T. Diprete, and A. McDaniel, "Gender Inequalities in Education," *ISERP Working Paper 07-15*, 2007.

Conti, G., J. Heckman, and S. Urzua, "The Education-Health Gradient," *American Economic Review*, 2010, (100), 234–238.

Cunha, F., I. Elo, and J. Culhane, "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation," *NBER Working Paper 19144*, 2013.

_, J. Heckman, and S. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *NBER Working Paper 15664*, 2010.

Jayachandran, S. and A. Lleras-Muney, "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines," *The Quaterly Journal of Economics*, 2009.

_ and I. Kuziemko, "Why Do Mothers Breastfeed Firls Less Than Boys: Evidence and Implications for Child Health in India," *Quaterly journal of Economics*, 2011, (126), 1485–1538.

_ and R. Pande, "Why Are Indian Children So Short? The Role of Birth Order and Son Preference," *American Economic Review*, 2017, (107(9)), 2600–2629.

Oaxaca, R., "Male-female differentials in urban labor markets," *International Economic Review*, 1973, *14*, 673–709.

Sen, A., "More than 100 Million Women Are Missing," *The New York Review of Books*, 1990.

Sommerfelt, A.E. and F. Arnold, "Sex differentials in nutritional status of young children," *New York : Population Division, United Nations*, 1998.

A Appendix

A.1 Variables changes with respect to Attanasio et al. (2020a)

My model is drawn from Attanasio et al. (2020a), and for comparability purposes, I use the authors' database. However, I made some changes to the variables computation:

1. I use percentage of correct answers to the PPVT rather than PPVT score, for reasons of test changes between rounds 3 and 4 that I detail in the data section
2. I scale variables by standard deviation rather than by root mean square, as R automatically scales by RMS for non-centered variables.
3. Finally, I use clothing expenses rather than books expenses for the parental investment anchor, as books expenses had a very bad signal to noise ratio for girls aged 8

The pooled model estimation is very close to results from Attanasio et al. (2020a), providing reassurance that our results are indeed comparable with theirs.

		Cognition			*Health*	
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	0.53 [0.45,0.62]	0.86 [0.77,0.91]	NA [NA,NA]	−0.11 [−0.19,−0.02]	−0.01 [−0.05,0.07]
Health	0.02 [0,0.04]	0.05 [0.02,0.08]	0.01 [−0.01,0.03]	0.47 [0.43,0.51]	0.9 [0.81,0.94]	0.91 [0.86,0.96]
Parental Cognition	0 [−0.05,0.05]	−0.01 [−0.04,0.02]	0.03 [0.02,0.06]	0 [−0.05,0.03]	0.06 [0.03,0.09]	0 [−0.04,0.03]
Parental Health	−0.02 [−0.07,0.01]	−0.03 [−0.05,0.01]	0.01 [−0.01,0.04]	0.24 [0.18,0.33]	0.05 [0.01,0.09]	0.06 [0.01,0.12]
Investment	1 [0.95,1.07]	0.46 [0.36,0.56]	0.09 [0.01,0.2]	0.29 [0.22,0.36]	0.11 [0.03,0.2]	0.04 [−0.05,0.09]
Complementarity	−0.16 [−0.39,0.44]	0.03 [−0.19,0.31]	−0.28 [−0.49,−0.11]	0 [−0.08,0.05]	0.19 [0.05,0.43]	−0.03 [−0.14,0.3]
Elasticity of Subst.	0.86 [0.72,1.77]	1.03 [0.84,1.45]	0.78 [0.67,0.9]	1 [0.93,1.05]	1.24 [1.05,1.75]	0.97 [0.88,1.43]
Log TFP	−0.02 [−0.04,0.01]	0.42 [0.37,0.45]	1.59 [1.52,1.66]	−0.34 [−0.39,−0.27]	0.11 [0.06,0.16]	−0.06 [−0.13,0.03]
Investment Residual	−0.91 [−1.01,−0.79]	−0.48 [−0.59,−0.33]	−0.1 [−0.22,−0.03]	−0.65 [−0.9,−0.45]	−0.12 [−0.23,0.01]	−0.02 [−0.09,0.08]
Number of Children	0 [0,0.01]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0.01 [0,0.03]	0 [−0.02,0.01]	0 [−0.01,0.01]
Older Siblings	0 [0,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.03 [−0.04,−0.01]	0 [−0.01,0.01]	0.01 [0,0.02]
Gender	0 [−0.01,0.01]	0.02 [0.01,0.02]	−0.01 [−0.01,0]	0 [−0.01,0.01]	0.01 [0,0.02]	0.02 [0,0.02]
Urban	0 [0,0.01]	−0.01 [−0.01,0]	−0.01 [−0.01,0]	0 [−0.01,0.01]	0.01 [0,0.01]	0 [−0.01,0.01]
Hindu	−0.01 [−0.02,−0.01]	−0.01 [−0.01,0]	0.02 [0.01,0.02]	0.01 [−0.01,0.02]	0 [−0.01,0]	0 [−0.01,0.01]
Muslim	0 [0,0]	0 [0,0]	0 [−0.01,0]	0 [0,0]	0 [0,0]	0 [0,0]
Mother's Age	0 [−0.01,0]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.02]	0 [−0.01,0.01]	−0.02 [−0.03,−0.01]
Scheduled Caste	0 [−0.01,0]	0.01 [0,0.02]	0 [0,0.01]	0 [−0.01,0.01]	0 [−0.01,0]	0 [−0.01,0.01]
Scheduled Tribe	0.03 [0.02,0.03]	−0.01 [−0.01,0]	0 [0,0.01]	0.01 [0.01,0.02]	−0.01 [−0.02,−0.01]	0 [0,0.01]
BC Caste	0 [−0.01,0]	0 [0,0.01]	0.01 [0,0.01]	−0.02 [−0.03,0]	0.01 [0,0.02]	0 [−0.01,0]

Results come from the estimation of equation 8 on 10 000 simulated individuals from the pooled sample
Confidence intervals are obtained through 100 bootstraps samples

Table 8: Skills production functions - pooled sample

	Age 5	Age 8	Age 12
Cognition	NA [NA,NA]	0.293 [0.11,0.44]	0.785 [0.21,1.14]
Health	0.028 [0.01,0.05]	−0.029 [−0.05,0.01]	0.059 [−0.04,0.17]
Parental Cognition	0.019 [−0.02,0.05]	−0.036 [−0.07,0.05]	−0.089 [−0.16,0.03]
Parental Health	0.001 [−0.02,0.03]	−0.028 [−0.08,0.02]	−0.059 [−0.17,0.04]
Price Clothes	0.028 [0.01,0.05]	0.007 [−0.04,0.03]	−0.024 [−0.1,0.04]
Price Notebook	0.008 [−0.01,0.02]	−0.043 [−0.08,−0.01]	0.092 [0.03,0.16]
Price Mebendazol	0.013 [0,0.03]	−0.03 [−0.07,0]	−0.029 [−0.08,0.02]
Price Food	0.005 [−0.01,0.02]	0.021 [−0.01,0.07]	−0.018 [−0.08,0.03]
Older Siblings	0.006 [−0.02,0.03]	−0.002 [−0.04,0.06]	0.067 [−0.05,0.15]
Number of Children	−0.037 [−0.05,−0.01]	−0.026 [−0.08,0.02]	−0.253 [−0.34,−0.13]
Gender	0 [−0.01,0.02]	−0.007 [−0.05,0.02]	−0.065 [−0.12,0.01]
Income	0.295 [0.22,0.36]	0.613 [0.46,0.68]	0.963 [0.64,1.11]
Urban	−0.06 [−0.07,−0.03]	−0.048 [−0.07,0]	−0.041 [−0.13,0.05]
Hindu	0.022 [0,0.04]	−0.028 [−0.08,0.03]	0.074 [0.01,0.14]
Muslim	0.012 [−0.02,0.03]	−0.048 [−0.09,0]	0.099 [0.01,0.17]
Mother's Age	0.017 [0,0.03]	0.017 [−0.01,0.04]	0.039 [0,0.11]
Scheduled Caste	−0.002 [−0.03,0.02]	−0.012 [−0.05,0.03]	−0.087 [−0.16,−0.01]
Scheduled Tribe	0.003 [−0.02,0.02]	−0.026 [−0.07,0.01]	−0.1 [−0.19,−0.04]
BC Caste	−0.022 [−0.05,0]	0.016 [−0.04,0.06]	−0.084 [−0.2,−0.02]

Confidence intervalsobtained through 100 bootstraps samples.

Table 9: Investment - pooled sample

A.2 Factors

A.2.1 Factor densities

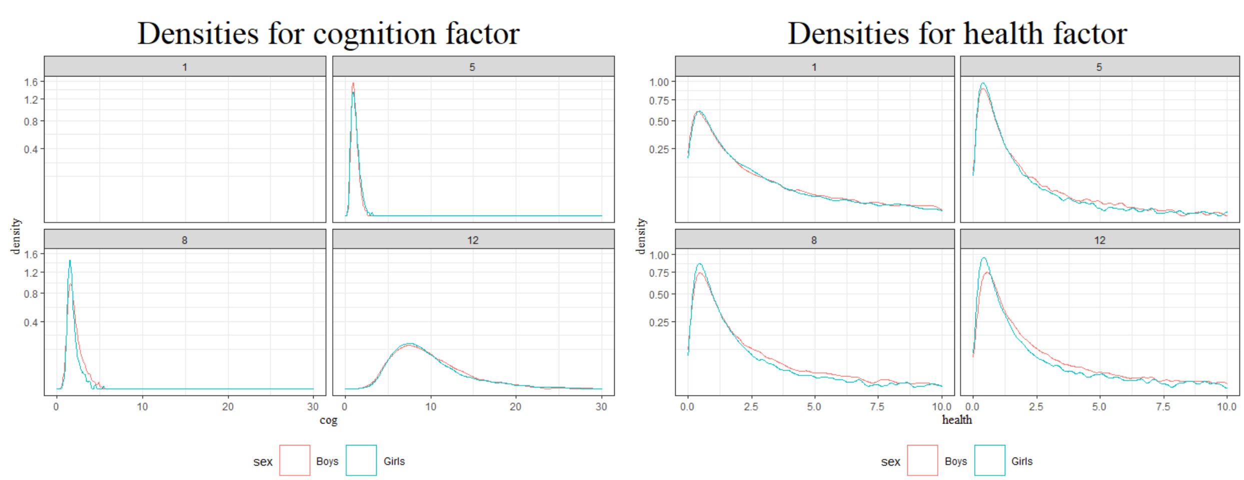

Figure 1: Factor densities across genders

This figures shows that at age 5, factors are always very comparable, as they are normalized and scaled on the same measures. When children get older, the factors evolve differently to represent different growth speeds.

A.2.2 Signal to noise for split samples

Measure-Cognition	Signal	Measure-Health	Signal
Age 12 PPVT (pc)	0.26	Age 12 Height Z-Score	0.61
Age 12 English	0.66	Age 12 Weight Z-Score	0.61
Age 12 Math	0.73	Age 8 Height Z-Score	0.56
Age 12 Language	0.5	Age 8 Weight Z-Score	0.74
Age 8 PPVT (pc)	0.12	Age 8 Health Status	0.04
Age 8 Math	0.34	Age 5 Height Z-Score	0.74
Age 8 EGRA	0.48	Age 5 Weight Z-Score	0.74
Age 5 PPVT (pc)	0.56	Age 5 Health Status	0
Age 5 CDA	0.4	Age 1 Height Z-Score	0.55
Younger Mom Education	0.79	Age 1 Weight Z-Score	0.79
Younger Dad Education	0.6	Age 1 Health Status	0.06
Literacy	0.4	Younger Mom Weight	0.58
NA		Younger Mom Height	0.13

Table 10: signal to noise - skill measure - girls sample

Measure-Investment	Signal
Age 12 Amount spent on clothing	0.27
Age 12 amount spent on books	0.16
Age 12 Amount spend on shoes	0.36
Age 12 Amount spent on uniform	0.14
Age 12 Meals in a day	0.01
Age 12 Food groups in a day	0.01
Age 8 Amount spent on clothing	0.42
Age 8 Amount spend on shoes	0.48
Age 8 Amount spent on uniform	0.1
Age 8 Meals in a day	0.03
Age 8 Food groups in a day	0.04
Age 5 Amount spent on clothing	0.46
Age 12 amount spent on books	0.2
Age 5 Amount spend on shoes	0.46
Age 5 Amount spent on uniform	0.2
Age 5 Meals in a day	0.07
Age 5 Food groups in a day	0.2

Table 11: signal to noise - investment measures - girls sample

Measure-Cognition	Signal	Measure-Health	Signal
Age 12 PPVT (pc)	0.4	Age 12 Height Z-Score	0.66
Age 12 English	0.73	Age 12 Weight Z-Score	0.69
Age 12 Math	0.67	Age 8 Height Z-Score	0.66
Age 12 Language	0.4	Age 8 Weight Z-Score	0.77
Age 8 PPVT (pc)	0.22	Age 8 Health Status	0.04
Age 8 Math	0.33	Age 5 Height Z-Score	0.73
Age 8 EGRA	0.64	Age 5 Weight Z-Score	0.7
Age 5 PPVT (pc)	0.48	Age 5 Health Status	0.01
Age 5 CDA	0.42	Age 1 Height Z-Score	0.53
Younger Mom Education	0.8	Age 1 Weight Z-Score	0.82
Younger Dad Education	0.54	Age 1 Health Status	0.07
Literacy	0.47	Younger Mom Weight	0.63
NA		Younger Mom Height	0.13

Table 12: signal to noise - skill measures - boys sample

Measure-Investment	Signal
Age 12 Amount spent on clothing	0.58
Age 12 amount spent on books	0.35
Age 12 Amount spend on shoes	0.21
Age 12 Amount spent on uniform	0.25
Age 12 Meals in a day	0.04
Age 12 Food groups in a day	0.01
Age 8 Amount spent on clothing	0.41
Age 8 Amount spend on shoes	0.38
Age 8 Amount spent on uniform	0.19
Age 8 Meals in a day	0.23
Age 8 Food groups in a day	0.18
Age 5 Amount spent on clothing	0.52
Age 12 amount spent on books	0.27
Age 5 Amount spend on shoes	0.44
Age 5 Amount spent on uniform	0.19
Age 5 Meals in a day	0.04
Age 5 Food groups in a day	0.26

Table 13: signal to noise - investment measures - boys sample

Here are the signal to noise tables for the split samples. The signal tends to be weaker than for pooled samples, which is normal since noise is a measure of variance and the sample has been split in two, but the measures are still quite high and there are no very low points different to the minima in the pooled sample tables now that we have dropped one problematic measure (book expenses at age 8). We can see that the signal for girls is weaker than that for boys, perhaps due to the difference in sample sizes.

A.2.3 Mixture weights, means, and factor distributions

	Girls		*Boys*	
	Mixture A	Mixture B	Mixture A	Mixture B
Weights	0.774 [0.76,0.99]	0.226 [0.01,0.24]	0.526 [0.5,0.56]	0.474 [0.44,0.5]
Mean Cognition Age 12	2.044 [1.93,2.1]	2.244 [1.97,2.27]	2.031 [2,2.07]	2.2 [2.17,2.23]
Mean Cognition Age 8	0.423 [0.39,0.49]	0.618 [0.39,0.65]	0.467 [0.44,0.5]	0.663 [0.62,0.7]
Mean Cognition Age 5	-0.053 [-0.06,0]	0.182 [0.03,0.22]	-0.067 [-0.1,-0.05]	0.075 [0.05,0.1]
Mean Health Age 12	-0.473 [-0.52,-0.26]	0.278 [-0.08,0.61]	-0.326 [-0.39,-0.27]	0.26 [0.18,0.33]
Mean Health Age 8	-0.373 [-0.44,-0.2]	0.24 [-0.13,0.47]	-0.399 [-0.47,-0.32]	0.236 [0.16,0.3]
Mean Health Age 5	-0.516 [-0.56,-0.35]	-0.009 [-0.34,0.15]	-0.541 [-0.61,-0.47]	-0.073 [-0.15,0]
Mean Health Age 1	-0.105 [-0.13,0]	0.36 [-0.05,0.47]	-0.228 [-0.28,-0.15]	0.253 [0.18,0.33]
Mean Investment Age 12	0.708 [0.67,1.02]	1.856 [1.08,1.99]	0.581 [0.51,0.7]	1.571 [1.34,1.73]
Mean Investment Age 8	0.109 [0.08,0.36]	0.777 [0.28,0.83]	0.133 [0.1,0.19]	0.488 [0.42,0.54]
Mean Investment Age 5	-0.08 [-0.12,0]	0.274 [-0.01,0.4]	-0.112 [-0.15,-0.07]	0.124 [0.08,0.17]
Mean Parental Cognition	-0.218 [-0.25,0]	0.745 [0.22,0.87]	-0.487 [-0.54,-0.41]	0.54 [0.48,0.61]
Mean Parental Health	-0.198 [-0.23,0]	0.678 [0.27,1.07]	-0.293 [-0.35,-0.22]	0.325 [0.25,0.42]
Mean Income Age 12	-0.184 [-0.22,0]	0.631 [0.09,0.71]	-0.315 [-0.36,-0.23]	0.35 [0.3,0.42]
Mean Income Age 8	-0.175 [-0.21,0]	0.597 [0.19,0.74]	-0.338 [-0.39,-0.26]	0.375 [0.3,0.45]
Mean Income Age 5	-0.21 [-0.24,0]	0.716 [0.27,0.83]	-0.356 [-0.41,-0.3]	0.394 [0.34,0.47]

Table 14: Mixture weights and means by gender

We also have the mixture weights by genders, which point to a more important departure from normality from girls. Other coefficients seem quite similar, with a bit less cognition age 8. We plot the distribution of the latent factors to show that the factor densities are overlapping a lot. Note that, for investment at age 12, trimmed the 100th quantile for readability purposes, some individuals being located at more than 1000 while the mean is around 5. We did not trim outliers in our estimations, and there are outliers in both samples.

A.3 Alternative specifications for skill production functions

A.3.1 Skills with income both in investment and production functions

	Cognition			*Health*		
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	0.24 [0.15,0.48]	0.74 [0.57,0.88]	NA [NA,NA]	−0.03 [−0.17,0.48]	0.07 [−0.09,0.88]
Health	0.02 [0,0.06]	0.04 [0,0.09]	0.01 [−0.07,0.04]	0.48 [0.39,0.58]	0.91 [0.8,0.09]	0.9 [0.79,0.04]
Parental Cognition	0.07 [−0.02,0.13]	−0.02 [−0.06,0.08]	0.05 [−0.03,0.09]	−0.03 [−0.09,0.06]	0.06 [0.03,0.08]	−0.04 [−0.07,0.09]
Parental Health	0.01 [−0.05,0.07]	−0.05 [−0.12,0.01]	0.01 [−0.02,0.09]	0.17 [0.07,0.37]	0.01 [−0.04,0.01]	0.14 [0.05,0.09]
Investment	0.71 [0.29,1.23]	0.93 [0.61,1.17]	1.01 [0.4,1.09]	0.88 [0.65,0.99]	−0.21 [−0.35,1.17]	−0.08 [−0.26,1.09]
Complementarity	0.21 [−0.2,0.67]	0.1 [−0.14,0.4]	−0.08 [−0.13,0.53]	−0.02 [−0.08,0.07]	0.23 [−0.08,0.4]	0.08 [−0.09,0.53]
Elasticity of Subst.	1.27 [0.83,2.95]	1.11 [0.88,1.66]	0.93 [0.88,2.13]	0.98 [0.93,1.07]	1.3 [0.92,1.66]	1.08 [0.92,2.13]
Log TFP	0.11 [−0.01,0.15]	0.53 [0.34,0.59]	1.63 [1.44,1.75]	−0.42 [−0.52,−0.28]	0.09 [−0.08,0.16]	−0.1 [−0.22,0.08]
Investment Residual	−0.77 [−0.96,−0.5]	−0.74 [−0.88,−0.53]	−0.2 [−0.4,−0.1]	−0.77 [−1.49,−0.41]	−0.02 [−0.27,0.17]	0.08 [−0.07,0.37]
Number of Children	0 [−0.01,0.01]	0 [−0.01,0.02]	0 [−0.02,0.01]	0.01 [−0.02,0.03]	−0.01 [−0.02,0.01]	−0.01 [−0.03,0.01]
Older Siblings	0.01 [−0.01,0.01]	0 [−0.02,0.01]	0 [−0.01,0.02]	−0.03 [−0.06,0]	0.02 [0,0.03]	0.01 [−0.01,0.03]
Income	−0.11 [−0.15,−0.02]	−0.21 [−0.25,−0.08]	−0.05 [−0.13,0]	0.04 [−0.01,0.11]	0.01 [−0.04,0.11]	0.09 [0.01,0.15]
Urban	0.01 [0,0.02]	0 [−0.01,0.02]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]
Hindu	−0.01 [−0.03,0]	−0.01 [−0.02,0.01]	0.01 [−0.01,0.02]	0 [−0.03,0.02]	0 [−0.01,0.02]	−0.01 [−0.03,0.02]
Muslim	0 [−0.01,0.01]	0 [0,0.01]	0 [−0.03,0]	0 [−0.02,0.01]	0 [0,0.01]	0 [−0.01,0.01]
Mother's Age	0 [−0.02,0.01]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.02]	0 [−0.02,0.02]	−0.02 [−0.04,0]
Scheduled Caste	−0.01 [−0.02,0]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.01,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0.02 [0.02,0.04]	0 [−0.02,0]	0 [−0.01,0.01]	0.01 [−0.01,0.02]	−0.01 [−0.02,0]	0 [−0.01,0.01]
BC Caste	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.02 [−0.03,0]	0.01 [0,0.03]	−0.01 [−0.02,0]

Confidence intervals are obtained through 100 bootstraps samples by gender

Table 15: Girls skills with income in both production and investment

	Cognition			*Health*		
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	0.36 [0.18,0.47]	0.83 [0.64,0.95]	NA [NA,NA]	−0.1 [−0.25,0.47]	0.06 [−0.02,0.95]
Health	0.02 [−0.01,0.04]	0.05 [0.02,0.09]	0 [−0.03,0.05]	0.48 [0.42,0.53]	0.86 [0.74,0.09]	0.87 [0.79,0.05]
Parental Cognition	0.06 [0.02,0.1]	0.06 [0.02,0.08]	0.04 [0.02,0.08]	0 [−0.07,0.07]	0.06 [0.02,0.08]	−0.02 [−0.06,0.08]
Parental Health	−0.01 [−0.07,0.04]	0.02 [−0.05,0.06]	0.02 [−0.02,0.08]	0.32 [0.2,0.43]	0.1 [0.05,0.06]	0.01 [−0.05,0.08]
Investment	0.71 [0.28,1.19]	0.51 [0.35,0.81]	1.13 [0.76,1.33]	0.67 [0.56,0.83]	−0.22 [−0.33,0.81]	0.2 [−0.32,1.33]
Complementarity	0.24 [−0.19,0.65]	0.36 [0.1,0.51]	−0.19 [−0.37,0.14]	0.01 [−0.07,0.07]	0.2 [−0.04,0.51]	−0.06 [−0.15,0.14]
Elasticity of Subst.	1.32 [0.83,2.77]	1.56 [1.11,2.03]	0.84 [0.73,1.16]	1.01 [0.93,1.08]	1.26 [0.96,2.03]	0.94 [0.87,1.16]
Log TFP	0.07 [0,0.13]	0.53 [0.44,0.58]	1.61 [1.54,1.69]	−0.31 [−0.41,−0.21]	0.11 [0.01,0.19]	−0.07 [−0.18,0.03]
Investment Residual	−0.92 [−1.08,−0.71]	−0.56 [−0.77,−0.4]	−0.14 [−0.31,0.04]	−0.47 [−0.86,−0.09]	−0.11 [−0.32,0.11]	−0.06 [−0.14,0.06]
Number of Children	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0.02 [0,0.04]	0 [−0.01,0.02]	0 [−0.01,0.02]
Older Siblings	0 [−0.01,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.01]	−0.02 [−0.04,0]	−0.01 [−0.02,0]	0.01 [−0.01,0.02]
Income	−0.08 [−0.12,−0.02]	−0.09 [−0.14,−0.02]	−0.08 [−0.16,0.03]	−0.02 [−0.07,0.06]	0.01 [−0.05,0.07]	0.01 [−0.04,0.1]
Urban	0.01 [0,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.01,0]
Hindu	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.02 [0.01,0.03]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]
Muslim	0 [0,0]	0 [0,0]	0 [−0.01,0]	0 [0,0]	0 [0,0]	0 [0,0]
Mother's Age	0 [−0.01,0.01]	0.01 [0,0.01]	−0.01 [−0.02,0]	0 [−0.02,0.01]	0 [−0.02,0.02]	−0.02 [−0.03,0]
Scheduled Caste	0 [−0.02,0]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.02,0.01]	0 [−0.02,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0.02 [0.01,0.03]	0 [−0.01,0.01]	0 [0,0.01]	0.02 [0.01,0.03]	−0.01 [−0.02,0]	0 [0,0.01]
BC Caste	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.01 [0,0.02]	−0.01 [−0.03,0]	0.01 [0,0.02]	0 [−0.01,0.01]

Confidence intervals obtained through 100 bootstraps samples by gender

Table 16: Boys' skills with income in both production and investment

When adding income into the equation, the coefficients have the same signs as the main production function estimates and their confidence intervals overlap. Income is only significantly positive for health at age 12 for girls and has the same signs as the main model and is slightly negative for cognition at two ages for girls and boys. One lead could be that given that parent's cognition is already controlled for, adding income might also proxy for them being away for work for longer, hence the negative coefficients.

A.3.2 Skills without a control function

	Cognition			*Health*		
Age	5	8	12	5	8	12
Cognition	NA $[NA,NA]$	0.71 [0.54,0.81]	0.89 [0.77,1.06]	NA $[NA,NA]$	−0.04 [−0.16,0.04]	−0.04 [−0.2,0.08]
Health	0.05 [0.01,0.09]	0.02 [0,0.1]	0.02 [−0.06,0.04]	0.5 [0.4,0.6]	0.9 [0.8,1.05]	0.9 [0.79,1.05]
Parental Cognition	0.09 [0,0.13]	0 [−0.01,0.11]	0.07 [−0.05,0.09]	0.06 [−0.01,0.12]	0.07 [0.03,0.13]	−0.01 [−0.05,0.06]
Parental Health	0.08 [−0.01,0.13]	0.01 [−0.05,0.05]	0.03 [0.01,0.1]	0.25 [0.11,0.43]	0.01 [−0.04,0.09]	0.14 [0.05,0.22]
Investment	1.09 [0.68,1.21]	1.92 [0.74,2.02]	0.96 [0.3,1.26]	0.76 [0.52,0.88]	−0.21 [−0.36,0.11]	−0.12 [−0.29,0.13]
Complementarity	−0.26 [−0.4,0.16]	−0.94 [−1.03,0.12]	−0.07 [−0.34,0.67]	−0.07 [−0.1,0.07]	0.23 [−0.08,0.36]	0.09 [−0.14,0.26]
Elasticity of Subst.	0.79 [0.72,1.19]	0.51 [0.49,1.14]	0.94 [0.7,2.08]	0.93 [0.91,1.08]	1.3 [0.93,1.57]	1.1 [0.87,1.36]
Log TFP	0 [−0.07,0.02]	0.46 [0.35,0.5]	1.7 [1.47,1.76]	−0.37 [−0.45,−0.26]	0.1 [0.01,0.19]	−0.05 [−0.15,0.08]
Investment Residual	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]
Number of Children	0.01 [−0.01,0.02]	0 [−0.01,0.02]	−0.01 [−0.03,0]	0.02 [−0.02,0.04]	−0.01 [−0.02,0]	−0.01 [−0.03,0.01]
Older Siblings	0.01 [−0.01,0.02]	−0.01 [−0.02,0.01]	0 [−0.01,0.02]	−0.03 [−0.06,0]	0.02 [0,0.03]	0.01 [−0.01,0.03]
Urban	−0.01 [−0.02,0]	−0.01 [−0.02,0]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]
Hindu	−0.01 [−0.03,0]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.02,0.02]	0 [−0.01,0.02]	0 [−0.03,0.02]
Muslim	0 [−0.01,0]	0 [−0.01,0]	0 [−0.03,0]	0 [−0.02,0.01]	0 [0,0.01]	0 [−0.01,0.01]
Mother's Age	−0.01 [−0.02,0.01]	0.01 [0,0.01]	−0.01 [−0.02,0]	0 [−0.02,0.02]	0 [−0.02,0.01]	−0.02 [−0.03,0]
Scheduled Caste	0 [−0.01,0]	0.01 [0.01,0.02]	0 [−0.01,0.01]	0.01 [−0.01,0.02]	−0.01 [−0.01,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0.03 [0.02,0.04]	−0.01 [−0.03,−0.01]	0 [−0.01,0.01]	0.01 [−0.01,0.02]	−0.01 [−0.02,0]	0 [−0.01,0.01]
BC Caste	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.02 [−0.03,0]	0.01 [0,0.02]	−0.01 [−0.02,0]

Confidence intervals are obtained through 100 bootstraps samples by gender

Table 17: Girls' skills without control function

Taking off the error from the investment function from the equation, we see that parental health has a positive coefficient at age 12 for girls, which it has not in the main specification. Other than that, the coefficients are all very similar, which is surprising given that we should have expected that not accounting for safety net would place less importance on lagged cognition, which is not the case.

		Cognition			*Health*	
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	0.72 [0.58,0.77]	0.96 [0.87,0.96]	NA [NA,NA]	−0.06 [−0.16,0.02]	0.07 [0.01,0.14]
Health	0.03 [0,0.06]	0.05 [0.01,0.08]	0.01 [0,0.06]	0.49 [0.42,0.54]	0.85 [0.74,0.91]	0.89 [0.8,0.95]
Parental Cognition	0.06 [0.03,0.1]	0.04 [0,0.07]	0.03 [0.02,0.08]	0.01 [−0.05,0.08]	0.07 [0.04,0.11]	0.01 [−0.03,0.06]
Parental Health	0.01 [−0.05,0.07]	0.02 [−0.02,0.07]	0 [−0.02,0.06]	0.33 [0.22,0.45]	0.11 [0.06,0.2]	0 [−0.04,0.06]
Investment	1.11 [0.5,1.22]	0.69 [0.43,1.07]	1.65 [0.54,1.85]	0.66 [0.54,0.79]	−0.26 [−0.35,0]	0.16 [−0.24,0.29]
Complementarity	−0.19 [−0.32,0.44]	0.21 [−0.16,0.47]	−0.69 [−0.89,0.37]	−0.01 [−0.07,0.06]	0.23 [−0.03,0.27]	−0.06 [−0.18,0.34]
Elasticity of Subst.	0.84 [0.76,1.73]	1.26 [0.86,1.89]	0.59 [0.53,1.58]	0.99 [0.93,1.07]	1.3 [0.97,1.38]	0.95 [0.85,1.51]
Log TFP	−0.02 [−0.05,0.02]	0.53 [0.47,0.58]	1.58 [1.55,1.65]	−0.33 [−0.41,−0.26]	0.13 [0.06,0.2]	−0.01 [−0.09,0.04]
Investment Residual	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]
Number of Children	0.01 [0,0.02]	0 [−0.01,0.01]	−0.01 [−0.02,0]	0.02 [0,0.04]	0 [−0.01,0.02]	0 [−0.01,0.02]
Older Siblings	0 [0,0.01]	−0.01 [−0.01,0]	0 [0,0.01]	−0.02 [−0.04,0]	−0.01 [−0.02,0]	0.01 [−0.01,0.02]
Urban	0 [−0.01,0.01]	−0.01 [−0.02,−0.01]	0 [−0.01,0]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.01,0]
Hindu	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.02 [0.01,0.03]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]
Muslim	0 [0,0]	0 [0,0]	0 [−0.01,0]	0 [0,0]	0 [0,0]	0 [0,0]
Mother's Age	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.01,0]	0 [−0.02,0.01]	0 [−0.01,0.02]	−0.02 [−0.03,0]
Scheduled Caste	0 [−0.01,0.01]	0.01 [0,0.02]	0.01 [0,0.02]	0 [−0.02,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0.02 [0.01,0.03]	−0.01 [−0.01,0]	0 [0,0.01]	0.02 [0.01,0.03]	−0.02 [−0.02,−0.01]	0 [−0.01,0.01]
BC Caste	−0.01 [−0.02,0]	0.01 [0,0.02]	0.01 [0,0.02]	−0.01 [−0.03,0]	0.01 [0,0.02]	0 [−0.01,0.01]

Confidence intervals obtained through 100 bootstraps samples by gender

Table 18: Boy's skills without control function

A.3.3 Income in neither production nor investment functions

	Cognition			*Health*		
Age	5	8	12	5	8	12
Cognition	NA [NA,NA]	0.32 [0.14,0.54]	0.81 [0.61,0.97]	NA [NA,NA]	0.05 [−0.11,0.17]	−0.01 [−0.15,0.19]
Health	0.03 [0,0.06]	0.03 [−0.01,0.09]	0.01 [−0.07,0.03]	0.48 [0.39,0.58]	0.89 [0.8,1.08]	0.9 [0.79,1.05]
Parental Cognition	−0.01 [−0.11,0.09]	−0.1 [−0.12,0.02]	0.04 [−0.09,0.07]	−0.01 [−0.05,0.08]	0.1 [0.05,0.18]	0 [−0.05,0.07]
Parental Health	−0.03 [−0.09,0.05]	−0.08 [−0.14,−0.01]	0.01 [−0.02,0.08]	0.18 [0.08,0.4]	0.04 [−0.02,0.11]	0.15 [0.05,0.23]
Investment	1.04 [0.49,1.27]	1.35 [0.88,1.45]	1.09 [0.49,1.17]	0.86 [0.62,0.96]	−0.12 [−0.3,0.08]	−0.11 [−0.29,0.12]
Complementarity	0 [−0.23,0.5]	−0.19 [−0.36,0.13]	−0.15 [−0.21,0.49]	−0.04 [−0.09,0.06]	0.09 [−0.07,0.31]	0.06 [−0.13,0.18]
Elasticity of Subst.	1 [0.81,2]	0.84 [0.73,1.15]	0.87 [0.82,1.96]	0.96 [0.92,1.07]	1.1 [0.94,1.46]	1.06 [0.88,1.22]
Log TFP	−0.01 [−0.06,0.03]	0.29 [0.19,0.38]	1.61 [1.32,1.72]	−0.37 [−0.45,−0.25]	0.14 [0.02,0.25]	−0.01 [−0.14,0.2]
Investment Residual	−0.86 [−1,−0.52]	−0.75 [−0.94,−0.45]	−0.13 [−0.4,−0.04]	−0.58 [−1.09,−0.16]	0.17 [−0.1,0.4]	0.06 [−0.07,0.38]
Number of Children	0 [−0.01,0.01]	0 [−0.02,0.01]	−0.01 [−0.02,0.01]	0.01 [−0.02,0.03]	−0.01 [−0.02,0.01]	−0.01 [−0.03,0.01]
Older Siblings	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.02]	−0.03 [−0.06,0]	0.02 [0,0.03]	0.01 [−0.01,0.03]
Urban	0 [−0.01,0.01]	−0.01 [−0.01,0.01]	−0.01 [−0.02,0]	0.01 [−0.01,0.02]	0 [−0.01,0.01]	0 [−0.01,0.02]
Hindu	−0.01 [−0.03,0]	−0.02 [−0.03,0]	0.01 [−0.01,0.02]	0 [−0.02,0.02]	0 [−0.01,0.02]	0 [−0.03,0.02]
Muslim	0 [−0.01,0]	0 [0,0.01]	0 [−0.03,0]	0 [−0.02,0.01]	0 [0,0.01]	0 [−0.01,0.01]
Mother's Age	−0.01 [−0.02,0.01]	−0.01 [−0.01,0.01]	−0.01 [−0.02,0]	0 [−0.01,0.02]	0 [−0.01,0.02]	−0.02 [−0.03,0]
Scheduled Caste	−0.01 [−0.02,0]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.01 [−0.01,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0.02 [0.02,0.04]	−0.01 [−0.02,0]	0 [−0.01,0.01]	0 [−0.01,0.02]	−0.01 [−0.02,0]	0 [−0.01,0.01]
BC Caste	0 [−0.01,0.01]	0 [−0.01,0.02]	0 [−0.01,0.01]	−0.02 [−0.03,0]	0.01 [0,0.02]	−0.01 [−0.02,0]

Confidence intervals are obtained through 100 bootstraps samples by gender

Table 19: Girls' skills without income in production or investment

Age		*Cognition* 8			*Health* 8	
	5	8	12	5	8	12
Cognition	NA $[NA,NA]$	0.25 [0.08,0.41]	1 [0.77,1.04]	NA $[NA,NA]$	−0.08 [−0.25,0.07]	0.1 [0.01,0.19]
Health	0.02 [−0.01,0.04]	0.06 [0.02,0.1]	0.07 [−0.01,0.14]	0.48 [0.42,0.53]	0.86 [0.75,0.91]	0.92 [0.79,0.96]
Parental Cognition	0.01 [−0.04,0.06]	0 [−0.06,0.05]	0.13 [0,0.2]	−0.01 [−0.06,0.06]	0.06 [0.04,0.11]	0.04 [−0.02,0.08]
Parental Health	−0.04 [−0.11,0.01]	−0.03 [−0.08,0.02]	0 [−0.05,0.05]	0.31 [0.21,0.41]	0.1 [0.06,0.18]	0 [−0.06,0.06]
Investment	1.18 [0.45,1.32]	0.88 [0.47,1.19]	0.74 [0.64,1.58]	0.69 [0.56,0.82]	−0.25 [−0.36,0]	−0.11 [−0.36,0.29]
Complementarity	−0.15 [−0.31,0.47]	0.08 [−0.17,0.45]	0.07 [−0.63,0.18]	0.01 [−0.07,0.08]	0.23 [−0.02,0.28]	0.16 [−0.19,0.45]
Elasticity of Subst.	0.87 [0.76,1.79]	1.09 [0.85,1.83]	1.07 [0.61,1.22]	1.01 [0.93,1.08]	1.3 [0.98,1.38]	1.19 [0.84,1.8]
Log TFP	−0.01 [−0.04,0.03]	0.39 [0.33,0.46]	1.76 [1.45,1.87]	−0.33 [−0.4,−0.26]	0.13 [0.04,0.23]	0.06 [−0.08,0.12]
Investment Residual	−0.93 [−1.1,−0.78]	−0.69 [−0.94,−0.48]	0.2 [−0.2,0.29]	−0.38 [−0.69,0]	−0.03 [−0.29,0.24]	0.1 [−0.06,0.18]
Number of Children	0 [−0.01,0.01]	0 [−0.02,0.01]	−0.02 [−0.03,0]	0.02 [0,0.04]	0 [−0.01,0.02]	0 [−0.02,0.01]
Older Siblings	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	−0.02 [−0.04,0]	−0.01 [−0.02,0]	0.01 [−0.01,0.02]
Urban	0.01 [0,0.01]	−0.01 [−0.02,0]	0 [−0.01,0]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.01,0]
Hindu	−0.01 [−0.02,0]	0 [−0.01,0.01]	0.02 [0.01,0.03]	0.01 [0,0.02]	0 [−0.01,0.01]	0 [−0.01,0.01]
Muslim	0 [0,0]	0 [0,0]	0 [−0.01,0]	0 [0,0]	0 [0,0]	0 [0,0]
Mother's Age	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.01,0.01]	0 [−0.02,0.01]	0 [−0.02,0.02]	−0.01 [−0.03,0]
Scheduled Caste	0 [−0.02,0]	0 [−0.01,0.01]	0.01 [0,0.02]	0 [−0.02,0.01]	0 [−0.02,0.01]	0 [−0.01,0.01]
Scheduled Tribe	0.02 [0.01,0.03]	0 [−0.01,0.01]	0 [0,0.01]	0.02 [0.01,0.03]	−0.01 [−0.02,0]	0 [−0.01,0.01]
BC Caste	−0.01 [−0.02,0]	0 [−0.01,0.01]	0 [0,0.02]	−0.01 [−0.03,0]	0.01 [0,0.02]	0 [−0.01,0.01]

Confidence intervals are obtained through 100 bootstraps samples by gender

Table 20: Boys' skills without income in production or investment

Here again, the coefficients are really similar to what can be seen in the main specification: although the instruments for investment are definitely weaker without income, dropping the latter does not modify the coefficients that are placed on them.

A.4 Alternative specifications for investment production functions

	Age 5	Age 8	Age 12
Cognition	NA [NA,NA]	0.194 [−0.04,0.4]	0.692 [−0.13,1.34]
Health	0.021 [−0.01,0.05]	0.011 [−0.05,0.08]	0.078 [−0.06,0.21]
Parental Cognition	0.153 [0.03,0.21]	0.222 [0.13,0.28]	0.155 [0.03,0.25]
Parental Health	0.069 [−0.01,0.13]	0.119 [0.03,0.21]	0.104 [−0.05,0.24]
Price Clothes	−0.008 [−0.08,0.08]	−0.27 [−0.39,−0.15]	0.082 [−0.23,0.3]
Price Notebook	−0.021 [−0.11,0.04]	−0.02 [−0.2,0.11]	0.45 [0.25,0.66]
Price Mebendazol	0.02 [−0.02,0.05]	−0.159 [−0.19,−0.04]	−0.158 [−0.25,−0.06]
Price Food	−0.077 [−0.22,0.12]	0.325 [0.03,0.66]	0.213 [−0.13,0.44]
Older Siblings	−0.011 [−0.05,0.01]	−0.026 [−0.1,0.02]	0.002 [−0.13,0.07]
Number of Children	−0.009 [−0.04,0.02]	−0.006 [−0.05,0.05]	−0.078 [−0.17,0.02]
Urban	−0.006 [−0.03,0.06]	−0.006 [−0.07,0.05]	0.103 [−0.03,0.2]
Hindu	0.024 [−0.05,0.13]	0.15 [0.03,0.23]	0.3 [0.05,0.49]
Muslim	0 [−0.01,0.01]	0.002 [−0.02,0.01]	0.018 [−0.02,0.03]
Mother's Age	0.075 [0,0.15]	0.199 [0.08,0.37]	−0.056 [−0.31,0.31]
Scheduled Caste	0.025 [−0.18,0.13]	−0.078 [−0.29,0.2]	−0.317 [−0.7,0.17]
Scheduled Tribe	−0.06 [−0.19,0.01]	−0.121 [−0.25,0.07]	−0.349 [−0.65,−0.06]
BC Caste	−0.044 [−0.21,0.03]	−0.124 [−0.26,0.12]	−0.275 [−0.76,0.12]

Confidence intervals obtained through 100 bootstraps samples

Table 21: Girls' parental investments without income as an instrument

	Age 5	Age 8	Age 12
Cognition	NA [NA,NA]	0.274 [−0.01,0.43]	0.849 [0.16,1.16]
Health	0.019 [0,0.05]	−0.029 [−0.07,0.02]	0.216 [0.05,0.39]
Parental Cognition	0.113 [0.07,0.15]	0.106 [0.05,0.18]	0.355 [0.21,0.55]
Parental Health	0.064 [0.02,0.11]	0.059 [0,0.12]	−0.031 [−0.24,0.13]
Price Clothes	0.073 [0.04,0.11]	−0.138 [−0.21,−0.05]	−0.174 [−0.48,0.18]
Price Notebook	−0.018 [−0.09,0.05]	−0.047 [−0.16,0.1]	0.396 [0.07,0.63]
Price Mebendazol	−0.018 [−0.04,0]	−0.036 [−0.08,0.04]	−0.009 [−0.09,0.06]
Price Food	−0.016 [−0.2,0.08]	0.728 [0.51,0.88]	0.144 [−0.28,0.37]
Older Siblings	−0.007 [−0.04,0.03]	−0.02 [−0.06,0.04]	0.008 [−0.08,0.14]
Number of Children	−0.02 [−0.04,0.01]	0.01 [−0.05,0.05]	−0.073 [−0.22,0]
Urban	−0.031 [−0.06,0]	−0.027 [−0.07,0.01]	−0.026 [−0.19,0.09]
Hindu	0.036 [−0.01,0.1]	0.053 [−0.14,0.17]	0.417 [0.19,0.63]
Muslim	0.001 [0,0.01]	0 [−0.01,0.01]	0.038 [0.01,0.07]
Mother's Age	0.074 [0.01,0.14]	0.059 [−0.03,0.18]	0.289 [−0.02,0.5]
Scheduled Caste	0.023 [−0.11,0.09]	0.117 [−0.05,0.2]	0.089 [−0.44,0.53]
Scheduled Tribe	−0.024 [−0.11,0.02]	−0.089 [−0.22,0.04]	−0.09 [−0.42,0.16]
BC Caste	0.008 [−0.08,0.07]	0.059 [−0.09,0.15]	0.009 [−0.47,0.37]

Confidence intervals obtained through 100 bootstraps samples

Table 22: Boy's parental investments without income as an instrument

Understandably, dropping income as an instrument makes parental cognition (a proxy for it) more positive. One measure of girls' health (age 8) also becomes significant. There is great variance in the cognition coefficients, but this was already the case for girls and in all but one instance (where the coefficient was already very low), the coefficients for boys are still non-zero. The price of food has a very big coefficient in the boys' investment function, which could mean that some complementarities between goods are not taken into account, as mentionned by Attanasio et. al. (2020). Finally, the sign and significance of belonging to a Hindu household changes for boys.