On The Production of Skills and the Birth Order Effect

Ronni Pavan*

Royal Holloway, University of London

September 10, 2013

Abstract

Data indicate that on average firstborn children outperform their younger siblings on measures such as test score, wages, educational attainment, employment, etc. Using data from the Children of the National Longitudinal Survey of Youth 1979, I also find evidence of a sizeable firstborn effect on many cognitive tests, a pattern that is robust to the inclusion of family level fixed effects and other controls. However, I also document considerable gaps in parental investment across birth order. Using a framework similar to Cunha and Heckman (2008) and Cunha et al. (2010), I estimate that differences in the provision of parental inputs across siblings can account for 20% to 45% of the gap in cognitive skills between firstborn children and their subsequent siblings. This framework can control for endogeneity in parental inputs, measurement error, missing observations, and for the dynamic impact of parental investments.

1 Introduction

Understanding what influences the cognitive development of children is a fundamental question for economists, sociologists and psychologists. One of the factors that can potentially shape the production function of achievement is family structure. In this paper, I focus on the relationship between birth order and cognitive skills. Data indicate that on average firstborn children outperform their younger siblings on measures such as test score, wages, educational attainment, employment, etc. As scientists we are interested in understanding whether the

*I would like to thank Josh Kinsler for helpful comments and discussions. All remaining errors are my own.
Contact the author at ronni.pavan@rhul.ac.uk

birth order effect is causal, and if so, what the sources of the effect are. Although several theories can explain why a causal relationship may emerge[1], one may also hypothesise that this relationship is instead driven by the fact that "low skill" families have a higher probability of having more children.

While in the psychology literature there is still a heated debate on whether a causal effect exists, the evidence presented in the economic literature seems to agree that indeed such an effect is present and sizeable. Interestingly, much less empirical research has been devoted to identifying the channels through which this effects originates. Notable exception are Price (2008), Hotz and Pantano (2011) and Lehmann et al. (2013).

Using data from the Children of the National Longitudinal Survey of Youth 1979 (CNLSY from now on), this paper contributes to the literature in several ways. First, I show evidence of the existence of the firstborn effect on many different measures of cognitive abilities. In particular, I find evidence of the effect in cognitive tests that are administered to very young children, indicating that the effect is likely to emerge very early in life[2]. In the CNLSY, first-born children score on average 0.1 to 0.2 standard deviations higher than their siblings. These results are obtained using family fixed effect estimation techniques in order to control for the fact that lower IQ mothers tend to have more children. Interestingly, the siblings of first born children start life at an advantage, as indicated by the fact that they have higher birth weight. However, this edge appears to dissipate quickly since early measures of cognitive skills like the Motor and Social Development Test already indicate a positive first-born effect[3].

Second, I show that a similarly sized birth order effect also appears on many measures of "parental investments". One possible explanation for why later born children perform poorly in cognitive tests relative to their first born siblings is that parents allocate fewer resources to them (resource dilution theory). To test this theory, I estimate family fixed effect regressions for several measures of parental inputs and find that indeed parents "invest" significantly more in firstborn children. This effect seems to be stronger among those measures that are

[1] In Section 2, I briefly review the theoretical and empirical literature on the birth order effect and other literatures related to this paper.

[2] Kessler (1991) and Rodgers et al. (2000) found no first-birth effect at very young ages with the same data, though they relied on a much smaller sample size. In the case of Rodgers et al. the smaller sample size is the result of a very strong sample selection.

[3] The results are consistent with the finding of Lehmann et al. (2013)

associated with younger children, such as whether a mother reads to his/her child.

Third and most importantly, I estimate an achievement production function in order to quantify the importance of parental disparities. Using the results from the estimation of my empirical model I quantify that the differences in family input across siblings can account for 20-45% of the cognitive gap. The model is similar to recent work by Cunha et al. (2010) and is advantageous in this setting as it allows the researcher to link the dynamic impact of partially unobserved parental inputs to the evolution of children achievements, controlling for the potential endogeneity of such inputs. An added benefit to the empirical model presented here is that it can account for the selection bias generated by optimal stopping models, as it directly models the fertility decision of mothers. The fixed effect models that are common in this literature cannot quantify nor eliminate this selection effect. My counterfactual simulations show that fertility is impacted by the quality of the firstborn child, however, the effect is small when compared to the total firstborn effect.

The paper most closely related to this one is recent work by Lehmann et al. (2013). They also seek to understand the role of parental investments in generating the firstborn effect. However, they largely follow the previous literature in estimating family fixed effects models. As noted previously, these models can't capture dynamics, selection into fertility, or allow for measurement error in the parental or children measures. As a result, their model underestimate the impact of parental inputs on the evolution of cognitive skills.

In Section 2, I review the exiting literature. In section 3, I present the data and the reduced form analysis. In section 4, I present the model and its estimation strategy. In section 5, I present the results. In section 6, I conclude.

2 Related Literature

Several early influential studies have found a negative relationship between birth order and intelligence (normally IQ tests)[4]. Many theories exist to explain this relationship, some of which suggest a causal impact of birth order on cognitive development. The resource dilution model (Blake (1981)) is essentially a modified dynamic version of Becker and Tomes (1976) quality-quantity trade-off model, where later born children have access to diluted resources

[4]See for example Belmont and Marolla (1973).

while older siblings have access to less constrained resources at least for a few years. The confluence model (Zajonc (1976)) suggests that the family structure–birth order, family size, and child spacing–has important influences on intellectual development in children. The admixture hypothesis is a formalisation of reverse causality that could lead to selection bias. It suggests that birth order and family size do not cause IQ (or other ability) differences, but rather are themselves caused by the distribution of intelligence among parents in the population (Velandia et al. (1978)). The optimal stoppage model predicts that a bad draw (i.e. a difficult to raise, problematic child) may lead to a decrease in subsequent fertility, and therefore this can induce selection as the last child tends to be of lower quality than the average.

Although there are very few empirical studies trying to identify a particular channel, several works have tested for the presence of a generic causal relationship between birth order and IQ. In psychology, researchers have analysed this theory for several decades leading to a plethora of works with contradicting findings. Focusing on the most recent contributions, Rodgers et al. (2000) find no evidence of birth order on intelligence using the CNLSY. This work has been criticised by Zajonc (2001) and particularly Armor (2001) who noted the extremely selected sample utilized by Rodgers et al. and the various manipulations of the dependent variable utilized for the analysis. Armor (2001) indeed finds a negative birth order effect using a different intelligence test in the same data set.[5] The discussion between Rodgers and coauthors and Zajonc and coauthors continues in other works using different data sets (see for example Wichman et al. (2006) and Zajonc and Sulloway (2007) and there does not seem to be a convergence between the two positions. In economics the situation is quite different. Although early contributions failed to find a significant relationship (Kessler (1991)), a recent wave of contributions seems to consistently find a negative birth order effect (Caceres-Delpiano (2006), Conley and Glauber (2006), Kantarevic and Mechoulan (2006), Booth and Kee (2009), Hotz and Pantano (2011), Black et al. (2005), Black et al. (2011), Lehmann et al. (2013)).

While economists generally agree that the birth order effect exists, there is much current debate about the mechanisms that originate this effect. Price (2008) finds evidence that parents spend more quality time with older siblings over the child's life and this is due to the fact that parents tend to equally split the quality time that have available across the existing

[5]Using the same data set Hotz and Pantano (2011), Lehmann et al. (2013) and my paper all find statistically and economically significant evidence of the presence of a negative both order effect.

siblings. Hotz and Pantano (2011) find evidence that parental discipline strategy differs across siblings, with younger siblings living in an environment with laxer rules. These works do not quantify how much of the firstborn gap can be explained by the channels they have identified. As previously discussed, Lehmann et al. (2013) find evidence of a birth order effect both in cognitive measures and parental investment measures. However when they control for the contemporaneous effect of such investment (plus prenatal maternal behaviour) in the cognitive measures, they fail to reduce significantly the birth order gap.

Somewhat tangential to the birth order effect literature, but critical for my study, is the literature on estimating the production function of achievement, often understood as test scores. Todd and Wolpin (2003) and Todd and Wolpin (2007) provide an excellent summary of the existing theoretical frameworks. They also review the related empirical results in term of impact of family inputs or school inputs on the evolution of test scores. In this paper I employ a new approach to modelling production functions developed Cunha and Heckman (2008) and Cunha et al. (2010) that treats true skills and family inputs as unobserved but assumes that noisy measures of each are available. Cunha and Heckman (2008) and Cunha et al. (2010) study the dynamic complementarities within and across cognitive and non-cognitive skills and quantify the importance of maternal skills or family inputs in the production of such skills. The advantage of using this latter framework is at least threefold. First, this method can account for the fact that what the econometrician observes are noisy measures of the true underlying factors. Second, different tests might be more or less noisy measures of the true underlying cognitive ability and therefore difficult to compare. The estimation procedure allows the data to determine which tests should receive the most weight. Third, many family characteristics and input measures are typically missing. This might cause the sample size to shrink significantly, as in a standard regression analysis a researcher needs to have information on both the left hand side variable and on all the right hand side variables in order to include the data point in the estimation. The methodology I employ naturally handles the presence of missing observation and allow the researcher to utilise more extensively the information available.[6]

[6] Another related literature is the one on the impact of family size on the development of children (Becker and Tomes (1976) provides a theoretical foundation). The empirical literature on family size finds contradicting results (see for example Black et al. (2005) and Black et al. (2010)) that depend on the type of instrument

3 Data and Descriptive Analysis

3.1 Data

The data utilized in this paper comes from the Children of the National Longitudinal Survey of Youth (CNLSY) sample. In this data set, women that were originally part of the National Longitudinal Survey of Youth (NLSY79) are interviewed every other year about their children, starting in 1986. The CNLSY records information on 11504 children from birth to 14 years old. This children are born from 4931 of the 6283 women interviewed in the NLSY79. After the children turn 15, they continue to be interviewed as part of the Young Adults of the NLSY79 (NLSY-YA). The CNLSY sample includes an oversample of black mothers and an oversample of poor white mothers. I utilise the sampling weights provided by the NLSY79.

Out of the original 11504 children, I exclude families with twins (462 kids), families with more than 3 children (3005), families with subsequent siblings that are more than 15 years apart (144), families with non-white mothers (3,397) and families with missing crucial information (9). The final sample consists of 2487 mothers and their 4785 children. There are 749 families with one child, 1178 with two children and 560 with three children.

For each household I have information about the demographics of the mother and of each sibling, such as year of birth and gender. In my empirical analysis I utilise a series of cognitive tests and measures of parental investments for each child. These variables contain information about different moments in the child's life. As measures of cognitive ability at birth I use gestation length and birth weight. The Motor and Social Development scale (MSD) measures dimensions of the motor, social, and cognitive development of children from birth to age 4. The Memory for Locations assessment was given in 1986 and 1988, to children ranging from 8 months to 3 years old and is a measure of a child's short-term memory. The Parts of the Body assessment was also given in 1986 and 1988, and it measures a 1- or 2-year-old child's receptive vocabulary knowledge. The Peabody Picture Vocabulary test (PPVT) is given to all children age three and over. It measures the children's vocabulary for English and provides an estimate of verbal ability. The Peabody Individual Achievement Test (PIAT) starts at age 5 and it is divided in three subtests: PIAT Mathematics, PIAT Reading Recognition and PIAT Reading Comprehension. These variables overlap to a large extent with the ones utilized by

utilized (twins versus gender composition of existing children).

Cunha et al. (2010) and a more accurate descriptions of them can be found in their data appendix or in the NLSY web pages (http://www.bls.gov/nls/nlsmrdat.htm). It should be noted that I use the raw scores for all tests, which are then standardised within the sample to make the interpretation of the results simpler and facilitate the estimation of the model, also because the percentile score of the PIAT Reading Comprehension provided by the NLSY is truncated for low scoring young children. Using the NLSY-YA I construct the numbers of years of education that children have by the age of 19. This variable, which is missing for a large fraction of children given that they have not reached that age by 2010 yet, provides some information on the cognitive abilities of children when they reach adulthood. I then include in my sample are the weight of the chid (at all ages, not just birth) and whether the child reads for fun. These last two variables are useful to check the robustness of some patterns and results.

The measures regarding parental investments are mainly those that describe the quality of a child's home environment. I utilise: whether the child has more than 10 books (asked to children of all ages), how often the mother reads to the child (children up to age 9), how often child gets out of house (up to age 2) or goes to an outing (from 3 to 5 years old), The amount of hours the child watches TV during weekdays or week-ends (age 3 and above), how often child is taken to museum (age 6 and above), how often child is taken to theater (age 6 and above), whether the child is taken to musical performances (age 6 and above), whether the family receives daily newspaper (age 6 and above), whether the child receives special lessons/activities (age 6 and above)

While the parental investment measures capture differences in the levels of inputs, some parents may simply be more efficient in translating those inputs into output. To allow for this in the model, I also select variables that measure mothers' cognitive ability. These variables are part of the Armed Services Vocational Aptitude Battery (ASVAB) that was administered to all mothers in 1980. In particular I use Arithmetic Reasoning, Mathematical Knowledge, Word Knowledge and Paragraph Comprehension.

In table 1, I present the sample statistics of all variables utilized in this paper both for the overall sample and by birth order. The variables are standardised (by age when applicable) within the sample to simplify comparisons and the estimation of the model.[7] The summary

[7]Different measures have different scales and different levels of dispersion. The firstborn effects on the

statistics provide some interesting insights. Of course the age of the mother at birth is an increasing function of birth order. As expected, the children's average age is instead inversely correlated with birth order. More interestingly the negative correlation between mothers' cognitive measures and fertility seems to appear only at the third child. The cognitive measures are in general lower for later born children, although they weigh more at birth. Interestingly the average weight after birth is instead larger for first born. We also see evidence of declining parental investments across birth cohorts in inputs such as number of books and whether the mother reads to the child. Finally it seems that later born children read less for fun and have fewer years of education at 19. In order to obtain a more refined estimate of the birth order effect on both parental investments and cognitive outcomes, I now turn to a fixed effect regression approach.

3.2 Reduced-Form Analysis

In this section, I present estimates of the firstborn effect on parental investments and child outcomes using family level fixed effect regressions. This type of estimation technique can control for all unobservables that might be correlated with birth order but are constant within the family. For example, a simple OLS regression for a cognitive test would deliver a biased estimate of birth order if larger families (larger birth orders on average) tend to have lower cognitive ability. The set of controls included in the regressions will depend on the type of dependent variable analysed but when applicable they include year dummies, age of the mother at birth, age of the child and gender. The standard errors are obtained by clustering at the family level. Unless otherwise stated, all variables have been standardised so the coefficient on first born can be interpreted in terms of standard deviations of the dependent variable.

Cognitive Measures. In the first panel of Table 2, I show the impact of being the first born on several measures of cognitive ability. Although we see a sizeable impact of birth order that is statistically significant and around 0.1 to 0.2 standard deviations in many variables, we can notice a few exceptions. The birth order effect is not significant in the PIAT math test, but is sizeable and significant in the two reading PIAT tests and in the PPVT. These

unstandardised versions of these variables are affected by these differences and therefore not directly comparable across each other.

results are consistent with the psychology literature which finds that the birth order effect is much stronger in verbal tests than in mathematical tests. The Memory for Locations and Body Parts tests show no significant results, however, these tests have few observations and were administered only in 1986 and 1988. The birth order effect is quite large and significant for the MSD overall, and is particularly sizeable for children under 2. The results for the MSD for very young children is in sharp contrast to the birth order effect on birth weight, the earliest measure of child well-being. First born children are actually smaller than there subsequent sibling. This pattern of results across the two measures indicate that the birth order effect appears at very early ages, although not at birth. In the last panel of Table 2 we can see though that although firstborn weigh less at birth, the direction of the birth order effect reverses early the life of these children. By the age of 4 they already weight around 0.1 standard deviations more than their younger siblings at the same age[8]. The variable "Years of education at age 19" is not standardised and therefore we can conclude that first born children have on average a quarter of a year more of schooling at age 19 than their younger siblings.

Although I report the firstborn effect over different ages only for the MSD, I have run fixed effect regressions interacting the first born dummy with age. For most measures the interaction is negative, statistically insignificant, and smaller than the pattern observed in the MSD. Exceptions are the interactions for the PIAT math and for the Body Parts tests which are both positive, although the latter is very imprecisely estimated.

Parental Investments. In the second panel of Table 2, I show the results for parental investments. Even in this case we see evidence of a first born effect in many measures, indicating that first born children receive more family inputs than their siblings. This can potentially explain part of the gap that we observe in the first panel of Table 2. The goal of the model presented in the next section is indeed to quantify the contribution of parental investments in the formation of the birth order cognitive gap. The firstborn effect seems to be stronger in inputs that are more common among young children, like number of books and whether the mother reads to the child, and is not as prevalent in inputs that are associated with older children, like whether the child is taken to the theater or museum. One might wonder whether this different attitude towards younger siblings in terms of verbal inputs has

[8]Lehmann et al. (2013) note though that first born children are less likely to be overweight at birth

a permanent impact on their habits later in life. In the last panel we see that first born children read for leisure on average half an hour more a week than their siblings (when the mean is 2.6 hours).

Summary. From Table 2, we can conclude that birth order does affect many tests of cognitive ability, and that this influence is present from very early in life (although not at birth). It also seems that family inputs are skewed towards older siblings, in particular when the first born are younger and probably still the only children in the family. Although these results are very informative, it is difficult to compare different tests because different tests might be differently able to capture the true underlying gap in cognitive ability. Furthermore, by looking at this table we cannot understand how important are the gaps in family input for the development of the gap in cognitive tests. Running a regression of cognitive tests controlling for family input would not be ideal for several reasons[9]. Parental inputs are hard to measure and hard to compare across each other, as we do not completely understand how and to what extend each parental behaviour affects the child's cognitive development. Hence we can think of our measures as contaminated by measurement errors. Because of this, directly including them in the regressions would understate the importance of family inputs. Second, the measures of parental investments are very often missing and this would make the sample size of a joint regression very small. Third, it would be difficult to understand the dynamic impact of each input, i.e. the effect that a variable has not only on the contemporaneous gap but also on its persistence. An additional drawback of this framework is that it cannot control for the possible bias introduced by the optimal stoppage behaviour discussed earlier. The framework that I introduce in the next section is indeed geared towards dealing with these aspects of the problem.

4 The Model

In this section I present a model of skill formation for young children. The main features of this model is that family structure is allowed to interact with the production function of cognitive skills and with its input determination. Skills and inputs are assumed to be unobserved by

[9]This is indeed the approach utilized by Lehmann et al. (2013).

the econometrician. Although the model is behavioural in its essence and I do not specify the preferences of individuals or their information set, the set up is more consistent with parents that observe and are aware of their children skills.

Children of age a and birth order j are characterised by a level of cognitive ability H_a^j. Mothers' cognitive ability is assumed to be constant over time and equal to H^m. I assume that maternal cognitive ability and the potential initial ability of each sibling come from a joint density with the cumulative density function $F_h(H_0, H^m; \theta)$ which depends on the parameter vector θ, where $H_0 = \{H_0^1, H_0^2, H_0^3\}$. Notice that this specification allows for both a correlation in the initial ability of siblings and that later born children draw their initial ability from a (marginal) distribution that is different from their older siblings.

The most important part of the model is the law of motion that explains the evolution of cognitive skills, i.e. the production function of skills:

$$H_a^j = \mu_{a,s}^j + \alpha_{1,a} H_{a-1}^j + \alpha_{2,a} H^m + \alpha_{3,a} I_{a-1}^j + \epsilon_a^j.$$

This function, which is similar in spirit to the one proposed by Cunha and Heckman (2008), allows for self productivity since past skills are allowed to influence future skills through an autoregressive component[10]. Mother's skills are allowed to directly affect a child's learning process. This feature captures the quality of the parental investments, rather than the quantity. I assume that parental investments are represented by a scalar unobserved (to the econometrician) variable I_a^j which also directly influences the evolution of skills. If $\alpha_{3,a}$ is positive then higher family inputs mean higher cognitive ability of the child in the next period. The constant $\mu_{a,s}^j$ is allowed to depend on both family size and the birth order of the child. This allows for the fact that larger families might have less efficient learning technologies or that later born siblings have a production function different from their older siblings. These effects are allowed to change over the life of the child. One common concern with production functions is that the innovations ϵ_a^j might be correlated with the inputs. For example, children who are likely to learn faster might be rewarded by their parents with a higher level of inputs. I tackle this issue by separating the innovation into $\epsilon_a^j = \alpha_{4,a}\pi + u_a^j$ where π is a time (and family) invariant random variable that can be correlated with family inputs. The newly defined shock

[10]The main differences with Cunha et al. (2010) is that we do not consider non-cognitive skills and we utilise a linear specification for the production function, like in Cunha and Heckman (2008).

u_a^j is now assumed to be independent across ages and to all inputs. I assume that the vector $u_a = \{u_a^0, u_a^1, u_a^2\}$ comes from a joint density $F_{u,a}(u_a;\theta)$, and therefore I allow this innovation to be correlated across siblings[11]. This can capture the fact that the accumulation pattern for the cognitive skills can be similar across siblings for reasons other than parental inputs (for example genetics).

The determination of family inputs, or parental investments, is crucial to the model. One possibility is that the amount of parental investment is a function of the family structure. For example, the "dilution" theory states that later born children have access to fewer resources, and therefore smaller parental investments than their older siblings. I assume that the input level for sibling j at age a is determined by the the following equation:

$$I_a^j = \eta_{a,s}^j + \beta_{1,a}H_a^j + \beta_{2,a}H^m + \beta_{3,a}FI_a^j + \beta_{4,a}\pi + e_a^j$$

where FI_a^j is the log family income of the child's family and $\eta_{a,s}^j$ is a constant which depends on both the order of the sibling j and the size of the family s[12]. Following CHS (2010), I include family income to control for the fact that families face financial constraint when choosing the optimal allocation of resources. Notice that the variable π that captures the endogeneity in the production function is indeed allowed to directly influence the amount of family inputs of the child. The input shock e_a^j is assumed to be independent across ages and to the other right hand side variables. I assume that the vector $e_a = \{e_a^0, e_a^1, e_a^2\}$ comes from a joint density $F_{e,a}(e_a;\theta)$, and therefore I allow this input shock to be correlated across siblings.

Not every family has three siblings and therefore we need to allow the family's fertility decision to be related to the unobserved variables of the model. The possibility that there is a spurious correlation between the cognitive level of a family and its size is what originated the large literature discussed earlier. Therefore, I allow the dynamic fertility decision (having one more child) to depend on the mother's cognitive ability:

$$Pr(s_{a+1} = n+1 | s_a = n) = F_{f,a}(H_a^n, H^m, \pi; \theta).$$

[11]The variable π is assumed to be independent of maternal ability. This is done without loss of generality given that we do not observe a direct measure of only this variable. I will discuss this further in the identification section.

[12]I also assume that family income FI_a^j has a law of motion that depends on maternal ability and past family income.

Notice that fertility may also depend on the variable π and the cognitive ability of the youngest child at the moment of the decision. This last feature can capture the "optimal stopping" behaviour that families might have. In particular, a negative correlation between family size (and therefore birth order) and the cognitive ability of the youngest child would be consistent with a family continuing to have children until a "bad" one arrives. The fertility model does assume however, that the fertility decision itself does not impact the cognitive skill of the existing children. The family size effects will only appear once the younger sibling is born[13]. The age at which a mother has the first child is also likely related to the unobservables of the model. This aspect of the fertility decision is exogenous to the model and is accounted for in the estimation procedure by controlling for the age of the mother at birth in all measurement equations.

At this point it is useful to make a summary of the channels through which the model can generate a birth order effect. The endogenous fertility decision allows for a couple of spurious (non-causal) correlations between birth order and cognitive ability. First, families with lower skills (through maternal skills or through π) might be more likely to have more children. Second, because fertility decisions depend on the skills of existing siblings, a birth order effect can emerge as a result of optimal stopping theory.

The structure of the input and production functions allows for several causal explanations for the emergence of the birth order effect. First, the gap at a certain age might be a consequence of the gap in the previous period because skills are persistent over time ($\alpha_{1,a}$). Second, the gap might be generated by different levels of inputs between younger and older siblings ($\eta^{j}_{a,s}$). These different input levels might be directly a function of birth order or might be a function of different family size between siblings[14]. Third, the gap might be generated by a different efficiency of the production function, captured by the constant ($\mu^{j}_{a,s}$). Also this effect might depend on both birth order and family size. This channel is more difficult to interpret as we do not fully understand or observe the fundamental parts of the production system. It could be related to unobservables inputs that are not captured by the variable

[13] I am ruling out the the possibility that the fact that a mother will have a new child in the future causes the cognitive ability of the existing children to grow differently.

[14] The first born child lives for a certain period in a family with only one child while this is never true for the second born child.

I_a^j or it could be related to the effect of the social interactions that siblings have. Lastly, the gap might be due to different initial conditions, i.e. to differences in the distribution of the initial draw. Also in this case, it is difficult in this framework to give a deep structural interpretation to this channel, although it likely relates to either the differences in behaviour of experienced pregnant women versus inexperienced pregnant women while pregnant, or to the differences that emerge in women's wombs after the first pregnancy and that affect the formation of future children[15].

4.1 Identification

As mentioned above the crucial variables (H^m, H_a, I_a, π) are assumed to be unobserved by the econometrician. Fortunately, our data set contains several noisy measures of these variables that can be used to obtain the identification of their distributions. I assume that the cognitive tests $M_{j,a}^h$ are noisy measures of the cognitive skills H_a^j, the parental investment measures $M_{j,a}^i$ are noisy measures of the family inputs I_a^j, the mother's ASVAB tests M^m are noisy measures of the maternal cognitive skills H^m. I also assume that the fertility decision $M_{j,a}^f$ of a family whose youngest child is of order j is a noisy measure of the cognitive skills of the j^{th} child, of maternal skills and of the family level variable π[16]. Following Cunha et al. (2010), we can utilise these measurements to obtain non-parametric identification of the joint density of all unobservables and therefore of the production function as well, which can be seen as a restriction on the overall density[17]. There are several requirements that need to be satisfied in order to apply their theorems. First, we need to have enough measurements. If the model contains L unobservables, then we need to observe $2L + 1$ measures of these variables. My data set satisfies this requirement. Second, I need to impose some normalisations because the scale of the unobservables is not identifiable. In order to simplify the estimation I assume

[15]Lehmann et al. (2013) finds evidence that women's behaviour while pregnant changes after the first pregnancy and, for example, they are less likely to reduce smoking or drinking. This nevertheless is not consistent with the fact that first born normally weight less than their younger siblings.

[16]Because fertility is a measure of both π and maternal ability, I assume that the two variables are independent. In order to separately identify the distribution of these two variables, I would need otherwise to observe some measures of π only.

[17]Note that variables at different age can be seen as different variables. For example the cognitive skill of the first born after one period H_1^1 is a different variable from the cognitive skill after two periods H_2^1.

that:

$$\begin{aligned} M^h_{j,a} &= \mu^h_a + \delta^h_a H^j_a + \epsilon^j_{j,a} \\ M^m &= \mu^m + \delta^m H^m + \epsilon^m \\ M^i_{j,a} &= \mu^i_a + \delta^i_a I^j_a + \epsilon^i_a \\ M^{f,*}_{j,a} &= \mu^f_a + \delta^f_{1,a} H^j_a + \delta^f_{2,a} H^m + \delta^f_{2,a}\pi + \epsilon^f_a \end{aligned}$$

and that $M^f_{j,a} = 1$ if $M^{f,*}_{j,a} > 0$. The first three line are meant to represent a vector of observables with at least 2 elements and 3 elements at least once (per child in the case of the first and third line). All noises (or uniquenesses) are assumed to be *i.i.d.* In order to identify the model I need to normalise $\delta = 1$ for one measurement equation for each unobservable. In the case of π, I instead set its variance equal to 1. I finally set the variance of the shock ϵ^f_a equal to 1 as well[18]. These normalisations, jointly with some regularity conditions, are sufficient to satisfy the requirements of theorem 1 of Cunha et al. (2010).

As discussed in Cunha et al. (2010), it is useful to map the human capital into an adult outcome in order to provide an economically useful metric. I utilise the highest grade completed at the age of 19 to anchor my model. In the CNLSY children are followed until they turn 15. Following CHS, I assume that after that age skills are constant and equal to H^j_{T+1}. Years of education at 19 (from the NLSY-YA instead) are then assumed to be equal to:

$$M^s = \mu^s + \delta^s H^j_{A+1} + \epsilon^s.$$

In order to have a complete non parametric identification of the distribution of H^j_{T+1}, I should observe at least 2 adult outcomes and I should normalise one of the δ^s. Instead, I assume that the production function at time T is equal to the one at time $T-1$ if not for its constant. These restrictions allow for the identification of the parameters $(\mu^s, \delta^s, \epsilon^s)$. The anchored version of the human capital can then be obtained by defining $\tilde{H}^j_a = \mu^s + \delta^s H^j_a$.

4.2 Estimation Strategy

Each child is followed from birth to 15 years of age. The interviews are held every other year which means that I have at most 9 observations per child, counting the adult outcome as

[18] I also assume that the mean in the production function can be decomposed as $\mu^j_{a,s} = \mu^{A,j}_a + \mu^B_s$ and that (without loss of generality) $\mu^{A,2}_a = \mu^{A,3}_a = 0$ and $\mu^B_1 = 0$. A similar assumption is also made for the mean of the input function $\eta^j_{a,s}$. This implies that these constant can only be studied in relative terms.

well. Hence, with a slight abuse of notation, I assume $a = 0$ for birth related variables then $a = 1, ..., 7$ during childhood and then $a = 8$ for the adult outcome. Although most parameters are indexed by a, I choose to restrict them to have a smaller set of parameters to estimate. For most cases, I select three stages $\tilde{a}$ where $\tilde{a} = 1$ if $a = \{0, 1\}$, $\tilde{a} = 2$ if $a = \{2, 3, 4\}$, and $\tilde{a} = 3$ if $a = \{5, 6, 7\}$, within which the parameters are assumed to be constant. For the production function, I instead select $a' = 1$ if $a = \{0\}$, $a' = 2$ if $a = \{1, 2, 3\}$, and $a' = 3$ if $a' = \{4, 5, 6, 7\}$, which provide a significantly higher value for the likelihood function and easier to interpret parameter estimates.

As mentioned in the data section, all variables (but years of education) are standardised within the sample, to facilitate the search for the optimal parameter vector. All cognitive and investment measurement equations include the same set of controls that are utilised in the fixed effect regressions (year, age, mother's age at birth, gender) in addition to the unobserved factors. The fertility equation controls for year, year of birth of the mother and the age of the last child, in addition to the unobserved factors. The family income equation also controls for year and year of birth of the mother in addition to maternal unobserved skills. Maternal ASVAB measurements control for birth year of the mother and for the age of the mother at the time of the first child in addition to maternal unobserved skills. In my model the fertility decision of the first child is exogenous as I only model whether families have two or three children and not at what age the first one arrives. Part of the correlation between age of the mother and skills could be due to the fact that younger mothers are likely to have different skills from older mothers, rather than a causal impact of age. In this paper this effect is net out by controlling for mother's age at birth in all measurements. Missing variables are assumed to be missing at random. As explained in the data section, my sample selection is such that no crucial variable like family size, year of birth, or gender of the child is missing[19].

I assume that all random variables come from normal distributions. This assumption does not seem to be very restrictive as noted in one of the robustness checks of Cunha et al. (2010). The model is estimated by using a quasi maximum likelihood approach. The term "quasi"

[19]Although I can potentially deal with missing family income by introducing one more unobserved variable in the model, I chose to keep the framework as simple as possible and I input the missing values using a linear projection based on past values. When family income is imputed I also increase the variance of the shock to the input to control for the prediction error.

refers to the fact that the model estimated is different form the model outlined in the previous section in one dimension. In the estimation routine, I assume that fertility follows a normal distribution, i.e. $M_{j,a}^{f,*} = M_{j,a}^{f}$, which makes it similar to a Linear Probability model. By assuming that fertility is a normal random variable rather than a binary variable, I will be able to implement a Kalman filter representation of the likelihood function, which enormously reduces the computational time. It is quite obvious that a binary variable is very different from a normal variable, and therefore I need to verify to what extent the model estimates can be mapped to the true structural parameters of the model. This is done by implementing a Monte Carlo analysis where I simulate the data assuming that $M_{j,a}^{f} = 1$ if $M_{j,a}^{f,*} > 0$ and $M_{j,a}^{f,*} = \mu_a^f + \delta_{1,a}^f M_a^j + \delta_{2,a}^f M^m + \delta_{2,a}^f \pi + \epsilon_a^f$, but I estimate the parameters $(\mu_a^f, \delta_{1,a}^f, \delta_{2,a}^f, \delta_{2,a}^f)$, plus the variance of the shock, assuming instead that $M_{j,a}^{f} = M_{j,a}^{f,*}$. The parameter vector chosen for the Monte Carlo analysis is similar to the actual estimates of the model, although parameters are restricted to have equal elements across different ages to make the estimation faster. I repeat the estimation on 10 simulated data sets of the same sample size as the real data, although I do not have missing variables in the simulated data, and I report the mean and standard deviations of the estimates of the most relevant parameters. The results are shown in Table 3.

In the first column I show the parameters that I utilise for the Monte Carlo analysis, while in the second column I report the average estimates and their standard deviations. There appear to be no economically significant bias in utilising the approximated likelihood function. The average estimates are very close to the structural parameters and even their dispersion is quite small. The impact of maternal ability in the production function is the most biased estimate but even in this case the distance between the true and the estimated parameter is not very large, being the coefficient quite small. The estimates for the fertility equation, which arguably should be the most affected by the approximation, are very close to the true parameters. In the third and fourth column I increase the size of the coefficients in the fertility equation, inducing therefore a much stronger selection that the data would suggest. In this case, we start seeing more evidence of biases that lead the estimated parameters to be half the size of the true parameters (and the constant nearly twice as large). It should be noted though that the estimates of the parameters relative to the production function are still very close to the true ones. I introduced the second set of parameters to show that a bias due to the approximation may indeed appear if the selection is strong, but it should be noted that the parameter estimates that are obtained in the next sections are much more consistent with the size of the coefficients in the first set. Thus, the potential bias due to the approximation is likely quite small making the simplification worthwhile given the large gain in computational efficiency.

5 Results

In this section I present the results obtained from the estimation of the model outlined above. The large number of parameter estimates (287) makes it unreasonable to present and discuss each one of them, although a table with all of them is available upon request. Instead, in the next few tables I summarise the most relevant estimates. In table 4 I show the estimated signal to noise ratio of each measure. A number close to 100% indicates that the measures provides a lot of information about the underlying unobserved variable, while a number closer to zero means that the measure is not very informative. Several messages emerge from this table. First, the ASVAB tests are quite good measures of the maternal cognitive ability. Birth weight is an excellent measure of skills at birth, although we will see later that those skills do not persist much. Early cognitive skills are measured imprecisely although the precision of the MSD increases quite a bit in the second stage. PIAT tests and PPVT are very good measures of skills. Parental investment measures are in general more noisy signals than the cognitive tests. The number of books and whether the mother reads to the child are particularly good measures of inputs when the child is young but they become less informative later in life. Most other measures for inputs contain some information about the underlying input.

In tables 5 and 6, I show the estimates for most parameters related to equations 1, 2 and 3. Firstborns seem to have a more efficient production function in the first stage of life. In stages 2 and 3 the production function of skills does not seem to differ substantially between first born and later born children. Although firstborn children start their life with fewer skills (which is what I show in table 7 and what was expected given the patterns in birth weight), they catch up very early on with their siblings as indicated by the large premium in the first period and skills depreciation. The relative size of the family has negative impact on the evolution of the skills, although the magnitude is relatively small. This is consistent with some of the results in the literature on family size, although my identification strategy is very different and does not require IV. I identify this effect by looking at how the slope of skill accumulation changes when new children are born. Even conditioning on families of the same size we can identify the effect by looking at families that have smaller or larger gaps between their siblings.

Cognitive skills of mothers are not important at young ages but they matter when kids are older (in stage 3). This is exactly the opposite of the impact of family inputs. They are important for young kids but not for older ones. Stage 2 seems to be the crucial period for parental investments. This period (in the production function) starts the year after birth and continues until the child is younger than 8 years old. Robustness checks not reported here indicate that the importance of parental inputs peaks towards the beginning of this period. It should be noted that I restricted the parameters of the production function to be positive for these inputs. As mentioned earlier, the persistence of cognitive skills is low at the beginning, but becomes larger over time. The evolution of skills is correlated across siblings, indicating that siblings experience similar trajectories in the evolution of skills, independently of other factors that may then reinforce this correlation, as parental inputs.

First born kids receive more family inputs although the gap is much larger for younger kids. Interestingly, larger families give more inputs. This is probably related to a certain complementarity in the provision of inputs. If I take one child to the museum it is relatively simple to take the other one as well. Inputs are mildly positively related to kids' skills and strongly related to maternal skills. The idiosyncratic shock is much larger when kids are younger. Shocks to inputs are correlated among siblings indicating, maybe, idiosyncratic parental preferences for investing more at a certain age rather than at another.

In the estimates of the fertility equation, we see evidence of an optimal stoppage behaviour because in the first and second stage the probability of having a second child is increasing in the skills of the first child. Interestingly the opposite is true in the third stage making the size of the total effect ambiguous. The estimates for the third child are an order of magnitude smaller.

An interesting feature of the fertility equation is that we do not see any negative relationship between mother's skills and fertility. This seems inconsistent with the fact that families with three kids have mothers with lower ASVAB tests, as seem in Table 1. In a set of OLS regressions that are not reported, I find that the relationship between size and ASVAB tests disappears once I condition on the mother's age at the time of the first child (which is lower for lower skill mothers). Hence, the model suggests that low skill mothers have more children because they have more time to accumulate additional children. The effect of early pregnancies on achievement is controlled for by including mother's age at birth in the cognitive measurement equations, family input measurement equations and maternal cognitive ability measure.

The standard deviation of the fertility equation is an extra parameter that is estimated but does not correspond to any structural parameter, and it is only reported in order to discuss this feature of the estimation. The variance of the fertility equation is in reality given by $p \times (1-p)$ where p is the probability of having one more child, but in the estimation fertility is treated as a continuous variable and therefore I estimate this variance as a free parameter rather than calculating what it should be given the other parameters. When I simulate the model this parameter is not utilised.

Not surprisingly the initial cognitive skills are correlated among siblings and they correlate with maternal skills, as seen in Table 7. First born children are born with fewer skills than their younger siblings. This is mainly driven by the fact that first born children weigh less at birth. It is not possible in this framework to understand why later born children have an advantage at birth but it might be related to either maternal experience (second time mothers know how to behave better while pregnant) or to the fact that women's wombs that have been already utilised are better suited for growing additional foetuses[20]. Unsurprisingly, years of education at 19 are strongly related to cognitive skills. The final parameter I would like to discuss is the dummy for first born children in the production function of skills for the final period, which is the period relative to the adult outcome. This variable is the last estimate of table 7. This parameter is quite large when compared to the first born dummy of the previous period, i.e. stage 3. It indicates that the first born effect does not disappear in adulthood and this is driven by the fact that we see a strong first order effect in the years of education at 19.

[20]Interestingly, the evidence presented by Lehmann et. al. (2013) is that second behave less appropriately when in their previous pregnancy.

5.1 The First Born Gap and Some Counterfactuals

In this section I discuss the magnitude of the firstborn effect implied by the estimates of the model and I quantify the sources of this effect. In order to do so I simulate a data set using the parameter estimates from above and calculate the implied gaps on the underlying unobserved variables. The simulated data set has the same sample size as the actual sample, although all observations are observed (i.e. no missing data). In the first column of table 8 I show the magnitude of the first born effect in family input provision. Young first born kids receive around 0.3 standard deviations more family inputs than their siblings. This gap reduces to about 0.2 standard deviations in stage 2 and 3. This result confirms what looked apparent from the fixed effect estimations. First born children receive more investments and this gap is larger when they are young.

In the second column I show the first born gap in cognitive skills measured in years of education at 19 years of age. As the parameter estimates suggested, first born children are born with 0.26 years of education deficit with respect to their younger siblings but soon this situation reverses and in stage 1 (birth excluded) they acquires an edge of nearly 0.5 years. This gap slowly decreases with age and in stage 3 it is half the size, at 0.25. In adulthood, the trend reverses and the gap reaches 0.34. The gap in adulthood increases as a result of the large first born dummy in the production function as indicated by Table 7, which is identified by the gap in years of education at 19. This parameter captures everything that happens between the age of 15 and the age of 19 and may also capture factors that affect educational achievement and are not contained in test scores. In a similar context, Chetty et al. (2011a) also finds that the impact of some factors (teacher quality in their case) fades away in test scores but reappears in long term outcomes. One may hypothesise that the presence of non-cognitive factors could explain this particular feature (Chetty et al. (2011b)), and this conjecture is indeed supported by the findings of Lehmann et al. (2013). Given that this parameter may also capture something that is related to parental behaviour but is missing in my model, I decide not to calculate the counterfactual gaps that I describe below for this final period.

In table 9, I compute a few counterfactual gaps to understand how different features of the model contribute to the first born effect. In this simulation exercise I look at the implied gaps at stage 1 (excluding birth) and at stage 2 and 3, excluding the adult stage for the above mentioned reasons. In the first line I report the baseline simulation from table 8. In the second line, I show how the cognitive gaps would be different had all children had a level of input that did not depend on family structure, i.e. setting both the firstborn dummy and the family size dummies equal to zero. The gap reduces by around 0.1 years of education at all stages. In percentage, this reduction goes from 20% in stage 1 to 45% in stage 3, where the cumulative effect of inputs is largest. These results indicate that indeed parental behaviour is responsible for the emergence of part of the gap. The third line shows an even larger reduction in the gap. In this exercise we eliminate only the first born effect, leaving the family size effect. This indicates that parents provide more family inputs in larger families and this compensates for the first born effect given that later born kids live in larger families. This feature is probably driven by some type of complementarity in input provision. As an example, when a parent goes to the museum, all children can benefit, so the more children a family has, the larger the benefit, while the cost arguably does not grow as much.

In line 4, I eliminate the effect of family structure from the production function of achievement. The gap in this case has a inverted U-shape. This is due to the fact that first born kids are born with a disadvantage and the first year dummy for first born is what reverses this pattern in their second year of life. Without this dummy the emergence of the first born gap is slower. As a result at stage 3 the gap is even smaller than what the previous experiment showed. Given that the impact of family structure on the production function can be seen as a "residual" explanation, this result suggests that although family inputs are important, other factors that are not captured by the measures present in the data are also crucial for explaining why first born children perform better. In line 5, we see evidence that the production function of achievement in larger families is less efficient. Once the effect of family size is removed, younger siblings perform better.

The last line of the table provides evidence that the optimal stoping model can indeed account for part of the firstborn gap, although the magnitude of this effect is small when compared to the other channels.

6 Conclusion

In this paper, I provide evidence of the existence of a sizeable birth order effect in both cognitive skills and parental investments. This evidence is robust to the inclusion of several controls and family level fixed effects. Using a methodology similar to the model proposed by Cunha and Heckman (2008) or Cunha et al. (2010), I then show that the difference in parental behaviour across different siblings can account for 20% to 45% of the birth order gap. I also find that the optimal stoppage behaviour, parents tend to keep having children until a difficult one is born, is unlikely to be an important driving force of the effect. The results indicate that a large fraction of the gap is still difficult to interpret.

In my model, the structure of the family is allowed to influence the choice of parental investments and the efficiency of the production function of achievement. One missing feature of the model I estimate is the presence of non-cognitive skills. These type of skills, that I decided to omit in this paper for computational reasons, might help understand better what generates the unexplained part of the gap.

References

Armor, D. (2001). On family size and intelligence. *American Psychologist*, 56(6/7):pp. 521–522.

Becker, G. S. and Tomes, N. (1976). Child endowments and the quantity and quality of children. *Journalof Political Economy*, 84(4):pp. S143–S162.

Belmont, W. and Marolla, F. A. (1973). Birth order, family size, and intelligence. *Science*, 182(4117):pp. 1096–1101.

Black, S. E., Devereux, P., and Salvanes, K. (2005). The more the merrier? the effects of family size and birth order on childrens education. *Quarterly Journal of Economics*, 120(2):pp. 669–700.

Black, S. E., Devereux, P., and Salvanes, K. (2010). Small family, smart family? family size and the iq scores of young men. *The Journal of Human Resources*, 45(1):pp. 33–58.

Black, S. E., Devereux, P., and Salvanes, K. (2011). Older and wiser? birth order and iq of young men. *CESifo Economic Studies*, 57(1):pp. 103–120.

Blake, J. (1981). Family size and the quality of children. *Demography*, 18(4):pp. 421–42.

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

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

Chetty, R., Friedman, J., and Rockoff, J. (2011a). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. NBER Working Paper 17699.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011b). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Conley, D. and Glauber, R. (2006). Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variations in fertility. *The Journal of Human Resources*, 41(4):pp. 722–737.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying, and estimating the technology for the formation of skills. *Journal of Human Resources*, 43(4):738–782.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Hotz, V. J. and Pantano, J. (2011). Strategic parenting, birth order and school performance. Working Paper.

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

Kessler, D. (1991). Birth order, family size, and achievement: Family structure and wage determination. *Journal of Labor Economics*, 9(4):413–426.

Lehmann, J.-Y. K., Nuevo-Chiquero, A., and Vidal-Fernandez, M. (2013). Birth order differences in early inputs and outcomes. IZA Discussion Paper No. 6755.

Price, J. (2008). Parent-child quality time: Does birth order matter? *The Journal of Human Resources*, 43(1):pp. 240–265.

Rodgers, J. L., Cleveland, H. H., van den Oord, E., and Rowe, D. C. (2000). Resolving the debate over birth order, family size, and intelligence. *American Psychologist*, 55(6):599–612.

Todd, P. and Wolpin, K. I. (2003). On the specification and estimating of the production production function for cognitive achievement. *Economic Journal*, 113(485):pp. F3–F33.

Todd, P. and Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school and racial test score gaps. *Journal of Human Capital*, 1(1):pp. 91–136.

Velandia, W., Grandon, G. M., and Page, E. B. (1978). Family size, birth order, and intelligence in a large south american sample. *American Educational Research Journal*, 15(3):pp. 399–416.

Wichman, A. L., Rodgers, J. L., and MacCallum, R. C. (2006). A multilevel approach to the relationship between birth order and intelligence. *Personality and Social Psychology Bulletin*, 32(1):pp. 117–127.

Zajonc, R. B. (1976). Family configuration and intelligence: Variations in scholastic aptitude scores parallel trends in family size and the spacing of children. *Science*, 192(4236):pp. 227–236.

Zajonc, R. B. (2001). Birth order debate resolved? *American Psychologist*, 56(6/7):pp. 522–523.

Zajonc, R. B. and Sulloway, F. J. (2007). The confluence model: Birth order as a within-family or between-family dynamic? *Personality and Social Psychology Bulletin*, 33(9):pp. 1187–1194.

Table 1: Summary Statistics

	Whole sample			First Child			Second Child			Third Child		
Demographics	Mean	Std	Obs.	Mean	Std	Obs.	Mean	Std	Obs.	Mean	Std	Obs.
Year	1995.66	6.13	30488	1994.67	5.92	14944	1996.42	6.14	11674	1997.90	6.13	3870
Child's Age	7.39	4.05	19958	7.66	4.01	9712	7.21	4.08	7758	6.92	4.07	2488
Child's Gender	0.48	0.50	4785	0.49	0.50	2487	0.47	0.50	1738	0.48	0.50	560
Mother's Age at Birth	26.71	5.71	4785	24.97	5.55	2487	27.91	5.22	1738	30.26	5.21	560
Mother's Cogn. Measures												
Arithmetic Reasoning	17.57	6.64	4539	17.57	6.64	2354	17.62	6.65	1654	17.39	6.66	531
Mathematical Knowl.	13.74	5.98	4539	13.71	5.95	2354	13.83	5.97	1654	13.60	6.16	531
Word Knowledge	26.89	6.54	4539	26.95	6.48	2354	26.97	6.50	1654	26.44	6.93	531
Paragraph Compr.	11.67	2.91	4539	11.68	2.88	2354	11.70	2.92	1654	11.57	3.04	531
Child's Cogn. Measures												
Birth Weight	120.01	23.34	4364	118.11	19.44	2384	121.82	21.19	1517	123.63	40.82	463
Gestation Length	38.70	2.11	4298	38.75	2.25	2290	38.61	2.00	1530	38.77	1.65	478
MSD	10.34	3.28	4833	10.67	3.15	2128	10.11	3.39	2017	9.97	3.25	688
Memory for Locations	6.83	3.23	1292	7.07	3.13	652	6.73	3.30	508	5.94	3.33	132
Body Parts	6.52	3.30	886	7.02	3.13	442	6.27	3.25	350	4.72	3.64	94
PPVT	85.57	35.43	7353	88.75	34.69	3857	82.93	35.83	2699	79.65	36.05	797
PIAT Math	41.40	17.94	12408	42.11	17.84	6354	40.96	18.00	4636	39.76	18.07	1418
PIAT Comprehension	40.44	17.11	12118	41.71	17.17	6193	39.52	17.01	4528	37.91	16.76	1397
PIAT Recognition	44.50	20.04	12363	45.72	19.94	6330	43.62	20.04	4616	42.13	20.08	1417
Parental Investments												
Number of Books	0.91	0.28	20130	0.93	0.25	9797	0.90	0.30	5595	0.88	0.33	2520
Mother Reads to Child	4.61	1.42	14240	4.76	1.37	6770	4.49	1.44	5595	4.44	1.49	1875
Child Out of House	4.28	1.17	3835	4.34	1.12	1644	4.22	1.20	1608	4.29	1.21	583
Child to an Outing	3.67	0.99	4441	3.69	0.99	2102	3.63	1.00	1760	3.72	0.99	579
Hours TV Week-Ends	3.14	2.60	15654	3.15	2.64	7133	3.15	2.57	6343	3.06	2.59	2178
Hours TV Week-Days	2.77	3.27	15654	2.83	3.34	7128	2.73	3.19	6348	2.72	3.24	2178
Child Taken to Museum	2.24	0.90	16288	2.29	0.91	8145	2.22	0.88	6205	2.12	0.87	1938
Child Taken to Theater	1.96	0.83	11866	1.97	0.83	6049	1.95	0.82	4449	1.93	0.83	1368
Child Taken to Musical	0.60	0.49	11880	0.57	0.50	6057	0.63	0.48	4452	0.65	0.48	1371
Daily Newspaper	0.54	0.51	11878	0.54	0.51	6053	0.54	0.51	4455	0.52	0.51	1370
Special Lessons/Activ.	0.73	0.45	11872	0.73	0.44	6050	0.73	0.44	4449	0.71	0.46	1373
Other Variables												
Weight	63.64	36.33	19959	65.83	36.62	9728	62.10	35.91	7743	60.06	36.02	2488
Reads for Fum	2.61	4.35	6175	2.83	4.34	2676	2.42	4.08	2592	2.52	5.02	907
HGC at 19	12.83	1.67	1646	13.01	1.63	940	12.71	1.60	560	12.08	1.97	146

Notes: Data from the CNLSY. The final sample consists of 2487 mothers and their 4785 children. There are 749 families with one child, 1178 with two children and 560 with three children.

Table 2: Fixed Effect Estimates of the First Born Effect

	First-Born Dummy	Std	t-stat	N. obs.	Control Dummies
Child's Cognitive Measures					
Birth Weight	-0.12***	0.05	-2.62	4360	BY,AM,G
Gestation Length	0.03	0.05	0.51	4296	BY,AM,G
MSD	0.16***	0.06	2.93	4832	AM,G,A,Y
MSD for less than 2 years old	0.24**	0.08	2.97	2863	AM,G,A,Y
Memory for Locations	-0.08	0.17	-0.50	1292	AM,G,A,Y
Body Parts	-0.08	0.16	-0.49	885	AM,G,A,Y
PPVT	0.18***	0.04	4.69	7353	AM,G,A,Y
PIAT Math	0.02	0.04	0.45	12408	AM,G,A,Y
PIAT Comprehension	0.14***	0.04	3.74	12118	AM,G,A,Y
PIAT Recognition	0.11***	0.04	2.78	12363	AM,G,A,Y
Parental Investments					
Number of Books	0.11***	0.03	4.32	20130	AM,G,A,Y
Mother Reads to Child	0.29***	0.02	12.15	14240	AM,G,A,Y
Child Out of House	0.13**	0.05	2.45	3833	AM,G,A,Y
Child to an Outing	0.15***	0.05	3.28	4439	AM,G,A,Y
Hours TV during Week-Ends	-0.09 ***	0.02	-3.93	15654	AM,G,A,Y
Hours TV during Week-Days	0.01	0.02	0.57	15654	AM,G,A,Y
Child Taken to Museum	0.03	0.02	1.64	16287	AM,G,A,Y
Child Taken to Theater	0.02	0.02	0.89	11866	AM,G,A,Y
Child Taken to Musical	-0.03	0.02	-1.11	11880	AM,G,A,Y
Daily Newspaper	0.01	0.02	0.25	11878	AM,G,A,Y
Special Lessons/Activities	0.02	0.03	0.81	11872	AM,G,A,Y
Other Variables					
Weight, less than 2 years old	-0.14	0.14	-1.02	545	AM,G,A,Y
Weight, 2-4 years old	0.12**	0.05	2.35	2562	AM,G,A,Y
Weight, 4-6 years old	0.14***	0.05	2.65	2860	AM,G,A,Y
Weight, 6-8 years old	0.18***	0.05	3.55	2997	AM,G,A,Y
Weight, more than 8 years old	0.12***	0.04	2.87	10993	AM,G,A,Y
Reads for Fun (Hours)	0.52***	0.13	3.87	13941	AM,G,A,Y
HGC at 19 (Years of Education)	0.27**	0.14	1.90	1646	BY,AM,G

Notes: These results are obtained using family level fixed effect estimation techniques. Standard errors are clustered at the family level. *,**,*** Indicate a coefficient is statistically significant at a 1, 5 and 10% significance level. BY = birth year, AM = age of the mother, G = gender, A = age of the child, Y = survey year.

Table 3: Monte Carlo Estimation

	True	Mean Estimates	True	Mean Estimates
Production Function Parameters				
FirstBorn dummy	0.030	0.027	0.030	0.025
		(0.007)		(0.010)
H_{t-1}	0.800	0.806	0.800	0.802
		(0.005)		(0.003)
H_m	0.100	0.078	0.100	0.079
		(0.005)		(0.008)
I_{t-1}	0.500	0.521	0.500	0.524
		(0.030)		(0.035)
π	-0.1	-0.111	-0.100	-0.112
		(0.010)		(0.011)
Fertility parameters				
Constant	0.150	0.156	0.150	0.237
		(0.003)		(0.007)
H_{t-1}	0.100	0.092	0.400	0.208
		(0.002)		(0.004)
H_m	-0.020	-0.018	-0.200	-0.113
		(0.004)		(0.006)
π	-0.040	-0.035	-0.200	-0.107
		(0.003)		(0.005)

Notes: This Monte Carlo analysis is performed estimating the model on 10 simulated data sets of the same size as the real data. The numbers in parenthesis are the standard deviations of the estimated coefficients. For expositional clarity, not all coefficients are reported.

Table 4: Estimated Signal to Noise Ratio of Each Measure

Var. Name	**Time Invariant**	**0-2 years old**	**3-9 years old**	**10-14 years old**
Maternal Skills				
Arithmetic Reasoning	72.8%			
Mathematical Knowl.edge	67.4%			
Word Knowledge	62.0%			
Paragraph Comprehension	53.9%			
Birth Outcome				
gestation Length	26.7%			
Birth Weight	99.8%			
Adult Measure				
Highest grade completed at 18	21.5%			
Child's Cognitive Measures				
Memory for Locations		5.9%	0.1%	
MSD		9.7%	24.0%	
Body Parts		10.6%	8.1%	
PPVT			25.2%	29.2%
PIAT Math			31.7%	33.7%
PIAT Reading Recognition			59.5%	45.0%
PIAT Reading Comprehension			62.4%	51.1%
Parental Inputs				
Special Lessons/Activities			12.9%	11.0%
Number of Books		25.3%	8.4%	5.0%
Child Taken to Museum			18.5%	19.1%
Child Taken to Musical			5.7%	9.7%
Daily Newspaper			6.8%	7.3%
Child Out of House		4.4%		
Child to an Outing			6.8%	
Mother Reads to Child		35.0%	14.3%	3.7%
Child Taken to Theater			19.2%	23.4%
Hours TV during Week-Ends			13.1%	11.5%
Hours TV during Week-Days			8.7%	9.6%

Notes: These numbers are obtained simulating data using the parameter estimates on a sample of the same size as the real data. Each number represent the fraction of the measurement variance that can be explained by the underlying unobserved factor, either skill or input.

Table 5: Parameter Estimates

Production Function	After Birth	Stage 2	Stage 3	Input Function	Stage 1	Stage 2	Stage 3
First Born Dummy	0.355	-0.018	0.006	First Born Dummy	0.206	0.091	0.095
	(0.017)	(0.007)	(0.005)		(0.019)	(0.006)	(0.006)
Family Size = 2	-0.024	=	=	Family Size = 2	0.085	=	=
	(0.003)	=	=		(0.004)	=	=
Family Size= 3	-0.034	=	=	Family Size= 3	0.074	=	=
	(0.004)	=	=		(0.005)	=	=
Past period Skills	0.058	0.665	0.814	Child's Skills	0.103	0.022	0.022
	(0.010)	(0.010)	(0.003)		(0.020)	(0.003)	(0.004)
Maternal Skills	0.000	0.000	0.119	Maternal Skills	0.203	0.171	0.223
	-	-	(0.012)		(0.014)	(0.008)	(0.008)
Investment	0.232	0.814	0.000	Family Income	0.091	0.027	0.036
	(0.028)	(0.019)	-		(0.010)	(0.002)	(0.003)
π	0.171	-0.242	0.034	π	0.293	0.288	0.309
	(0.012)	(0.005)	(0.003)		(0.013)	(0.005)	(0.005)
Std Shock	0.118	0.385	0.111	Std Shock	0.334	0.002	0.013
	(0.011)	(0.004)	(0.002)		(0.015)	(0.001)	(0.001)
Correlation	0.222	=	=	Correlation	0.426	=	=
	(0.010)	=	=		(0.023)	=	=

Notes: Standard errors are in parenthesis. See section 3 for a description of the data utilized and section 4 for a description of the ML estimation procedure. The coefficients for family size dummies and for the correlation in the shocks are restricted to be the same across all ages.

Table 6: Parameter Estimates for the Fertility Equation

	Second Child	Third Child
Previous Child's Skills (Stage 1)	0.084	-0.001
	(0.009)	(0.005)
Previous Child's Skills (Stage 2)	0.075	-0.007
	(0.006)	(0.003)
Previous Child's Skills (Stage 3)	-0.045	-0.004
	(0.015)	(0.010)
Maternal Skills	0.005	0.002
	(0.004)	(0.002)
π	-0.025	-0.007
	(0.004)	(0.002)
Std Shock*	0.153	0.058
	(0.006)	(0.001)

Notes: Standard errors are in parenthesis See section 3 for a description of the data utilized and section 4 for a description of the ML estimation procedure. *= This parameter does not exist in the structural model as the standard deviation of the shock is equal to $\sqrt{p(1-p)}$, where p is the probability of having an extra child. See section 4.2 and 5 for a discussion.

Table 7: Other Parameter Estimates

Distribution Initial Draw		*Adult Related Parameters*	
Std Children	0.991	Adult Measurement	
	(0.016)	Constant	12.587
Std Mothers	0.813		(0.026)
	(0.015)	Child's Skill	1.121
Covariance between Children	0.433		(0.031)
	(0.014)	Std Measurement Error	1.995
Covariance with Mother	0.123		(0.037)
	(0.012)	Production Function of Adult Skills	
First Born Dummy in Initial Mean	-0.221	First Born Dummy	0.171
	(0.019)		(0.051)

Notes: Standard errors are in parenthesis. See section 3 for a description of the data utilized and section 4 for a description of the ML estimation procedure.

Table 8: Firstborn Effects

	Gaps in Parental Inputs (standard deviations)	Gaps in Child's Skills (years of eduction at 19)
Birth	0.36	-0.26
Stage 1	0.30	0.49
Stage 2	0.17	0.37
Stage 3	0.20	0.25
Adult		0.36

Note: The firstborn effects reported here are either the average gaps between the cognitive skills of firstborns and their siblings, measured in terms of years of education at 19, or the average gaps between the parental investments for firstborns and their siblings, measured in standard deviations. These numbers are obtained simulating data using the parameter estimates on a sample of the same size as the real data.

Table 9: Counterfactuals Firstborn Gaps in Cognitive Skills

	Stage 1	Stage 2	Stage 3
Baseline	0.49	0.37	0.25
No Firstborn or Size effect in Inputs	0.39	0.22	0.14
Percent Reduction	19.9%	40.7%	44.9%
No Firstborn effect in Inputs	0.36	0.15	0.10
Percent Reduction	26.3%	58.5%	61.6%
No Firstborn or Size effect in Production	0.14	0.21	0.11
Percent Reduction	72.1%	43.1%	53.9%
No Firstborn effect In Production	0.18	0.25	0.16
Percent Reduction	62.5%	31.5%	37.2%
No Optimal Stopping	0.47	0.33	0.22
Percent Reduction	3.1%	10.3%	11.8%

Note: The firstborn effects reported here are the average gaps between the cognitive skills of firstborns and their siblings. They are measured in terms of years of education at 19. These numbers are obtained simulating data using the parameter estimates on a sample of the same size as the real data.