Sibling Differences in Educational Polygenic Scores: How do Parents React?*

Anna Sanz-de-Galdeano
University of Alicante and IZA
Anastasia Terskaya
Universidad de Alicante

February 15, 2019

Abstract

We study whether parents' investment decisions exacerbate or mitigate differences in their children's genetic predisposition for education. Parental investment decisions depend both on parental preferences regarding inequality in the distribution of their children's quality and on how costly it is for parents to add to their children's quality by investing in their human capital (or the price effect). Our empirical strategy allows us to isolate the effects of parental preferences regarding equality from the price effect, a distinction that cannot be made when relying on sibling or twin fixed-effects models. Importantly, recent advances in molecular genetics allow us to use genetic variants that predict educational attainment as a measure of children's endowments. Individuals' genetic makeup is fixed at conception, so these indicators cannot be affected by parental investment decisions. We find evidence that parents of nontwin siblings display inequality aversion and, given the absolute endowment level of one child, they invest more in him/her if his/her sibling is better-endowed. Parents of twins instead display neutral preferences regarding equality, possibly because it is difficult to provide differential parental investments across children of the same age.

JEL classification: D13, D64, J13, J24

Keywords: Intrahousehold allocation of resources; educational polygenic scores; parental investments; endogenous fertility; Add Health.

*Authors' contact: Anna Sanz-de-Galdeano and Anastasia Terskaya, FAE, Universidad de Alicante, Carretera de San Vicente s/n, 03080 San Vicente – Alicante, Spain. Emails: anna.sanzdegaldeano@gmail.com and anastasiaterskaya@gmail.com. Sanz-de-Galdeano is also affiliated with CRES-UPF. She acknowledges financial support from the Spanish Ministry of Economy and Competitiveness Grant ECO2017-87069-P. We are grateful to Lola Collado and Daniel Rees, as well as participants at the Applied Economics Seminar at the University of Alicante, and the Econometrics Seminar Series at the University of Melbourne for their helpful comments.

1 Introduction

Is the family an equalising agent? Do parents exacerbate or mitigate differences in children's endowments by reallocating resources within the family? These are crucial questions for both academics and policy makers, as parental responses should be considered when designing policies aimed at fostering human capital and reducing inequalities among children.

The literature analysing how parental investments are related to their children's endowments is vast, and it has continuously grown since the seventies. In their seminal contribution, Becker and Tomes (1976) propose a model of resource allocation within the family, and analyse how parental investments are affected by differences in their children's ability or other aspects of their endowments. They show that, if the cost to parents of adding to children's quality by investing in their human capital is negatively related to their endowments (that is, if such cost is higher for less able children), parents may will reinforce differences in children's endowments by investing more in better-endowed children. In contrast, Behrman *et al.* (1982) develop a general preference model that introduces parental aversion to inequality in the distribution of their children's quality. In their framework, the degree of parental inequality aversion is key in determining whether parents will follow a compensating strategy (devoting more resources to a child with a smaller endowment) or a reinforcing strategy (devoting more resources to his/her better-endowed sibling).[1]

The subsequent empirical literature has so far reached mixed conclusions on whether parents compensate or reinforce differences in their children's endowments. Some studies have found evidence of parental compensatory behaviour (Behrman *et al.*, 1982; Pitt *et al.*, 1990; Bharadwaj *et al.*, 2018; Terskaya, 2019), while others have found that parents follow a reinforcing strategy (Datar *et al.*, 2010; Aizer and Cunha, 2012; Hsin, 2012; Frijters *et al.*, 2013; Rosales-Rueda, 2014). Interestingly, Yi *et al.* (2015) provide evidence that, when faced with differences in health endowments among their children, parents react by compensating in terms of health investments, while they instead reinforce inequalities through their human capital investment decisions. This lack of consensus is due to several factors:

i the use of different data sets for different contexts;

ii the use of different measures of children's endowments and/or parental investments; and last but not least:

iii the varying ways in which different authors have dealt with the numerous identification challenges that arise when empirically investigating how parents react to differences in their children's endowments.

This paper studies whether parents' investment decisions exacerbate or mitigate differences in their children's genetic predisposition to educational attainment while taking into

[1]Since Behrman *et al.* (1982) assume that the cost of adding to quality is unrelated to children's endowments, evidence that parents follow a compensating strategy can be used to infer that parents are inequality averse within their framework.

account that parental investment decisions depend both on parental preferences regarding inequality in the distribution of their children's quality and on how costly it is for parents to add to their children's quality by investing in their human capital (or the price effect). We deploy an empirical strategy that allows one to isolate the effects of parental preferences regarding equality from the price effect, a distinction that cannot be made when relying on sibling or twin fixed-effects models. Importantly, recent advances in molecular genetics allow us to use genetic variants that predict educational attainment as a measure of children's cognitive endowments. These indicators are not only strong predictors of cognitive outcomes, but they are also fixed at conception and hence cannot be affected by parental investment decisions. We also address the potential endogeneity of fertility in families with nontwin siblings (i.e., parents sequentially decide whether to have more children depending on the endowments of previous children) by focusing on first-born children and conditioning on their own absolute endowment levels. We find evidence that parents of nontwin siblings display inequality aversion and, for a given level of absolute cognitive endowment of one child, they invest less in him/her if his/her sibling is better-endowed. Parents of twins instead display neutral preferences regarding equality, perhaps because it is harder for them to invest differently across their children.

Our paper adds to the existing literature in several important ways:

First, we use educational polygenic scores as an indicator of children's propensity to educational achievements. This indicator is not only interesting per se, as we will argue later, but it also allows one to circumvent reverse causality issues. In particular, endowment indicators measured during childhood may be the result of prior parental (both post- and pre-natal) investments,[2] while endowment indicators measured at birth (such as birthweight) may reflect pre-natal investment decisions (Del Bono *et al*., 2012; Almond and Currie, 2011; Currie, 2011 among others provide extensive evidence on the effect of prenatal environment on children's at-birth endowments).[3] In contrast individuals' genetic endowments are fixed at conception and hence cannot be the consequence of parental investment choices.

Second, we focus on parental responses to differences in children's cognitive rather than health endowments, while most previous studies have focused on the latter. This may be due to the fact that at-birth measures of endowments (such as birthweight), which are less likely to suffer from reverse causality than indicators measured later on in life, are not frequently available in the cognitive domain. Be that as it may, with a few exceptions (Ayalew, 2005; Frijters *et al*., 2013), we know much less about parental responses to differences in their children's cognitive endowments than about how parents react when they face differences in their chil-

[2]For instance, Rosales-Rueda (2014) and Yi *et al.* (2015) use siblings' and twins' variation, respectively, in the exposure to health shocks during early childhood as a measure of children's endowments and study whether parents invest more in healthier children (following a reinforcing strategy) or if, in contrast, they invest more in their siblings who are in worse health (following a compensating strategy). Ayalew (2005) relies on siblings' variation in the scores of the standard Raven's Colored Progressive Matrix (CPM) test to measure cognitive endowments.

[3]Most of the studies relying on at-birth indicators focus on birthweight (Datar *et al*., 2010; Hsin, 2012; Cabrera-Hernández, 2012; Restrepo, 2016; Abufhele *et al*., 2017; Bharadwaj *et al*., 2018). Adhvaryu and Nyshadham (2016) use variation in in utero exposure to a iodine supplementation programme and Zweimueller and Halla (2014) use in utero exposure to the radioactive fallout from the Chernobyl accident.

dren's health endowments. However, parents may not necessarily respond in the same way to their children's differences in cognitive versus health endowments.

Third, previous theoretical contributions have highlighted that parental investment decisions do not only depend on parental preferences regarding inequality in the distribution of their children's quality (Behrman *et al.*, 1982), but also on how costly it is for parents to add to their children's quality by investing in their human capital (Becker and Tomes, 1976). If such costs differ among children, even inequality averse parents may follow a reinforcing strategy if the cost of investing in their lower-endowed child is sufficiently higher than the cost of investing in their higher-endowed child (Terskaya, 2019). Therefore, evidence based on family fixed effects models (which essentially compares parental investments across children within the family) is not informative on whether patents are inequality averse or not, an issue that has so far been empirically overlooked. In order to investigate this, we study how parental investments in one child are affected by the divergence between his/her endowment and that of his/her sibling while holding constant the child's own cognitive endowment, which serves as a proxy for the costs to adding to his/her quality faced by the parents. That is, our empirical model aims at answering the following question: do parents invest more or less in children who are more or less able than their siblings, but who are otherwise comparable in terms of their own ability and hence the costs their parents face if they invest in them?

Fourth, we deal with the potential endogeneity of fertility decisions (which may affect studies based on nontwin siblings) by focusing on twins and on first-born children. Studies based on nontwin siblings are likely affected by endogenous fertility because parents' decisions to have more children may depend on the endowments of previous children (see Ejrnæs and Pörtner 2004 among others). In order to address this issue we focus on first-born children and study how parents respond to their relative endowments (with respect to those of their later-born siblings) while conditioning on their own absolute endowment levels. Studies based on twins comparisons cannot be affected by endogenous fertility, but we analyze twins separately because previous evidence suggests that parents of twins are less likely to be responsive to their endowment differences (Bharadwaj *et al.*, 2018) if, for instance, they find it more costly to implement favouritism among their twin children than parents of nontwin siblings (Almond and Mazumder, 2013).

Finally, our study also adds to an emerging literature that aims at integrating genetics and the social sciences. For instance, recent contributions have studied the association between educational polygenic scores and human capital accumulation (Domingue *et al.*, 2015; Papageorge and Thom, 2018), labor market outcomes (Papageorge and Thom, 2018) and wealth at retirement (Barth *et al.*, 2018). However, there is still much to learn regarding the mechanisms through which genetic endowments affect socioeconomic outcomes, and whether their impact is reinforced or mitigated by environmental factors in different contexts. To our knowledge, this is the first study showing that relative genetic endowments have an impact on how much parents invest in their children. This suggests that the effect of individuals genetic predisposition for education on future outcomes is not only direct, but it may also operate through intra-household investment decisions.

2 Empirical Strategy

Our primary goal is to estimate how parents' investment decisions are affected by differences in their children's genetic endowments linked to future cognitive outcomes. In particular, we want to distinguish between two potential mechanisms that may induce parents to follow a reinforcing strategy (or to invest more in their better-endowed children than in his/her lower-endowed siblings) versus a compensating strategy (that is, to invest more in children with a lower relative endowment).

1. The cost of adding to children's quality or the price effect. The cost of investing in high ability children may differ from the cost of investing in their less able siblings (Becker and Tomes, 1976).

2. Parental preferences for equality. If, within a family, children's endowments differ, inequality averse parents will try to attenuate such differences through their investment decisions. In contrast, if parents care more about efficiency than equality, their investment decisions will aim at maximizing their children's total expected earnings, hence allocating more resources to children with higher returns to inputs (Behrman *et al.*, 1982).

Importantly, the comparison of parental investments devoted to children with different initial endowments that sibling or twin fixed-effects models provide only allows one to identify the composite impact of parental preferences regarding equality *and* the price effect. Actually, Terskaya (2019) shows that even inequality averse parents might reinforce differences between their children if the cost of investing in them is sufficiently lower for higher endowed children than for their lower-endowed siblings. In other words, neither following a compensating strategy necessarily implies that parents are inequality averse nor following a reinforcing strategy necessarily implies that parents only care about efficiency.

2.1 Parental preferences regarding equality versus the price effect

We propose an alternative empirical strategy to siblings and twins fixed-effect models that allows one to disentangle the effect of parental preferences regarding equality in the distribution of their children's quality from the price effect. Our strategy involves identifying the impact on parental investment decisions of children's relative (with respect to their siblings) genetic endowments predictive of cognitive outcomes while holding children's own (absolute) genetic endowments constant (that is, by holding prices or parents' costs of adding to their children's quality constant). In particular, we estimate the following model:

$$PI_{if} = \beta_0 + \beta_1 * EPGS_{if} + \beta_2 * (EPGS_{if} - EPGS_{jf}) + X'_{if}\alpha + S'_{jf}\delta + F'_f\gamma + u_{if} \tag{1}$$

, where PI_{if} is a parental investment indicator for child i in family f; $EPGS_{if}$ stands for child i's education polygenic score (that is, our measure of the absolute educational genetic endowment for child i); $EPGS_{jf}$ denotes the education polygenic score of child j, with subscript j denoting child i's sibling; X'_{if} and S'_{jf} are vectors of individual-level characteristics of children

i and j, respectively, that may affect parental investment decisions; and F'_f is a vector of family-level characteristics (shared by children i and j) that may also influence parental investment choices. Note that $(EPGS_{if} - EPGS_{jf})$ is our indicator of child i's relative cognitive genetic endowment, as it is the difference between i's endowment and his/her sibling j's endowment. Although we have so far generally referred to i and j in equation (1) as "siblings", throughout this section we will distinguish between nontwin siblings and twin siblings, as identifying the effect of interest presents more challenges in the case of nontwin siblings .

As we are controlling for child i's own (absolute) endowment ($EPGS_{if}$), β_2, which is our main coefficient of interest, measures the effect of parental preferences regarding equality in the distribution of children's quality on parental investment decisions. For any given level of child i's endowment ($EPGS_{if}$), $\beta_2 < 0$ is consistent with parental inequality aversion, as it indicates that i's parents will invest less (more) in him/her if child i is higher-endowed (lower-endowed) than his/her sibling j. In contrast, $\beta_2 > 0$ is consistent with parents valuing more efficiency than equality, as they invest more (less) in child i if his/her endowment is higher (lower) than that of his/her sibling j. Finally, $\beta_2 = 0$ is consistent with parents having neutral preferences regarding equality in the distribution of their children's quality.

As for β_1 in equation (1), this parameter is informative on the price effect or parents' costs of adding to their children's quality. In particular, $\beta_1 > 0$ would imply that, for any given level of inequality in siblings' endowments $(EPGS_{if} - EPGS_{jf})$, parents invest more in children whose own (absolute) endowment ($EPGS_{if}$) is higher. Note that, since $(EPGS_{if} - EPGS_{jf})$ is held constant in (1), a positive value of β_1 could not be attributed to parental preferences for efficiency over equality.[4]

In the following sections we highlight the identification challenges involved in the estimation of (1) and we discuss how we address them.

2.2 Reverse causality

An important challenge faced by studies analysing the effect of children's endowments on parental investment decisions is reverse causality. Even within families, endowment indicators measured during childhood may be the result of prior parental (both post- and pre-natal) investments, while the endowment indicators measured at birth often used (such as birthweight) may be the consequence of pre-natal investment decisions. This not an issue with educational polygenic scores, as individuals' genetic makeup is fixed at conception.

2.3 Unobserved parental genes

Despite the fact that genetic lotteries occur within families (Fletcher and Lehrer, 2011; Domingue *et al.*, 2015), parental genes (which we do not observe) affect children's genes as well as (potentially) parental investments. However, the fact that we observe both siblings' genes allows

[4] It is worth stressing that, in order to interpret β_1 as a price effect, one must hold sibling differences in endowments $(EPGS_{if} - EPGS_{jf})$ constant, and not just the absolute endowment of each child's sibling ($EPGS_{jf}$) because parents may also respond to inequalities among their children.

Table 1: Parental Preferences for Children's Quality and Fertility Decisions

Endowment of 1st child	High				Low			
Parental preferences for high ability children	Strong		Indifferent		Strong		Indifferent	
Decision to have a 2nd child	no	no	maybe	maybe	yes	yes	maybe	maybe
Endowment of 2nd child (relative to the 1st)			higher	lower	higher	lower	higher	lower

us to estimate their correlation with parental genes. This is because children's genes are a function of parental genes plus some random component which is uncorrelated across siblings.Hence, the only source of correlation between siblings' genes are parental genes. This allows us to compute the magnitude of the bias of $\hat{\beta}_2$ due to the omission of parental genes. The analytical derivation, included in Appendix A, indicates that, in the worst case scenario, the true β_2 will be about 60% of our estimated $\hat{\beta}_2$.

Note that we also control for parental characteristics (e.g., parental socioeconomic status, education, etc.) that likely reflect parental genes and our results barely change. This suggests that, in practice, we are probably not too close to the worst scenario previously described.

Additionally, we also use an indicator of parental investment that is relative (capturing differences across siblings) rather than absolute (i.e. focused on the investment allocated to each child). The advantage of this indicator is that it measures relative parental investments and therefore it should be unaffected by factors shared by siblings, such as parental socioeconomic status (which we control for) or parental genes (which we do not observe). As shown in Section 3 our results are robust to using this relative parental investment measure as an outcome.

2.4 Endogenous fertility

There is still an additional issue that studies analysing parental responses to children's endowment differences must confront: the potential endogeneity of fertility. If fertility were exogenously fixed or randomly allocated one could compare (regardless of birth order) the parental investments made in equally endowed children with differently endowed siblings. However, parents' decisions to have more children may depend on the endowments of previous children. While this is not an issue for the analysis of twins, for whom we will estimate equation (1) as it is, it may well be a problem for analyses based on nontwin siblings ' comparisons.

In fact, Ejrnæs and Pörtner (2004) show that parents who strongly prefer children with high genetic endowments will stop having children earlier if they already have a high ability child. In contrast, if parents are indifferent towards their children's endowments, their decision to keep on having children will be independent of the endowments of their previous children.

Table 1 illustrates that highly endowed children with highly endowed older siblings were born to parents who are indifferent towards their children's endowments (indifferent parents, for short). In contrast highly endowed children with low-endowed older siblings could have been born both to parents with strong preferences for high ability children or to indifferent parents. Therefore, the comparison of second-born children with the same absolute level of ability but who differ in terms of their sibling's endowments (or, in other words, who differ in terms of their ability relative to that of his/her siblings) is complicated by the fact that these children are born to parents with systematically different preferences regarding their offsprings' endowments. On the bright side, Table 1 also illustrates that first-born children with the same absolute endowment levels but who differ in terms of their sibling's endowments are born to parents with similar preferences. As a consequence, one can circumvent the endogenous fertility issues that affect the analysis of nontwin siblings by focusing on first-born children while conditioning on their absolute endowment levels. This gives the following version of equation (1):

$$PI_{1f} = \beta_0 + \beta_1 * EPGS_{1f} + \beta_2 * (EPGS_{1f} - EPGS_{2f}) + X'_{1f}\alpha + S'_{2f}\delta + F'_f\gamma + u_{1f} \qquad (2)$$

, where subscripts i and j have been replaced by subscripts 1 and 2, with 1 referring to first-born children and 2 denoting their next younger siblings.

Hence, we estimate equation (1) for a sample of twins, and equation (2) for a sample of first-born children with at least one younger sibling. Note also that analysing the investment decisions of parents of twins and nontwin siblings separately is advisable. This is because previous studies suggest that parents of twins are less likely to be responsive to their endowment differences (Bharadwaj *et al.*, 2018) if, for instance, they find it more costly to implement favouritism among their twin children than parents of nontwin siblings (Almond and Mazumder, 2013).[5]

3 Data and Descriptive Statistics

3.1 The Add Health Dataset

We use data from The National Longitudinal Study of Adolescent to Adult Health (Add Health in what follows), which is a nationally representative longitudinal survey of U.S. 7^{th} to 12^{th} graders during the school year 1994/95 drawn from a stratified sample of 80 high schools and 52 middle schools. Within each school and grade, a random sample of approximately 17 males and 17 females, as well as an oversample of siblings and specific minorities were selected in 1994/95 (hereafter Wave I with No. Obs.= 20,745, ages 12–20 years), which constitutes the so called in-home sample. Subsequent interviews were conducted in 1996 (hereafter Wave II with No. Obs.= 14,738, ages 13-21 years), in 2001/02 (hereafter Wave III with No. Obs.= 15,197, ages 18-26 years) and in 2008 (hereafter Wave IV with No. Obs.=15,701, ages 24-32 years).

[5] Abufhele *et al.* (2017) also find that parents in Chile do not invest differently in twins with different birth weight.

The in-home survey of Add Health collects information on respondents' behaviours during adolescence and early adulthood, as well as information on their relationship with their parents and siblings, which allows us construct indicators of parental investments and parental favouritism. Another crucial advantage of using Add Health data in our analyses is that this dataset also contains extensive genetic information for the sample of siblings, which provides us with indicators for their genetic endowments.

3.2 Genome-Wide Data to Measure Genetic Endowments

The Add Health sibling pairs sample was genotyped via Oragene saliva collection with the Illumina Human Omni Quad chip at Wave IV of the study (see McQueen *et al*. 2015 for details). The siblings' genetic database included 1,886 individuals with valid data on 940,862 single nucleotide polymorphisms (SNPs),[6] which were subsequently used to construct (among others) a single indicator that predicts educational attainment. We will mainly use the term educational polygenic score (EPGS in equations (1) and (2) in Section 2) to refer to this indicator.[7]

Polygenic scores summarize the genetic propensity of an individual to a particular trait. The approach Add Health used to calculate polygenic scores is based on recent advances in genetics that rely on genome-wide association studies (GWAS). GWAS analyse the association between an outcome of interest (a phenotype) and each of a large number of SNPs through a data-mining approach (see Belsky and Israel 2014 for details). In particular, GWAS consist in regressing an outcome of interest (such as years of schooling) against a very large number of individual SNPs, and adopting conservative p-value thresholds for identifying genome-wide significant associations.

The first large-scale GWAS of educational attainment was conducted by Rietveld *et al*. (2013), and it analysed data on more than 100,000 individuals. Rietveld *et al*. (2013) identified several SNPs that were strongly associated with educational attainment even after strict adjustments for multiple hypothesis testing aimed at avoiding finding false significant results.[8] SNPs in the Add Health Sibling Pairs genetic database were matched to the SNPs analysed in Rietveld *et al*. (2013) and, for each of these SNPs, a score was calculated as the number of education associated alleles multiplied by the corresponding effect-sizes estimated in the original GWAS.

Many of these SNPs are likely to be involved in biological processes related to cognitive processes, such as learning and long-term memory, and neuronal development or function, which suggests that the EPGS is closely related to cognitive ability. Hence, we will use the terms EPGS, "cognitive polygenic score", and "cognitive genetic endowment" interchangeably. Domingue *et al*. (2015) have already shown that the EPGS based on the results of Rietveld *et al*. (2013) predicts educational attainment in the AddHealth sample. In Section 3.5 we con-

[6]A SNP is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population.

[7]Polygenic scores are also frequently referred to as polygenic risk scores, genetic risk scores, or genome-wide scores.

[8]Rietveld *et al*. (2014) have replicated these results.

firm that this is indeed the case and, we show that the EPGS is also a strong predictor of individuals' grades as well as their scores on the the Peabody Picture Vocabulary test (PPVT).

3.3 Parental Investments

We construct several alternative measures of parental investments. Our first set of measures is based on questions about teenagers' relationship with their parents included in the in-home questionnaire administered in Wave I. Adolescents were asked similar questions about their relationship with their mother and their father. In particular, we consider the following binary outcomes: *i) In the past 4 weeks went to a movie, play, museum, concert, or sports event with the mother (father); ii) In the past 4 weeks had a talk about a personal problem were having with the mother (father); iii) In the past 4 weeks talked about school work or grades with the mother (father); and iv) In the past 4 weeks worked on a project for school with the mother (father); In the past 4 weeks talked about other things were doing in school with the mother (father).* Using these variables, we construct three indicators: a parental investment index based on questions involving both parents; a maternal investment index based on questions involving the mother; and a paternal investment index based on questions involving the father.[9]

To construct summary indexes, we follow Kling *et al.* (2007). Each summary index variable, $Y*$, is constructed as the unweighted average of all standardized outcomes:

$$Y* = \frac{\sum_k Y*_k}{K} \text{ , where } Y*_k = \frac{Y_k - \mu_k}{\sigma_k}$$

, and Y_k is the k^{th} component of the index, μ_k denotes its mean and σ_k its standard deviation.

Additionally, we also use a measure of parental favouritism or relative parental investment based on the following question addressed to Wave I respondents from the sibling sample about each sample sibling:[10] *"Think of all the things your parents do for you and NAME. Do you think that you or your NAME receive more attention and love from your parents? Would you say that NAME receives: 1 a lot more; 2 a little more; 3 the same amount; 4 a little less; 5 a lot less".* Therefore, the variable takes a higher value if the respondent believes that his/her parents favour him/her more than his/her sibling. The advantage of this indicator is that it measures relative parental investments and therefore it should be unaffected by factors shared by siblings, such as parental socioeconomic status or parental genes. We will use the terms "favouritism indicator" or "relative parental investment" to refer to this variable in what follows.

3.4 Sample Restrictions

For our analysis we use the Add Heath Sibling Pairs sample with available EPGS, which includes 1,886 individuals from 1,113 families. In 380 families only one sibling was genotyped,

[9] In only one parent is present in the household the parental investment index only includes information regarding the teenagers' relationship with him/her.

[10] By sample sibling we mean a sibling who was also interviewed in Wave I of Add Health. The question is asked as many times as sample siblings an individual has. If the respondent has more than one sample sibling, we take an average of the answers.

Table 2: Summary Statistics

	First-borns		Twins	
	Mean	Std. Dev.	Mean	Std. Dev.
EPGS. Normalized	0.000	1.000	0.000	1.000
Sibling's EPGS. Normalized	0.066	0.981	0.000	1.000
EPGS -Sibling's EPGS. Normalized	-0.066	0.883	0.000	0.830
EPGS is higher than sibling's EPGS	0.474	0.500	0.500	0.501
Age	17.187	1.298	15.842	1.627
Age -Sibling's age	2.089	1.100	0.000	0.000
Female	0.513	0.500	0.500	0.501
Sibling is female	0.536	0.497	0.500	0.501
Rural	0.297	0.456	0.204	0.404
Black	0.267	0.443	0.219	0.414
White	0.541	0.499	0.650	0.478
Sibling is white	0.546	0.498	0.650	0.478
Sibling is black	0.261	0.439	0.219	0.414
Total family income before tax 1994. In hundred thousands	0.427	0.540	0.460	0.358
Resident parent college graduate	0.198	0.399	0.219	0.414
Both parents live in hh	0.689	0.463	0.715	0.452
SES index (normalized)	0.000	1.000	0.000	1.000
N. Observations	595		274	

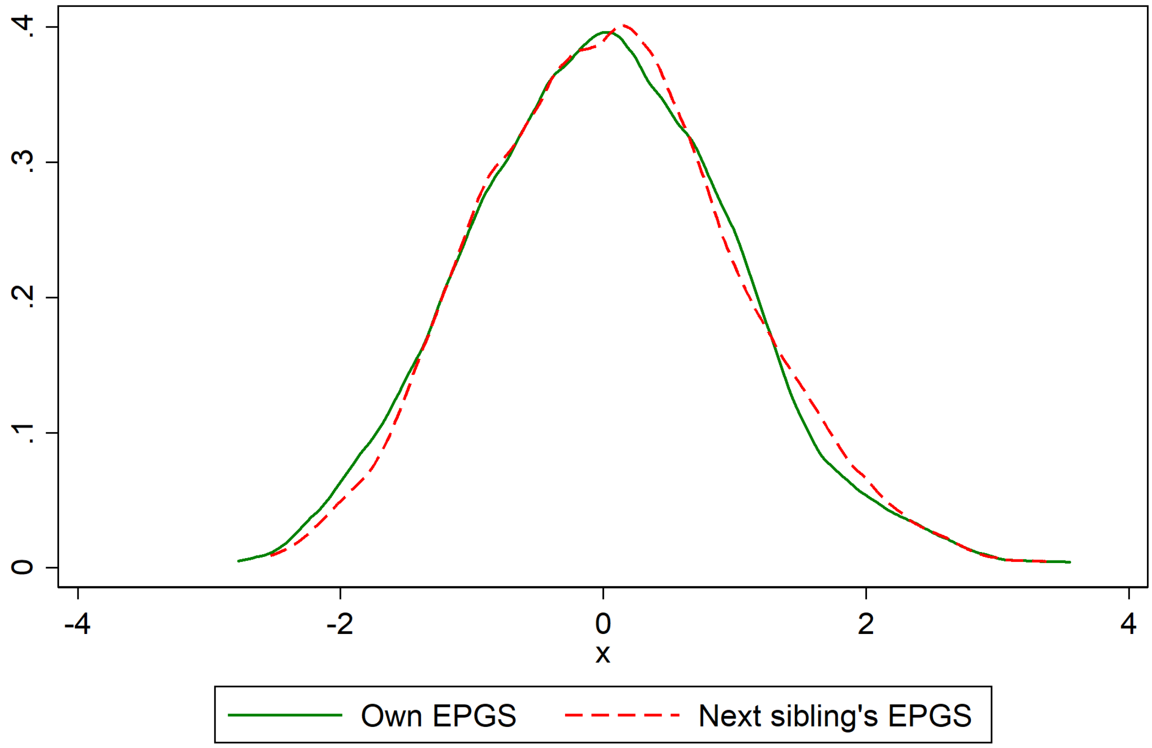
Note: EPGS is the educational polygenic score provided by AddHealth for the sibling sample. Normalized variables have mean 0 and standard deviation 1.

Table 3: Educational Polygenic Scores, Educational Attainment and PPVT Scores

	(1)	(2)	(3)	(4)
Dependent variable:	Years of education	PPVT	Overall GPA	High-School Drop-Out
EPGS. Normalized	0.421***	2.515***	0.184***	-0.0349***
	(0.0801)	(0.460)	(0.0364)	(0.0107)
Observations	869	833	645	869
R-squared	0.061	0.205	0.188	0.022
Incremental R-squared	0.027	0.024	0.033	0.011
Incremental R-squared (% of R-squared)	44%	12%	18%	50%

Note: OLS coefficient estimates and their associated robust standard errors in parentheses. All regressions include the following controls: age, age squared, a female dummy, and race indicators. EPGS is the educational polygenic score provided by AddHealth for the sibling sample. It is normalised to have mean 0 and standard deviation 1. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Figure 1: Education Polygenic Score (Normalized). Kernel Density Estimate



Note: Polygenic scores are computed by AddHealth in Wave IV. First-born and twins are included. No.obs: 869

but our aim is to study how parental investment decisions are affected by the existence of differences across their children's endowments. Hence, we drop these 380 observations, which leaves us with 732 families (595 with nontwin siblings and 137 with twins). We estimate equation (1) using the sample of twins (No. Obs.=274), and we estimate equation (2) using the sample of first-borns from with singleton children.

3.5 Descriptive Statistics

Table 2 separately describes the samples of first-born singletons and twins (hereafter first-borns' sample and twins' sample). Note that we standardize EPGS so it has mean 0 and the standard deviation 1 in each separately analysed sample.

Figure 1 presents a plot of the (kernel-smoothed) densities of each teenager's own EPGS and his/her next younger sibling's (or twin's) EPGS in our sample of first-borns and twins.[11] The distribution of Add Health respondents' EPGS is approximately normal, and it does not significantly vary by birth order

Table 2 (which displays summary statistics for the siblings' EPGS and for all the control variables used in our analyses) shows that, on average, first-borns in our sample have slightly lower EPGS (0.066 standard deviations lower) than their next younger siblings, but this difference is not significant at standard levels of testing.

[11]In the first-borns' sample, the next younger sibling's EPGS is rescaled using the mean and the standard deviation of first-borns' EPGS, so that the scales of the two resulting variables are comparable.

First-borns in our sample are on average 17 years old, while twins are 16 years old on average. The next younger siblings of first-born teenagers are on average 2 years younger. Approximately 50% of both first-borns and twins are female. In the first-borns' sample, 29.7% live in a rural area, 26.7% are black and 54.1% are white; these figures are 20.4%, 21.9%, and 65.0%, respectively, in the sample of twins. 19.8% of first-borns and 21.9% of twins have at least one college educated parent, and the majority of them live with both parents (68.9% of first-borns and 71.5% of twins).

To measure parental socio-economic status (SES) we construct an index based on parental education, parental occupation prestige, household income, and household receipt of public assistance following Belsky *et al.* (2018), and we also standardize it so that it has mean 0 and standard deviation 1 in each sample.[12] We describe the construction of this index in Appendix A.

Table 3 shows that, as expected, individuals EPGS' have a strong association with educational attainment (years of schooling), grades, and PPVT scores in our sample, even after controlling for individual and family characteristics, such as family SES. We find that a 1 standard deviation increase in the EPGS is associated with 0.3 additional years of education, a result that is in line with the estimates from Domingue *et al.* (2015).

Table 4 describes the main outcomes measuring parental investments in the two samples. 25.5% and 20.3% of first-borns and twins, respectively, think that they receive less attention or love from their parents than their siblings. While only 12% (15%) of mothers of first-borns (twins) worked on a project with their child, 65.3% (69%) talked about school with him/her in the past 4 weeks. Table 4 also suggests that mothers are more likely than fathers to provide inputs to their children.

Finally, Table 5 presents balancing tests, that is, the results of regressing each of our control variables X', S' and F' in equations (1) and (2) on $EPGS_{1f} - EPGS_{2f}$ for the sample of first-borns (see equation (2)) and on $EPGS_{if} - EPGS_{jf}$ (see equation (1)) for the sample of twins. Consistent with the idea that genetic variation across siblings resembles a lottery, none of these associations are significant.

4 Main Results

Our main results are displayed in Table 6. Columns 1 and 2 show estimates of our main coefficient of interest, $\hat{\beta}_2$, as well as of $\hat{\beta}_1$, obtained from estimating equation (2) using the sample of first-born children (with and without covariates). Columns 3 and 4, in turn, focus on the sample of twins and display coefficient estimates obtained from estimating equation (1) with and without covariates, respectively. As expected, adding covariates barely alters our coefficient estimates, which is consistent with our previous balancing tests results and with genetic variation across siblings being as good as random. We will first discuss our results for nontwin siblings and then move on to the evidence for twins.

Regardless of the parental investment indicator used, our results indicate that parents of

[12] Belsky *et al.* (2018) construct a similar score using Add Health data to study social-class mobility.

Table 4: Summary of Outcomes

	First-borns			Twins		
	N	Mean	SD	N	Mean	SD
Thinks that receives less than sibling	423	0.255	0.437	217	0.203	0.403
Thinks that receives the same as sibling	423	0.686	0.465	217	0.760	0.428
Thinks that receives more than sibling	423	0.059	0.236	217	0.037	0.189
Favouritism. Normalized	423	0.000	1.000	217	-0.000	1.000
Parental Investment Index	583	0.000	0.496	272	0.057	0.519
Maternal Investment Index	568	-0.000	0.595	265	0.082	0.602
Paternal Investment Index	411	-0.000	0.680	195	0.048	0.725
Parental Investment Indexes Components						
Maternal Investment Index Components						
Attended cultural/sports event with mother. W1	568	0.236	0.425	265	0.298	0.458
Talked about a personal problem with mother. W1	568	0.423	0.494	265	0.385	0.487
Talked about school with mother. W1	568	0.653	0.476	265	0.691	0.463
Worked on a project with mother. W1	568	0.120	0.325	265	0.151	0.359
Talked about other school things with mother. W1	568	0.546	0.498	265	0.623	0.486
Paternal Investment Index Components						
Attended cultural/sports event with father. W1	411	0.212	0.409	195	0.277	0.449
Talked about a personal problem with father. W1	411	0.195	0.396	195	0.185	0.389
Talked about school with father. W1	411	0.540	0.499	195	0.533	0.500
Worked on a project with father. W1	411	0.097	0.297	195	0.133	0.341
Talked about other school things with father. W1	411	0.499	0.501	195	0.477	0.501

Note: Favouritism is a categorical variable that takes values ranging from from 1 (thinks that sibling receives a lot more attention) to 5 (thinks that he/she receives a lot more attention than sibling). It is normalised to have mean 0 and standard deviation 1. W1 stands for Wave I of AddHealth.

nontwin siblings display inequality aversion because $\hat{\beta}_2$ is always negative and statistically significant at standard levels of testing (Columns 1 and 2, second row of all panels). That is, after conditioning on each first-born child's own absolute endowment level (as measured by his/her genetic predisposition for education or educational polygenic score, $EPGS_1$), parents invest less (more) in him/her if he/she is better (worse) endowed than his/her sibling. We find that if sibling differences in their endowments ($EPGS_1$-$EPGS_2$) increase by one standard deviation, parental, maternal and paternal investments decrease in the better-endowed child by 0.13 (Panel A, Column 2, first row), 0.10 (Panel B, Column 2, first row) and 0.15 (Panel C, Column 2, first row) standard deviations of the corresponding investment indexes. These effects are statistically significant, but let us now put their magnitude in perspective. We find that these effects are not only statistically significant but quantitatively relevant too because they represent a 65%, a 48%, and a 72% of the (positive) impacts that a standard deviation increase in families' socioeconomic status has on parental, maternal and paternal investments in their children, respectively.[13]

[13]We estimate that one standard deviation increase in family socioeconomic status is associated with about a

Table 5: Balancing Tests. Correlations between Educational Polygenic Score Differences between Siblings and Individual and Household Characteristics

	First-borns		Twins	
	Coefficient	SE	Coefficient	SE
Age	0.0616	0.0531	1.09e-08	0.102
Age squared	2.101	1.787	3.42e-07	3.259
Age - Siblings' age	0.0501	0.0469	-	-
Female	0.000745	0.0206	-0.00987	0.0303
Rural	0.00255	0.0205	0.00987	0.0303
Black	-0.0111	0.0187	0.0124	0.0199
White	-0.000601	0.0176	-3.50e-09	0.0213
Sibling is female	-0.00541	0.0204	1.88e-09	0.0282
Sibling is white	-0.00881	0.0204	-	-
Sibling is black	0.00163	0.0173	-	-
Total family income before tax 1994. In hundred thousands	-0.0428	0.0307	-0.00310	0.0206
Resident parent college graduate	0.00127	0.0172	-6.79e-10	0.0253
Both parents live in hh	0.00285	0.0189	2.54e-09	0.0260
SES index (normalized)	-0.00760	0.0431	-0.00446	0.0627
Observations	595		274	

Note: The table displays OLS coefficients and their associated robust standard errors obtained after regressing each variable on sibling differences in EPGS (normalized to have mean 0 and standard deviation 1). All individual and family characteristics are measured in Wave I of AddHealth. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

In section 2.3 we have acknowledged that we do not observe parental genes. Genes are randomly allocated across siblings, so $EPGS_1 - EPGS_2$ is uncorrelated with parental genes. However, $EPGS_1$ is likely correlated with parental genes. Since $EPGS_1$ and $EPGS_1 - EPGS_2$ are correlated, if parental genes directly affect parental investments (even after controlling for $EPGS_1$, $EPGS_1 - EPGS_2$ and other covariates), the omission of parental genes in (2) may bias both $\hat{\beta}_1$ and $\hat{\beta}_2$. However, as shown in Appendix A, even in the worst-case bias scenario induced by the omission of parental genes, not only we would consistently estimate the sign of β_2, but its magnitude would still be sizeable, as it would amount to a 58%, a 62%, and a 54% of the true effects for the parental, maternal, and paternal indexes, respectively, which would in turn translate into a 38%, a 23%, and a 39% of the (positive) impacts that a standard deviation increase in families' socioeconomic status has on these parental investment indicators. Crucially, the result obtained when using the favouritism indicator, which captures parental relative investment decisions, confirms that parents display inequality aversion. This is reassuring because this indicator is relative and hence should be unaffected by family-level unobserved factors shared by siblings (such as parental genes). In particular, we

0.2 standard deviation increase in our parental investment indicators. These estimates, not reported for ease of exposition, are available upon request from the authors.

Table 6: The Effect of Sibling Differences in Educational Polygenic Scores on Parental Investments

	First-Borns		Twins	
	(1)	(2)	(3)	(4)
Panel A: Parental Investment Summary Index. Normalized				
EPGS - Sibling's EPGS. Normalized	-0.122***	-0.129***	0.0282	0.0807
	(0.0446)	(0.0475)	(0.0722)	(0.0713)
EPGS. Normalized	0.116***	0.122**	0.0574	-0.0640
	(0.0435)	(0.0522)	(0.0625)	(0.0755)
Observations	583	583	272	272
R-squared	0.015	0.072	0.005	0.102
Panel B: Maternal Investment Summary Index. Normalized				
EPGS - Sibling's EPGS. Normalized	-0.0987**	-0.102**	0.00387	0.0411
	(0.0456)	(0.0493)	(0.0678)	(0.0659)
EPGS. Normalized	0.0852*	0.0860	-0.0307	-0.117
	(0.0456)	(0.0545)	(0.0617)	(0.0750)
Observations	568	568	265	265
R-squared	0.009	0.071	0.001	0.119
Panel C: Paternal Investment Summary Index. Normalized				
EPGS - Sibling's EPGS. Normalized	-0.138**	-0.146**	0.0436	0.129
	(0.0542)	(0.0568)	(0.0820)	(0.0841)
EPGS. Normalized	0.142***	0.141**	0.155*	-0.0442
	(0.0517)	(0.0583)	(0.0787)	(0.0947)
Observations	411	411	195	195
R-squared	0.020	0.069	0.031	0.125
Panel D: Favouritism Indicator (Relative Parental Investment). Normalized				
EPGS - Sibling's EPGS. Normalized	-0.115**	-0.137***	-0.00359	0.0690
	(0.0512)	(0.0523)	(0.0778)	(0.0750)
EPGS. Normalized	0.128**	0.155***	0.0366	-0.133
	(0.0511)	(0.0573)	(0.108)	(0.109)
Observations	423	423	217	217
R-squared	0.016	0.066	0.001	0.059
Individual and family controls		YES		YES

Note: OLS coefficient estimates and their associated robust standard errors in parentheses. The regressions in Columns (2) and (4) include the following controls: age, age squared, a female dummy, race indicators, a rural area dummy, total family income, an indicator for whether at least one parent is a college graduate, an indicator for whether both parents live in the household, and a socoeconomic status index. EPGS is the educational polygenic score provided by AddHealth for the sibling sample (normalised to have mean 0 and standard deviation 1). EPGS - Sibling's EPGS is the difference in EPGS between siblings (that is, between first-borns and their next younger sibling in the sample of first-borns, and between twins in the twins sample), also normalized to have mean 0 and standard deviation 1. *** $p < 0.01$, $**p < 0.05$, $*p < 0.1$

find that if sibling differences in their endowments ($EPGS_1$-$EPGS_2$) increase by one standard deviation, the favouritism indicator decreases for the better-endowed child by 0.14 standard deviations (Panel D, Column 2, first row). Hence, the magnitude of our estimated parental inequality aversion parameter for the sample of firstborns is remarkably similar for the relative favouritism indicator and for the other three parental investment indicators. [14] Additionally, the inclusion of other family-specific controls shared by siblings observed in our data such as parental education, family income and socioeconomic status leaves our estimates of both β_1 and β_2 virtually unaltered. All this is suggestive evidence that the bias potentially induced by the omission of parental genes is unlikely to be large.

Interestingly, the coefficient on each first-born child's own educational polygenic score (Columns 1 and 2, second row of all panels) is in general positive, sizeable and statistically significant, suggesting that parental costs of adding to their children's quality matter, as they invest significantly more the better-endowed the child is. For instance, if a child's educational polygenic score increases by one standard deviation (holding constant his/her endowment difference with respect to his/her sibling), the parental investment indicator increases by about 0.12 standard deviations (Panel A, second row). Note however that we are cautions about giving $\hat{\beta}_1$ a causal interpretation because one cannot bound the extent of its potential bias due to the omission of parental genes as we did with $\hat{\beta}_2$ in Appendix A. However, there are two main reasons why we believe that our evidence clearly supports the notion that the price effect is positive and parents find it less costly to invest in better-endowed children. First, the estimate of β_1 we obtain when using our relative measure of parental investment (Panel D, Columns 1 and 2, second row) is also positive and similar in magnitude to the estimates obtained when using the other three parental investment indicators (Panels A, B, and C, Columns 1 and 2, second row). Second, as it was the case with $\hat{\beta}_2$, our results for $\hat{\beta}_1$ barely change when we add observed family characteristics shared by siblings to the model, as the comparison between the second row of Columns 1 and 2 of all panels of Table 6 reveals.

Our finding that the price effect is likely positive implies that, if it is large enough, even inequality averse parents may choose to follow a reinforcing or a neutral strategy. This is a relevant result for the literature on intra-household resource allocation, as it may explain why many previous empirical studies relying on family fixed-effects models have found that parents often follow a reinforcing strategy (Datar *et al.*, 2010; Aizer and Cunha, 2012; Hsin, 2012; Frijters *et al.*, 2013; Rosales-Rueda, 2014).[15]

In sum, our evidence for nontwin siblings is clearly supportive of parents displaying in-

[14] This is consistent with the findings of Agostinelli and Wiswall (2016), who show that maternal cognitive ability has no significant impact on how much parents invest in their children after controlling for family income and for children's cognitive ability (see their Table 3). They measure children's cognitive skills using several sub-scales of the Peabody Individual Achievement Test (PIAT) and the Peabody Picture Vocabulary Test (PPVT), while their maternal ability measure is based on sub-scales of the Armed Services Vocational Aptitude Battery (ASVAB).

[15] Actually, if we use our sample of nontwin siblings to estimate a family fixed-effects model such as:

$$PI_{if} = \alpha_0 + \alpha_1 * EPGS_{if} + X'_{if}\delta + \rho_f + u_{if} \text{ ,where } \rho_f \text{ is a family fixed effect.}$$

We find that parents follow a neutral strategy. These results are available upon request form the authors.

equality aversion such that, for a given level of a child's genetic predisposition for education, they reallocate resources to invest more in him/her if his/her sibling is better-endowed. This finding is very much in contrast with our evidence for twins, whose parents instead display neutral preferences regarding the distribution of quality among their children (see Table 6, Columns 3 and 4, second row). This is also in line with Bharadwaj *et al.* (2018), who find that parents of nontwin siblings in Chile follow a compensating strategy regarding initial health, while parents of twins are not responsive to their endowment differences in health at birth. [16]. One potential explanation for our contrasting results for parents of twins and of nontwin siblings might be that it can be difficult for the former to invest differently across their children because they are exactly the same age (Almond and Mazumder, 2013). Indeed, Bharadwaj *et al.* (2018) lay out a model of human capital accumulation and parental investments that incorporates as a novel component a public good and spillovers dimension in the provision of parental investments within the household, which is likely to be greater for children who are very close in age, and hence provides one rationalization for the differences observed between parents of twins and nontwin siblings. For instance, if a parent helps out with homework or plays with one twin it is difficult to prevent the other twin from participating to some extent. In our context, this implies that parents of twins may also be inequality averse, but they may be unable to invest differentially across their children even if they wished to.

Finally, we investigate whether the degree of inequality aversion we have previously uncovered for parents of nontwin siblings varies along the parental investment distribution. Table 7 and Figure 2 summarize the results of estimating equation (2) using unconditional quantile regression methods. The results indicate that parents of nontwin siblings start displaying inequality aversion shortly before the median of the parental investment index distribution. That is, "low investors" do not significantly react to endowment differences across their children, while "high investors" do.[17]

[16]Note however that Bharadwaj *et al.* (2018) differs from our work in several important ways. First, they use twin and fixed-effects models, which do not allow one to isolate the impact of parental preferences regarding inequality from the price effect. Second, they rely on birthweight as a measure of children's endowments, and variation in birthweight (like genetic variation) cannot be the consequence of pre-natal parental investment decisions, but it could (unlike genetic variation, as genes are fixed at conception) be the consequence of post-natal investment decisions. Additionally, our endowment indicator (educational polygenic scores) is more directly and strongly linked to cognitive outcomes.

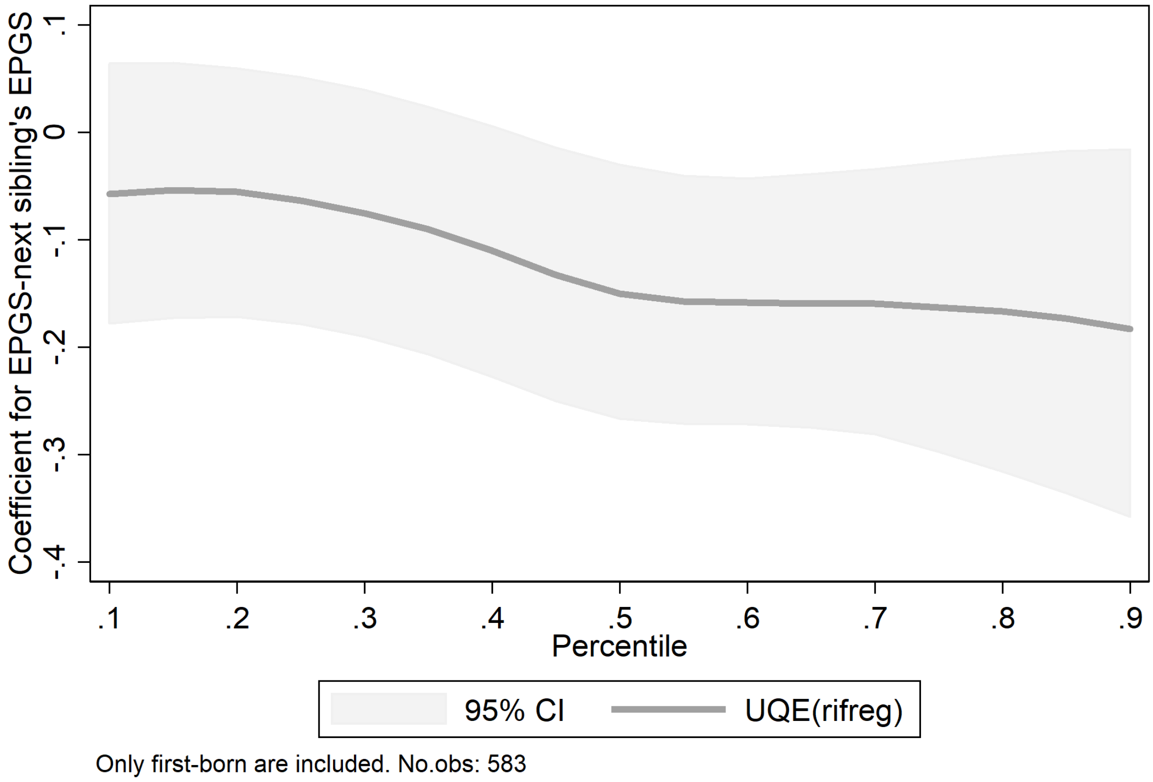
[17]Note however that Table 7 and Figure 2 display results only results for our parental investment index and for parents of nontwin siblings. The pattern of results for the other three investment indicators we have used is the same. We also find that parents of twins display neutral preferences not only at the mean, as shown in Table 7, but all along the parental investment distribution regardless of the investment indicator used.

Table 7: The Effect of Sibling Differences in Educational Polygenic Scores along the Distribution of Parental Investments Index (Normalized). Unconditional Quantile Regressions. Firstborns Only.

	(1) 15th Centile	(2) 30th Centile	(3) 40th Centile	(4) 50th Centile	(5) 60th Centile	(6) 70th Centile	(7) 80th Centile	(8) 90th Centile
EPGS - Sibling's EPGS. Normalized	-0.0806	-0.0247	-0.153**	-0.159***	-0.188***	-0.124**	-0.162**	-0.232**
	(0.0533)	(0.0585)	(0.0616)	(0.0567)	(0.0545)	(0.0582)	(0.0748)	(0.0983)
EPGS. Normalized	0.0324	0.0305	0.110	0.148**	0.207***	0.212***	0.223***	0.230**
	(0.0622)	(0.0673)	(0.0675)	(0.0619)	(0.0603)	(0.0640)	(0.0846)	(0.110)
R-squared	0.097	0.078	0.084	0.071	0.061	0.073	0.051	0.061

Note: Unconditional quantile regression estimates and their associated standard errors in parentheses. All regressions include the following controls: age, age squared, a female dummy, race indicators, a rural area dummy, total family income, an indicator for whether at least one parent is a college graduate, an indicator for whether both parents live in the household, and a socioeconomic status index. EPGS is the educational polygenic score provided by AddHealth for the sibling sample (normalised to have mean 0 and standard deviation 1). EPGS - Sibling's EPGS is the difference in EPGS between siblings (that is, between first-borns and their next younger sibling in the sample of first-borns, and between twins in the twins sample), also normalized to have mean 0 and standard deviation 1. No. observations: 583. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Figure 2: The Effect of Sibling Differences in Educational Polygenic Scores along the Distribution of Parental Investments.



5 Falsification Tests

5.1 Placebo Tests

In order to check that our results are not driven by chance we run placebo tests. First, we match each firstborn's EPGS ($EPGS_1$) with a randomly chosen EPGS of a second born child ($EPGS_2^P$) and compute the difference $EPGS_1 - EPGS_2^P$. Then, using this placebo variable we estimate our main specification for firstborns. [18] We repeat this this procedure 500 times in order to obtain a distribution for the estimate of the coefficient of $EPGS_1 - EPGS_2^P$. We find that, in line with our results being genuine, the coefficient of $EPGS_1 - EPGS_2^P$ is significant at the 5% level in less than 5% of our placebo regressions.

5.2 "Too Early" Parental Responses

Parents cannot possible observe differences between children when the firstborn child is too young and their second-born child is unlikely to have been born yet. Therefore, another placebo test of our main result consists in checking whether differences in siblings' EPGS affect "too early" early parental investments, such as breastfeeding. We use a retrospective question from the Add Health parental questionnaire that asked mothers how long each of

[18]We conduct this test only for firstborns since the results for twins are not significant.

their children participating in the in-home interview was breastfed. We define an indicator variable which takes the value zero if the mother's answer is *"(He/ she) was not breastfed"* and one if she reports that the child was breastfed to some extent. Then we estimate equations (1) and (2) using this variable as an outcome. As expected, we do not find any significant effect of sibling differences in EPGS on the probability of having been breastfed (for the sample of firstborns $\hat{\beta}_2 = -0.0161$ with $SE(\hat{\beta}_2) = 0.0241$).

6 Conclusions

To be written

Appendix A Parental investment equation with unobserved parental genes

Let a structural parental investment equation for firstborns be written as:

$$PI_{1f} = \beta_0 + \beta_1 EPGS_{1f} + \beta_2(EPGS_{1f} - EPGS_{2f}) + \gamma EPGS_{pf} + \epsilon_{1f} \tag{A.1}$$

, where f indexes families, and $EPGS_{pf}$ denotes parental EPGS (the average of maternal and paternal EPGS). $\mathbb{E}(\epsilon_{1f}|X_f) = 0$, where $X_f = \{EPGS_{1f}, EPGS_{2f}, EPGS_{pf}\}$.

Since children inherit their genes from their parents, we can write:

$$EPGS_{1f} = EPGS_{pf} + v_{1f} \tag{A.2}$$

$$EPGS_{2f} = EPGS_{pf} + v_{2f} \tag{A.3}$$

Note that v_{1f} and v_{2f} are uncorrelated across siblings because genetic lotteries occur within families (Fletcher and Lehrer, 2011; Domingue *et al.*, 2015) or, in other words, the allocation of genotypes across siblings is as good as random. Hence:

$$Cov(v_{1f}, v_{2f}) = Cov(EPGS_{pf}, v_{1f}) = Cov(EPGS_{pf}, v_{2f}) = 0$$

To assess the size of the potential bias induced by the fact that we do not observe parental genes, we express $EPGS_{pf}$ as a function of $EPGS_{1f}$ and $EPGS_{2f}$. Let us define a linear projection:

$$L(EPGS_p|EPGS_1, EPGS_2) = \delta_1 EPGS_1 + \delta_2 EPGS_2 \tag{A.4}$$

, where:

$$\begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}(EPGS_1^2) & \mathbb{E}(EPGS_1 EPGS_2) \\ \mathbb{E}(EPGS_1 EPGS_2) & \mathbb{E}(EPGS_2^2) \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{E}(EPGS_1 EPGS_p) \\ \mathbb{E}(EPGS_2 EPGS_p) \end{pmatrix}$$

Solving this we obtain:

$$\delta_1 = \frac{\mathbb{E}(EPGS_2^2)\mathbb{E}(EPGS_1 EPGS_p) - \mathbb{E}(EPGS_1 EPGS_2)\mathbb{E}(EPGS_2 EPGS_p)}{\mathbb{E}(EPGS_1^2)\mathbb{E}(EPGS_2^2) - \mathbb{E}(EPGS_1 EPGS_2)^2} \tag{A.5}$$

$$\delta_2 = \frac{\mathbb{E}(EPGS_1^2)\mathbb{E}(EPGS_2 EPGS_p) - \mathbb{E}(EPGS_1 EPGS_2)\mathbb{E}(EPGS_1 EPGS_p)}{\mathbb{E}(EPGS_1^2)\mathbb{E}(EPGS_2^2) - \mathbb{E}(EPGS_1 EPGS_2)^2} \tag{A.6}$$

EPGS are standardized to have mean 0 and standard deviation 1, which implies that:

$$\mathbb{E}(EPGS_p) = \mathbb{E}(EPGS_1) = \mathbb{E}(EPGS_2) = 0 \tag{A.7}$$

$$\mathbb{E}(EPGS_1^2) = \mathbb{E}(EPGS_2^2) = 1 \tag{A.8}$$

From (A.2) and (A.3) we obtain:

$$\mathbb{E}(EPGS_1 EPGS_p) = \mathbb{E}(EPGS_2 EPGS_p) = \mathbb{E}(EPGS_1 EPGS_2) \tag{A.9}$$

Finally, substituting (A.9) into (A.5) and (A.6) we obtain that:

$$\delta_1 = \delta_2 = \delta = \frac{\mathbb{E}(EPGS_1 EPGS_2) - \mathbb{E}(EPGS_1 EPGS_2)^2}{1 - \mathbb{E}(EPGS_1 EPGS_2)^2} \tag{A.10}$$

Let us rewrite equation (A.4) as a function of $(EPGS_{1f} - EPGS_{2f})$:

$$L(EPGS_p|EPGS_1, EPGS_2) = \delta EPGS_1 - \delta(EPGS_1 - EPGS_2 - EPGS_1)$$
$$L(EPGS_p|EPGS_1, EPGS_2) = 2\delta EPGS_1 - \delta(EPGS_1 - EPGS_2)$$

Therefore:

$$EPGS_{pf} = 2\delta EPGS_{1f} - \delta(EPGS_{1f} - EPGS_{2f}) + e_{1f} \tag{A.11}$$

,where $\mathbb{E}(e_{1f}|X) = 0$. Substituting (A.11) into (A.1) we obtain :

$$PI_{1f} = \beta_0 + (\beta_1 + 2\delta\gamma)EPGS_{1f} + (\beta_2 - \delta\gamma)(EPGS_{1f} - EPGS_{2f}) + \gamma e_{1f} + \epsilon_{1f} \tag{A.12}$$

, where $\mathbb{E}(\gamma e_{1f} + \epsilon_{1f}|X) = 0$.

Therefore, estimating equation (A.1) with omitted $EPGS_p$ yields the following estimates:

$$\tilde{\beta}_1 = \beta_1 + 2\delta\gamma$$
$$\tilde{\beta}_2 = \beta_2 - \delta\gamma$$

Let us assume that $\beta_1 \geq 0$ or that the price effect is non negative. That is, we assume that it is not less costly for parents to invest in a lower-endowed child than in better-endowed child. Then:

$$\beta_1 = \tilde{\beta}_1 - 2\delta\gamma \geq 0 \Leftrightarrow \gamma \leq \tfrac{\tilde{\beta}_1}{2\delta}$$

This implies that:

$$\beta_2 \leq \tilde{\beta}_2 + \frac{\tilde{\beta}_1}{2} \tag{A.13}$$

(A.13) gives us the true β_2 in the "worst-case bias scenario", that is, when our estimate of β_2 has the largest possible bias due to the omission of parental genes. This inequality allows us to compute the size of β_2 in the "worst-case bias scenario" for each of the three absolute measures of parental investments used in the paper. The first two columns of Table

Table A.1: Worst-Case Bias Scenario due to the Omission of Parental Genes

	$\tilde{\beta}_1$	$\tilde{\beta}_2$	Upper bound of β_2	$\frac{\beta_2 * 100\%}{\tilde{\beta}_2}$
Parental Investment Index	0.116	-0.138	-0.080	58%
Maternal Investment Index	0.085	-0.112	-0.069	62%
Paternal Investment Index	0.142	-0.156	-0.085	54%

A.1 display $\tilde{\beta}_1$ and $\tilde{\beta}_2$ obtained after reestimating our main specification without normalizing $EPGS_1 - EPGS_2$ so as to be consistent with the calculations presented in this Appendix.[19]. Column 3, in turn, presents the upper bound or the worst-case scenario values of β_2 computed using (A.13). They are negative for the three (non-relative) parental investment indicators, which is consistent with parents being inequality averse. Moreover, as shown in Column 4 of Table A.1, these worst-case scenario true values amount to sizeable shares (between 54% and 62%) of our estimated values displayed in Column 2.

Appendix B Socioeconomic Index Construction

Following Belsky *et al.* (2018) we constructed a family socioeconomic status indicator using information on Add Health participants' parents collected at the Wave I interview. We useed information on parental education, parental occupation, household income, and household receipt of public assistance.

We constructed parental years of schooling using one question addressed to parents (mostly to mothers) at Wave I (*"How far did you go in school?"*), as well as questions addressed to children at Wave I about both their resident mother and father (*"How far in school did she(he) go?"*). The maternal years of schooling variable is based on mothers' answers if they participated in the parental interview, and on their children's answers otherwise. The paternal years of schooling variable was constructed analogously. Paternal education was then computed as the average of paternal and maternal years of schooling.

We used children's answer to a question regarding both their father's and their mother's occupation (*"What kind of work does she (he) do?"*) to construct an occupational prestige indicator. In particular, we assigned occupational prestige scores based on the National Opinion Research Center (NORC) occupational classification.[20] We then computed a parental occupational prestige score as the average of mothers' and fathers' prestige scores.

Family income is based on the following question addressed to parents at Wave I: *"About how much total income, before taxes did your family receive in 1994? Include your own income, the income of everyone else in your household, and income from welfare benefits, dividends, and all other sources."*

[19] Hence, these figures are not equal to the estimates reported in Panels A, B and C of Table 6. Note also that we do not do this exercise for our relative parental investment or favouritism indicator because it is unaffected by the omission of family-specific factors like parental genes.

[20] http://ibgwww.colorado.edu/~agross/NNSD/prestige%20scores.html

As for household receipt of public assistance, we relied on the following question asked to children at Wave I regarding both their mother and their father: *"Does she (he) receive public assistance, such as welfare?"*.

Finally, we conducted principal components analysis of parental education, parental occupational attainment, family income, and household receipt of public assistance to produce a factor score. The first principal component explained 46.2% of the variance. We used loadings on this component to compute a socioeconomic status index, and then we standardized it to have mean 0 and standard deviation 1.

References

Abufhele, A., Behrman, J. and Bravo, D. (2017): "Parental preferences and allocations of investments in children's learning and health within families," *Social Science & Medicine*, 194, 76–86.

Adhvaryu, A. and Nyshadham, A. (2016): "Endowments at birth and parents' investments in children," *The Economic Journal*, 126, 781–820.

Agostinelli, F. and Wiswall, M. (2016): "Estimating the Technology of Children's Skill Formation," Tech. rep., National Bureau of Economic Research.

Aizer, A. and Cunha, F. (2012): "The production of human capital: Endowments, investments and fertility," Tech. rep., National Bureau of Economic Research.

Almond, D. and Currie, J. (2011): "Killing me softly: The fetal origins hypothesis," *Journal of economic perspectives*, 25, 153–72.

Almond, D. and Mazumder, B. (2013): "Fetal origins and parental responses," *Annu. Rev. Econ.*, 5, 37–56.

Ayalew, T. (2005): "Parental preference, heterogeneity, and human capital inequality," *Economic Development and Cultural Change*, 53, 381–407.

Barth, D., Papageorge, N. W. and Thom, K. (2018): "Genetic Endowments and Wealth Inequality," Tech. rep., National Bureau of Economic Research.

Becker, G. S. and Tomes, N. (1976): "Child endowments and the quantity and quality of children," *Journal of political Economy*, 84, S143–S162.

Behrman, J. R., Pollak, R. A. and Taubman, P. (1982): "Parental preferences and provision for progeny," *Journal of Political Economy*, 90, 52–73.

Belsky, D. W., Domingue, B. W., Wedow, R. *et al.* (2018): "Genetic analysis of social-class mobility in five longitudinal studies," *Proceedings of the National Academy of Sciences*, 115, E7275–E7284.

Belsky, D. W. and Israel, S. (2014): "Integrating genetics and social science: Genetic risk scores," *Biodemography and social biology*, 60, 137–155.

Bharadwaj, P., Eberhard, J. P. and Neilson, C. A. (2018): "Health at birth, parental investments, and academic outcomes," *Journal of Labor Economics*, 36, 349–394.

Cabrera-Hernández, F. (2012): "The accident of birth: effect of birthweight on educational attainment and parent's compensations among siblings," .

Currie, J. (2011): "Inequality at birth: Some causes and consequences," *American Economic Review*, 101, 1–22.

Datar, A., Kilburn, M. R. and Loughran, D. S. (2010): "Endowments and parental investments in infancy and early childhood," *Demography*, 47, 145–162.

Del Bono, E., Ermisch, J. and Francesconi, M. (2012): "Intrafamily resource allocations: a dynamic structural model of birth weight," *Journal of Labor Economics*, 30, 657–706.

Domingue, B. W., Belsky, D. W., Conley, D. *et al.* (2015): "Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health," *AERA open*, 1, 2332858415599972.

Ejrnæs, M. and Pörtner, C. C. (2004): "Birth order and the intrahousehold allocation of time and education," *Review of Economics and Statistics*, 86, 1008–1019.

Fletcher, J. M. and Lehrer, S. F. (2011): "Genetic lotteries within families," *Journal of health economics*, 30, 647–659.

Frijters, P., Johnston, D. W., Shah, M. *et al.* (2013): "Intrahousehold resource allocation: do parents reduce or reinforce child ability gaps?" *Demography*, 50, 2187–2208.

Hsin, A. (2012): "Is biology destiny? Birth weight and differential parental treatment," *Demography*, 49, 1385–1405.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75, 83–119.

McQueen, M. B., Boardman, J. D., Domingue, B. W. *et al.* (2015): "The national longitudinal study of Adolescent to Adult Health (Add Health) sibling pairs genome-wide data," *Behavior genetics*, 45, 12–23.

Papageorge, N. W. and Thom, K. (2018): "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study," Working Paper 25114, National Bureau of Economic Research.

Pitt, M., Rosenzweig, M. and Hassan, M. N. (1990): "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries," *American Economic Review*, 80, 1139–56.

Restrepo, B. J. (2016): "Parental investment responses to a low birth weight outcome: who compensates and who reinforces?" *Journal of Population Economics*, 29, 969–989.

Rietveld, C. A., Conley, D., Eriksson, N. *et al.* (2014): "Replicability and robustness of genome-wide-association studies for behavioral traits," *Psychological science*, 25, 1975–1986.

Rietveld, C. A., Medland, S. E., Derringer, J. *et al.* (2013): "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment," *Science*, 340, 1467–1471.

Rosales-Rueda, M. F. (2014): "Family investment responses to childhood health conditions: Intrafamily allocation of resources," *Journal of Health Economics*, 37, 41–57.

Terskaya, A. (2019): "Parental Human Capital Investment Responses to Children's Disability," Working paper, Fundamentos del Análisis Económico (FAE), Universidad de Alicante.

Yi, J., Heckman, J. J., Zhang, J. *et al.* (2015): "Early Health Shocks, Intra-household Resource Allocation and Child Outcomes," *The Economic Journal*, 125, F347–F371.

Zweimueller, M. and Halla, M. (2014): "Parental Response to Early Human Capital Shocks," Working paper, Universitaet Duisburg-Essen, Competent in Competition and Health.