I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12375

Sibling Differences in Educational Polygenic Scores: How Do Parents React?

Anna Sanz-de-Galdeano
Anastasia Terskaya

MAY 2019

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12375

Sibling Differences in Educational Polygenic Scores: How Do Parents React?

Anna Sanz-de-Galdeano
University of Alicante and IZA

Anastasia Terskaya
University of Alicante

MAY 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

We take advantage of recent advances in behavioral genetics to revisit a classic question in economics: how do parents respond to children's endowments and to differences in endowments among siblings? 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 effect 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, we use genetic variants that predict educational attainment as a measure of children's educational endowments. Individuals' genetic makeup is fixed at conception, so these indicators cannot be affected by parental investment decisions. We find evidence that parents of non-twin 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. In contrast, parents of twins do not significantly react to endowment differences among their children, likely because it is difficult to provide differential parental investments across children of exactly the same age.

JEL Classification: D13, D64, J13, J24

Keywords: intra-household allocation of resources, educational polygenic scores, parental investments, endogenous fertility, add health

Corresponding author:
Anna Sanz-de-Galdeano
Universidad de Alicante
Carretera de San Vicente s/n
03080 San Vicente
Alicante
Spain
E-mail: anna.sanzdegaldeano@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, Daniel Hamermesh and Daniel Rees, as well as seminar participants at University of Alicante, University of Melbourne, University of Padova, Universidad Alberto Hurtado, Universidad Pablo de Olavide, and Universidad Autónoma de Madrid for useful discussions and comments. We would also like to thank the participants to the 3rd IZA Workshop on Gender and Family Economics.

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 children's endowments is vast, and it has continuously grown since the seminal contributions of Becker and Tomes (1976) and Behrman *et al*. (1982). 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 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 children with lower endowments) or a reinforcing strategy (devoting more resources to their better-endowed siblings).[1]

The subsequent empirical literature has so far reached mixed conclusions on whether parents compensate or reinforce differences in their children's endowments.[2] Some studies have found evidence of parental compensatory behavior (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 to be expected because different authors have used different measures of children's endowments and/or parental investments in different contexts, and, perhaps more importantly, this literature presents several identification challenges that have been dealt with in various ways.

This article combines the traditional literature on intra-household resource allocation with recent advances in behavioral genetics to study how parents respond to children's educational genetic endowments and to differences in educational genetic endowments among siblings. We extend the previous empirical literature on intra-household resource allocation in three ways. First, we take into account that parental investment decisions depend both on parental preferences regarding inequality in the distribution of their children's quality (Behrman *et al*., 1982), as well as on how costly it is for parents to add to their children's quality by investing

[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.

[2] See Almond and Mazumder (2013) for a review.

in their human capital—or the price effect (Becker and Tomes, 1976). Importantly, evidence based on family fixed effects models, which compare parental investments across children within the family, is not informative on whether parents are inequality averse or not, as even inequality averse parents may follow a reinforcing strategy if the cost of investing in lower-endowed children is sufficiently higher than the cost of investing in higher-endowed children (Terskaya, 2019). To address this issue, we estimate 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 endowment, which serves as a proxy for the costs to adding to his/her quality faced by the parents. That is, we pose the following questions: do parents react to children's endowment levels? Do they 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 endowment and hence the costs their parents face if they invest in them? Distinguishing parental preferences for equality versus efficiency from the price effect is important for policy design and for assessing the effectiveness of compensatory interventions designed to help disadvantaged children because parents will reinforce or attenuate (depending on their preferences) the impact of such interventions by reallocating resources within the family.

Second, we use educational polygenic scores—a summary measure of genome-wide genetic variants that predict educational attainment—as an indicator of children's educational genetic endowments. Not only these genetic indexes robustly predict educational attainment (Domingue *et al.*, 2015; Papageorge and Thom, 2019), but they also allow 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,[3] while endowment indicators measured at birth (*e.g.*, birth weight) may reflect pre-natal investment decisions (Del Bono *et al.*, 2012; Almond and Currie, 2011; Currie, 2011, among others). In contrast, individuals' genetic endowments are fixed at conception and hence cannot be the consequence of parental investment choices. In addition, we deal with the potential endogeneity of fertility decisions—which may affect studies based on non-twin siblings because parents' decisions to have more children may depend on the endowments of previous children—by focusing on twins and on firstborn children.[4] We analyse twins separately because previous studies suggest that it is harder for parents to invest differentially within twins than within singleton siblings (Bharadwaj *et al.*, 2018).

Third, we focus on parental responses to differences in children's educational genetic endowments rather than health endowments or shocks, while, with some exceptions (Ayalew, 2005; Frijters *et al.*, 2013; Garcia Hombrados, 2017), most previous studies have focused on the latter. This may be due to the fact that at-birth measures of endowments (other than birth

[3] 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 in order to 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.

[4] See Ejrnæs and Pörtner (2004) among others.

weight), which are less likely to suffer from reverse causality than indicators measured later on in life, are not readily available.[5] Be that as it may, we know much less about parental responses to differences in their children's educational endowments than about how parents react when they face differences in their children's health endowments. However, parental responses could differ across these dimensions, so our study adds to the meager previous literature that focuses on educational endowments.

We rely on data from the sibling sample of the National Longitudinal Survey of Adolescent Health, a nationally representative of youth in the 7th through 12th grades in the United States in 1995. We find evidence that American parents of non-twin siblings display inequality aversion because, given a child' absolute level of educational genetic endowment, they invest less in him/her if his/her sibling is better-endowed, while also holding constant other sibling-level and family specific characteristics. Parents of twins instead do not significantly respond to endowment differences among their children. This consistent with the idea that it is more difficult to invest differently across siblings who are closer together in age than among siblings who are farther apart, so the role played by public goods in parental investments is expected to be larger for the former than for the latter (Bharadwaj *et al.*, 2018).[6] We also find suggestive evidence that the price effect is positive for parents of singleton siblings, as they invest more in genetically better-endowed children—holding endowment differences among siblings constant, as well as other sibling and family level characteristics.

Finally, our study also speaks to a broad and emerging literature that aims at integrating genetics and the social sciences (Beauchamp *et al.*, 2011; Lehrer and Ding, 2017). For instance, recent contributions have studied the association between educational polygenic scores and human capital accumulation (Domingue *et al.*, 2015; Papageorge and Thom, 2019), labor market outcomes (Papageorge and Thom, 2019), and wealth at retirement (Barth *et al.*, 2019). 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.

The remainder of the paper is organized as follows. In Section 2, we discuss our empirical strategy, focusing on how we exploit the availability of genetic data to address the empirical challenges involved in disentangling the price effect from the impact of parental preferences for equality versus efficiency. In Section 3 we describe the Add Health dataset used, we describe our measures of genetic predisposition to educational attainment, and we show that

[5]Most of the studies relying on at-birth indicators focus on birth weight (Datar *et al.*, 2010; Hsin, 2012; Cabrera-Hernández and Orraca-Romano, 2016; 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.

[6]We refer to non-monetary investments that can simultaneously involve more than one child and hence incorporate a spillovers dimension.

they correlate with several education-related indicators in our sample. Section 4 discusses our estimation results, Section 5 presents falsification tests, and 6 concludes.

2 Empirical Strategy

Our primary goal is to estimate how parents' investment decisions are affected by children's genetic predisposition for education. 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 versus efficiency *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) educational genetic endowments while holding children's own (absolute) 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:

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

, 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 in family f, 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. Both our indicators of parental investments and our regressors are described in detail in Section 3. Note that $(EPGS_{if} - EPGS_{jf})$ is our measure 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 non-twin siblings and twins because identifying the effect of interest presents more challenges in the case of non-twin 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 it would indicate that parents 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, $\beta_1 > 0$ implies that, for any given level of inequality in siblings' endowments $(EPGS_{if} - EPGS_{jf})$, parents will 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 cannot not be attributed to parental preferences for efficiency over equality.[7]

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 birth weight) 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.

[7]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.

2.3 Unobserved parental genes

Despite the fact that genetic lotteries occur within families (Fletcher and Lehrer, 2011; Domingue *et al.*, 2015; Davey Smith and Hemani, 2014), 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 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 would be about 60% of our estimated $\hat{\beta}_2$.

In Section 4 we show that our estimated coefficients of interest barely change when controlling parental characteristics (*e.g.*, parental socioeconomic status, education, *etc.*). In line with this evidence, we also follow Altonji *et al.* (2005) in calculating how strong selection on unobservables would have to be in order to fully account for our estimated effects of interest, and we find that selection on unobservables (such as parental genes) would have to be at least 30 and 18 times stronger than selection on observables to attribute our entire OLS estimates of β_1 and β_2 to selection effects, respectively.

Additionally, we 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 (described in detail in Section 3.3) 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. In Section 4 we show that 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 non-twin 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 are

Table 1: Parental Preferences for Children's Ability 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

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 firstborn 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 non-twin siblings by focusing on firstborn 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 firstborn 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 firstborn children with at least one younger sibling. Note also that analysing the investment decisions of parents of twins and non-twin siblings separately is advisable, as previous studies suggest that parents of twins are less likely to be responsive to their endowment differences because it is more difficult for them to implement favouritism among their twin children than for parents of siblings of different ages (Bharadwaj *et al.*, 2018).

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. 7th to 12th 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 first interviewed in 1994/95 (Wave I, ages 12–20 years), which constitutes the so called in-home sample. Subsequent interviews were conducted in 1996 (Wave II), in 2001/02 (Wave III), and in 2008 (Wave IV, ages 24-32 years). Information on all our variables of interest was collected at Wave I with the exception of genetic information, as genotyping was performed at Wave IV.

For our analysis we use the Add Heath Sibling Pairs sample with available educational polygenic scores (described in Section 3.2), which includes 1,886 individuals from 1,113 families. In 380 families only one sibling was genotyped. Since our aim is to study how parental investment decisions are affected by the existence of differences across children's endowments, we drop these 380 observations, which leaves us with 732 families (595 with non-twin 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 firstborns (No. Obs.=595).

The in-home survey of Add Health collects information on respondents' relationship with their parents, which allows us construct indicators of parental investments and parental favouritism. Another crucial advantage of using Add Health data in our analysis is that this dataset also contains genetic information for the Add Heath Sibling Pairs sample. Individuals' educational polygenic scores (our measure of individuals' genetic predisposition for education) are described in Section 3.2.

Importantly, the Add Health in-home and parent questionnaires also provide detailed information on individual and family background characteristics, such as age, sex, and race of respondents and their siblings, family structure, parental education, and socioeconomic status. To measure parental socioeconomic 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).[8] We describe the construction of this index in Appendix B, while Table C.2 in Appendix C displays summary statistics for our independent variables in the samples of firstborn singletons and twins.

Another crucial advantage of Add Health for our purposes is that its in-home survey collects information on respondents' relationship with their parents, which allows us to construct indicators of parental investments and parental favouritism (described in Section 3.3).

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),[9] 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.[10]

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

[9] 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.

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

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, while 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.[11] 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.[12]

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 related to cognitive ability. Hence, one may interpret EPGS as proxies for genetic cognitive endowments or genetic cognitive ability. Note, however, that, as Papageorge and Thom (2019) point out, EPGS may well incorporate genetic variants that do not operate exclusively through cognitive channels. Hence, we refer to EPGS as educational genetic endowments or genetic predisposition to educational attainment throughout the manuscript.

Figure C.1 in Appendix C plots the (kernel-smoothed) densities of Add Health responents' EPGS and those of their nearest younger siblings or twins for firstborns and twins, respectively.[13] The distribution of Add Health respondents' EPGS is approximately normal, and it does not significantly vary by birth order.

Table 2 shows that, as expected, individuals' EPGS have a strong association with educational attainment (years of schooling) in our sample of siblings. EPGS explain almost 5% of the observed variance in educational attainment, while a 1 standard deviation increase in EPGS is associated with almost 0.5 additional years of schooling if no controls are included (Column 1 of Table 2), and with about 0.3 additional years of schooling if one controls for both individual and family characteristics (Column 2 of Table 2), such as family SES among others. This result is in line with the evidence from Domingue *et al.* (2015), Barth *et al.* (2019)

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

[12] See the Add Health documentation (`https://www.cpc.unc.edu/projects/addhealth/documentation/restricteduse/datasets/GRS_EDU.pdf`) and Domingue *et al.* (2015) for details on educational polygenic score constructed by the Add Health for the sibling pairs sample. See also Papageorge and Thom (2019) and Barth *et al.* (2019) for a comprehensive discussion of GWAS, the computation of educational polygenic scores and their interpretation.

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

and Papageorge and Thom (2019).

The same pattern holds when analysing other indicators related to educational achievement such as scores on the Peabody Picture Vocabulary Test (PPVT),[14] grades, and the probability of dropping out of high school (Columns 3-8 of Table 2).

Since it is known that genetic variation across siblings resembles a lottery, it is worth checking whether differences in endowments among siblings are indeed uncorrelated with our individual and family control variables. Table 3 presents covariate balance 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 firstborns (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 is as good as random, none of these associations are significant.

3.3 Parental Investments

We use 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.[15]

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:[16] *"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*

[14] At the beginning of Wave I interview, Add Health respondents were given a computerized and abridged version of the PPVT, an age-specific test used to assess verbal ability and receptive vocabulary.

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

[16] 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: Educational Polygenic Scores, Years of Schooling and Other Education-Related Indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Years of education		PPVT		Overall GPA		High-School Drop-Out	
EPGS. Normalized	0.498***	0.296***	4.650***	2.075***	0.278***	0.160***	-0.0414***	-0.0261**
	(0.0717)	(0.0702)	(0.435)	(0.436)	(0.0325)	(0.0349)	(0.00988)	(0.0102)
Controls included		yes		yes		yes		yes
Observations	869	869	833	833	645	645	869	869
R-squared	0.047	0.274	0.105	0.295	0.094	0.295	0.018	0.086

Note: OLS coefficient estimates and their associated robust standard errors in parentheses. The following controls are included: age, age squared, the female dummy, race indicators, the rural area dummy, total family income, the indicator that at least one parent is a college graduate, the indicator that both parents live in the household, and the socioeconomic status index. EPGS is the educational polygenic score provided by Add Health for the sibling sample. It is normalized to have mean 0 and standard deviation 1. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

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.

Table C.3 in Appendix C describes the main outcomes measuring parental investments as well as their components.

4 Main Results

Our main results are displayed in Table 4. 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 firstborn 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 non-twin siblings and then move on to the evidence for twins.

Regardless of the parental investment indicator used, our results indicate that parents of non-twin 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 firstborn 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

Table 3: Balancing Tests. Correlations between Educational Polygenic Score Differences Between Siblings and Individual and Household Characteristics

	firstborns		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 at Wave I of Add Health. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

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.[17]

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 in-

[17]We estimate that one standard deviation increase in family socioeconomic status is associated with about a 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 4: The Effect of Educational Polygenic Scores and Sibling Differences in Educational Polygenic Scores on Parental Investments

	firstborns		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, and race indicators for respondents and their siblings, 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 Add Health for the sibling sample (normalized to have mean 0 and standard deviation 1). EPGS - Sibling's EPGS is the difference in EPGS between siblings (that is, between firstborns and their next younger sibling in the sample of firstborns, and between twins in the twins sample), also normalized to have mean 0 and standard deviation 1. *** $p < 0.01$, $** p < 0.05$, $*p < 0.1$

duced 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. Moreover, 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 β_2 virtually unaltered. A complementary way to assess the extent to which unobservables may drive our results relies on postulating that selection on observables is informative about selection on unobservables as suggested by Altonji *et al.* (2005). We find that the impact of unobserved factors would have to be at least 18 times stronger, as compared to observed factors, in order to explain away the relationship between sibling differences in educational genetic endowments ($EPGS_1$-$EPGS_2$) and parental investments. This makes it unlikely that unobservable factors can account for our results.

Consistent with the evidence presented so far, 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 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.[18]

As for the price effect, the coefficient of firstborn 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). Absent further evidence, one should be cautions in giving $\hat{\beta}_1$ a causal interpretation because it is not possible to 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 indicates 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

[18]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).

obtained when using the other three (absolute) 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 4 reveals. This similarity is in line with the large magnitudes of the Altonji *et al.* (2005) ratios we obtain for all our parental investment regressions. For instance, the ratio of the covariance between unobservables and EPGS and the covariance between observables and EPGS must be at least 30 to explain away the price effect for the parental investment index.

Our finding that the price effect is positive implies that, if it is large enough, even inequality averse parents may choose to follow a reinforcing or a neutral strategy. Actually, if we use our sample of firstborns to estimate a family fixed- effects model we find that their parents follow a neutral strategy despite being inequality averse, which suggests that the price effect and the parental inequality aversion mechanisms offset each other in the US.[19] This is a relevant result for the literature on intra-household resource allocation, as it may explain why previous empirical studies relying on family fixed-effects models have often found that parents follow a reinforcing strategy even in developed countries, like the US, with well-established credit markets and old-age pension systems (Datar *et al.*, 2010; Aizer and Cunha, 2012; Hsin, 2012; Frijters *et al.*, 2013; Rosales-Rueda, 2014).

In sum, our evidence for non-twin siblings is clearly supportive of parents displaying inequality 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 4, Columns 3 and 4, second row). One potential explanation for our contrasting results for parents of twins and of non-twin siblings might be that it can be difficult for the former to invest differently across their children because they are exactly the same age. 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. This dimension is likely to be greater for children who are very close in age, which provides a rationalization for the differences we observe between parents of twins and non-twin 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. This implies that, even if parents of twins were inequality averse, they may be unable to invest differentially across their children even if they wished to. Additionally, they may find it hard to invest more in their better-endowed twin even if it is less costly for them to do so than investing in their lower-endowed twin. This also likely explains why there is no significant price effect among the parents of twins.

[19] In other words, if we estimate the following equation:

$$PI_{if} = \alpha_0 + \alpha_1 * EPGS_{if} + X'_{if}\delta + \rho_f + u_{if}$$

, where ρ_f is a family fixed effect, we find that $\hat{\alpha}_1$ does not significantly differ from zero. These results are available upon request form the authors.

Finally, we investigate whether the price effect and the degree of inequality aversion and the price effect we have previously uncovered for parents of non-twin siblings varies along the parental investment distribution. Table 5 and Figure 1 summarize the results of estimating equation (2) using unconditional quantile regression methods (Firpo *et al.*, 2009). The results indicate that parents of non-twin siblings start displaying inequality aversion shortly before the median of the parental investment index distribution. Along the same lines, the price effect grows along the parental investment index distribution and it only turns statistically significant around the median. That is, “low investors” do not significantly react to endowment differences across their children, while “high investors” do, and they do so more strongly the more they invest. This suggests that parental behavioral responses are only detectable once parents exceed an investment level that allows them to be aware of children’s endowments and endowment differences among children.[20]

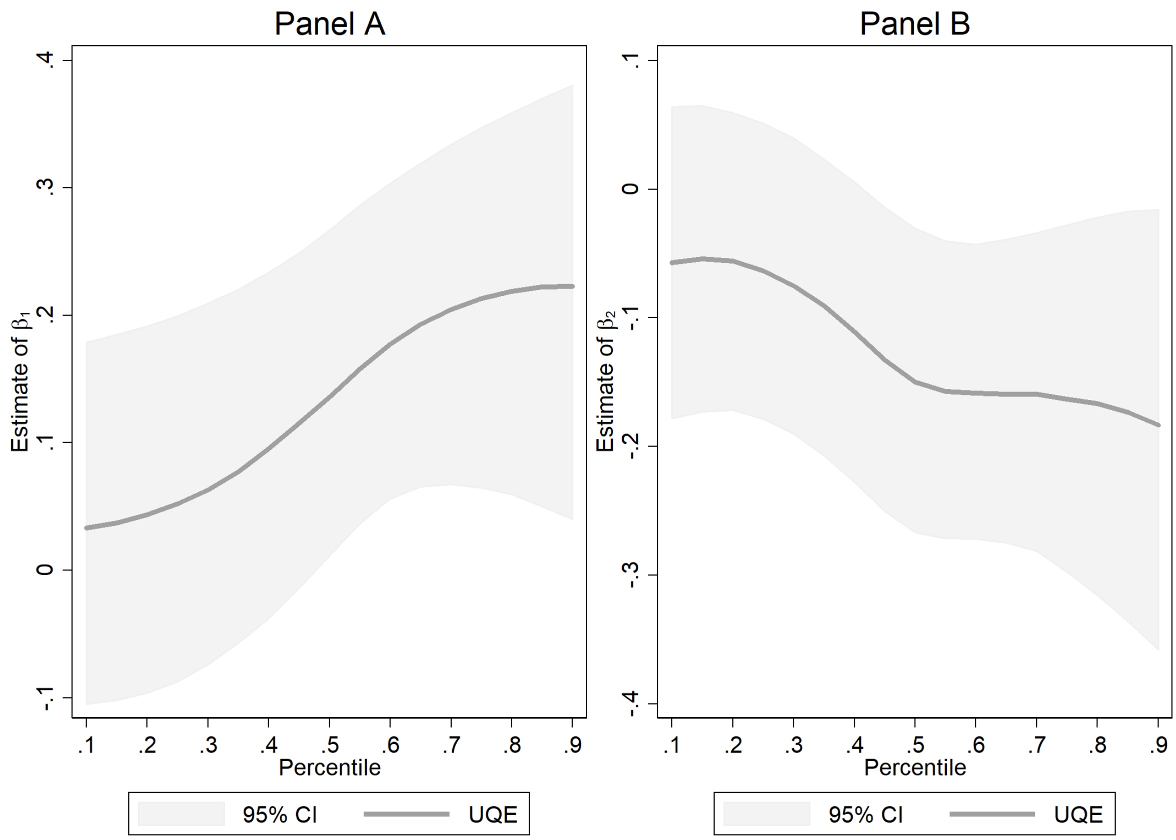
[20] Table 5 and Figure 1 display results for our parental investment index and for parents of non-twin siblings. The pattern of results for the other three investment indicators we have used is the same. We also find that parents of twins display neutral preferences not only at the mean, as shown in Table 5, but all along the parental investment distribution regardless of the investment indicator used.

Table 5: Parental Preferences Regarding Equality vs. Efficiency and the Price Effect Along the Parental Investment Distribution. Unconditional Quantile Regression Estimates. firstborns Sample.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	15th Centile	30th Centile	40th Centile	50th Centile	60th Centile	70th Centile	80th Centile	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. Dependent variable: parental investment index (normalized to have mean 0 and standard deviation 1). The following controls are included: age, age squared, a female dummy, and race indicators both for each individual and his/her sibling, 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 Add Health for the sibling sample (normalized to have mean 0 and standard deviation 1). EPGS - Sibling's EPGS is the difference in EPGS between siblings (that is, between firstborns and their next younger sibling in the sample of firstborns), also normalized to have mean 0 and standard deviation 1. No. observations: 583. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Figure 1: Parental Preferences Regarding Equality vs. Efficiency and the Price Effect Along the Parental Investment Distribution.



Note: The figure shows estimated unconditional quantile effects of $EPGS_1$ (Panel A) and $EPGS_1 - EPGS_2$ (Panel B) and their associated confidence intervals. Dependent variable: parental investment index (normalized to have mean 0 and standard deviation 1). See the note to Table 3 for the full list of controls included in the regression. $EPGS_1$ and $EPGS_1 - EPGS_2$ are also normalized. Sample of firstborns. No. observations: 583.

5 Falsification Tests

5.1 Placebo Tests

In order to check that our results for firstborns are not driven by chance we run placebo tests. To obtain placebo versions of $EPGS_1$ and $EPGS_1 - EPGS_2$ we replace their actual values with those from randomly individuals from our sample of firstborns (denoted by $EPGS_1^P$ and $EPGS_1^P - EPGS_2^P$).[21] We then estimate equation (2) using $EPGS_1^P - EPGS_2^P$ and $EPGS_1^P$ and including covariates. We repeat this procedure 500 times in order to obtain distributions of the estimated coefficients of $EPGS_1^P$ and $EPGS_1^P - EPGS_2^P$. We find that, in line with our

[21] We conduct this test only for firstborns since we find that parents of twins do not significantly react to endowment differences among their children.

results regarding inequality aversion and the price effect being genuine, the coefficients of $EPGS_1^P - EPGS_2^P$ and $EPGS_1^P$ are significant at the 5% level in less than 5% of our placebo regressions. Figure C.2 in Appendix C summarizes the results of these estimations by displaying the placebo t-value distributions of the tests $\beta_1 = 0$ and $\beta_2 = 0$.

5.2 "Too Early" Parental Responses

Parents are unlikely to be aware of children's endowment levels very soon after birth. Additionally, parents cannot possibly 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 EPGS and sibling differences in EPGS impact "too early" or "too soon after birth" parental investment indicators, such as breastfeeding, that are unlikely to be affected by the behavioral responses of parents. We use a retrospective question from Add Health parental questionnaire that asked mothers how long each of 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, the price effect is not significant, and we do not find any significant effect of sibling differences in EPGS on the probability of having been breastfed (for the sample of firstborns) ($\hat{\beta}_1 = -0.020$ with $SE(\hat{\beta}_1) = 0.027$ $\hat{\beta}_2 = -0.0161$ with $SE(\hat{\beta}_2) = 0.0241$).

6 Conclusions

We take advantage of recent advances in behavioral genetics to revisit a longstanding question in economics, namely: how do parental investment decisions respond to children's endowments and to endowment differences among children? We provide new evidence that American parents of non-twin children display inequality aversion. In particular, we show that they invest more (less) in equally endowed children if their genetic predisposition for educational attainment is higher (lower) than that of their siblings. Hence, parental investments may be able to reduce the effect of inequalities in genetic endowments. We also find evidence that the price effect is positive, as parents invest more in genetically better-endowed children *ceteris paribus*. Interestingly, in contrast with our findings for parents of non-twin children, we also find that American parents of twins display neutral preferences for efficiency versus equality and that the price effect is not significant for them.

Our findings are important for evaluating the role of the family in shaping inequality as well as the effectiveness of compensatory policies. In our application, American parents of singleton siblings follow neither a compensating nor a reinforcing strategy (that is what we actually find when we estimate family fixed effects models) because the (positive) price effect and the parental inequality aversion mechanisms offset each other. However, it is still the case that the majority of American parents display inequality aversion. Hence, interventions that help lower-endowed children and increase their endowments will in turn induce their inequality averse parents to invest less in their "compensated" children and act less as equalizing agents, attenuating the impact of compensatory programs as a consequence.

Suppose instead that parents followed a reinforcing strategy. If the price effect is positive because it is less costly for parents to invest in more able children than in their less able siblings, a reinforcing strategy may stem from inequality averse parents whose degree of inequality aversion could not offset the price effect, from neutral parents, or from parents who favour efficiency over equality. Compensatory policies will be most effective in the latter case, as interventions that help lower-endowed children will in turn induce parents who favour efficiency to invest more in them, while the opposite would happen if parents were inequality averse, and no parental behavioral response would take place if parents displayed neutral preferences.

Taken as a whole, our evidence suggests that further research is needed to look into the black box of intra-household dynamics and the nature of parental investments for different types of families. Twin fixed effects models are widely used by social scientists as a means to control for unobserved factors shared by twins (including pre-natal investments), and twin births are often used as an exogenous source of variation in family size under the premise that twin births are quasi-random and have no direct impact (except through fertility) on the outcome under analysis. However, it is well known that twins are a particular fraction of the population and results that rely upon twin differences are unlikely to be generalizable. For instance, Bhalotra and Clarke (2019) show that indicators of maternal health, health-related behaviors, and the pre-natal health environment are systematically positively associated with the probability of a twin birth in both rich and poor countries and also among women who do not use IVF. As for the distinctive features of post-natal parental investments in twins versus singleton siblings, Bharadwaj *et al.* (2018) conjecture that when siblings are close in age the degree of spillovers in parental investments is greater and it is therefore hard for parents to differentially invest across twins. Our findings that parents of singleton siblings display inequality aversion while parents of twins do not significantly respond to endowment differences among children are in line with this notion, and they also suggest that twins estimates likely overstate the effect of initial endowments on later-life outcomes in the general population.

Finally, our results highlight the idea that early life conditions not only affect later-life outcomes directly, but also indirectly through intra-household allocation effects. This idea is not new (see for instance Cunha *et al.* (2010) and Yi *et al.* (2015) among others), but our paper is the first to provide direct evidence that educational genetic endowments shape parental investment decisions. As the costs of genotyping technologies continue to fall, the number of social surveys incorporating genetic markers is growing, and so is the number of studies incorporating genomic data in economic analyses. Our evidence suggests that caution must be taken when interpreting reduced form estimates of genetic endowment effects on later-life outcomes. Additionally, using educational polygenic scores as instrumental variables may not be uncontroversial even if genetic variants are fixed at conception. Not only genes that affect some phenotypes (such as obesity or depression) might directly affect other outcomes of interest as well (*i.e.* educational attainment), as Cawley *et al.* (2011) illustrates, but our results also indicate that genetic endowments may also affect later-life outcomes through parental behavioral responses.

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," Working Paper 22442, National Bureau of Economic Research.

Aizer, A. and Cunha, F. (2012): "The Production of Human Capital: Endowments, Investments and Fertility," Working Paper 18429, 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.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113, 151–184.

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. (2019): "Genetic Endowments and Wealth Inequality," *Journal of Political Economy (forthcoming)*.

Beauchamp, J. P., Cesarini, D., Johannesson, M. *et al.* (2011): "Molecular Genetics and Economics," *Journal of Economic Perspectives*, 25, 57–82.

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.

Bhalotra, S. R. and Clarke, D. (2019): "Twin birth and maternal condition," *The Review of Economics and Statistics (forthcoming)*.

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. and Orraca-Romano, P. (2016): "The accident of birth: effect of birthweight on educational attainment and parent's compensations among siblings," Working paper, Centro de Investigación y Docencia Económicas. Available at https://www.academia.edu/34744698/The_accident_of_birth_effects_of_birthweight_on_educational_attainment_and_parents_compensations_among_siblings.

Cawley, J., Han, E. and Norton, E. C. (2011): "The validity of genes related to neurotransmitters as instrumental variables," *Health Economics*, 20, 884–888.

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

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.

Davey Smith, G. and Hemani, G. (2014): "Mendelian randomization: genetic anchors for causal inference in epidemiological studies," *Human Molecular Genetics*, 23, R89–R98.

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.

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.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.

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.

Garcia Hombrados, J. (2017): "Cognitive Skills and Intra-Household Allocation of Schooling," Working paper, Department of Economics, University of Sussex Business School. Available at https://www.sussex.ac.uk/webteam/gateway/file.php?name=wps-18-2017.pdf&site=24.

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.

Lehrer, S. F. and Ding, W. (2017): "What is the role for molecular genetic data in public policy?" , *IZA World of Labor*, 2017: 395.

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. (2019): "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study," *Journal of the European Economic Association (forthcoming)*.

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. Available at https://drive.google.com/file/d/1KWFzgQY88-ZDZMYsITnHcxyOy5OSKeQM/view.

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," CINCH Working Paper Series 1401, Universitaet Duisburg-Essen, Competent in Competition and Health.

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	$\frac{\beta_2}{\tilde{\beta}_2} * 100$
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.[22]. 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 sizable 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 used 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.[23] 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."*

[22]Hence, these figures are not equal to the estimates reported in Panels A, B and C of Table 4. 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.

[23]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.

Appendix C Descriptive Statistics

Table C.2: Summary Statistics of Regressors

	firstborns		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.	-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	

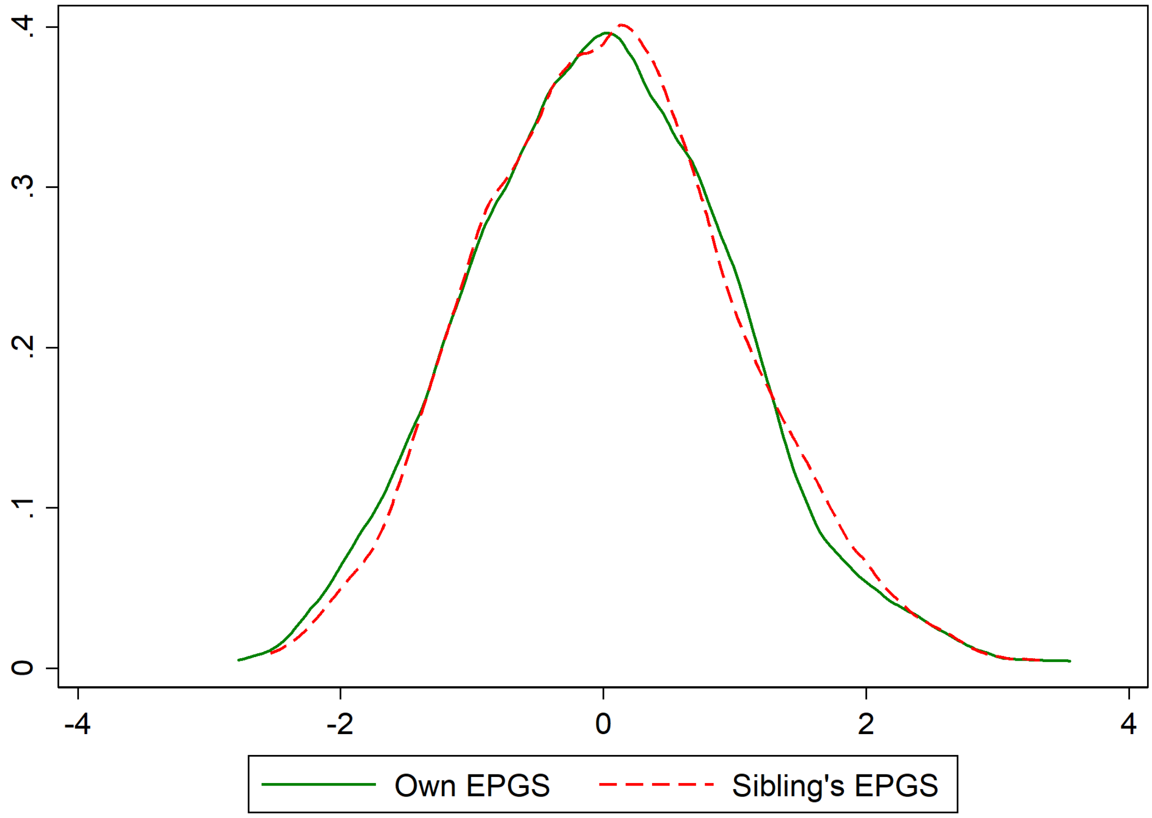
Note: EPGS is the educational polygenic score provided by Add Health for the sibling sample. Normalized variables have mean 0 and standard deviation 1.

Table C.3: Summary of Outcomes

	firstborns			Twins		
	N	Mean	SD	N	Mean	SD
favouritism (Normalized)	423	0.000	1.000	217	-0.000	1.000
Parental Investment Index (Normalized)	583	0.000	0.496	272	0.057	0.519
Maternal Investment Index (Normalized)	568	-0.000	0.595	265	0.082	0.602
Paternal Investment Index (Normalized)	411	-0.000	0.680	195	0.048	0.725
Parental Investment Indexes Components						
Maternal Investment Index Components						
Attended cultural/sports event with mother	568	0.236	0.425	265	0.298	0.458
Talked about a personal problem with mother	568	0.423	0.494	265	0.385	0.487
Talked about school with mother	568	0.653	0.476	265	0.691	0.463
Worked on a project with mother	568	0.120	0.325	265	0.151	0.359
Talked about other school things with mother	568	0.546	0.498	265	0.623	0.486
Paternal Investment Index Components						
Attended cultural/sports event with father	411	0.212	0.409	195	0.277	0.449
Talked about a personal problem with father	411	0.195	0.396	195	0.185	0.389
Talked about school with father	411	0.540	0.499	195	0.533	0.500
Worked on a project with father	411	0.097	0.297	195	0.133	0.341
Talked about other school things with father	411	0.499	0.501	195	0.477	0.501

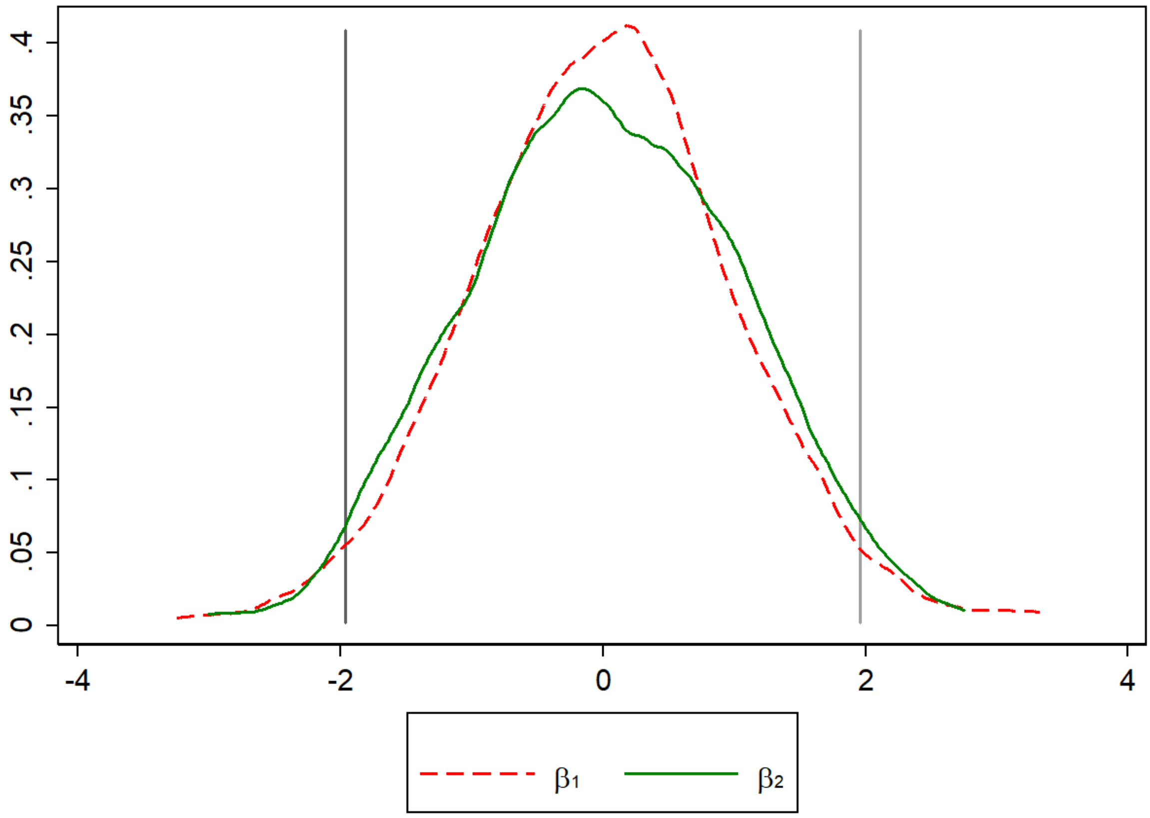
Note: favouritism is a categorical variable that takes values ranging from 1 (thinks that sibling receives a lot more attention) to 5 (thinks that he/she receives a lot more attention than sibling). It is normalized to have mean 0 and standard deviation 1.

Figure C.1: Educational Polygenic Scores (Normalized). Kernel Density Estimates



Note: firstborns and twins are included in the sample. No. obs.: 869. This Figure displays kernel-smoothed densities of Add Health respondents' own EPGS and those of their next younger siblings' or twins' EPGS for firstborns and twins, respectively.

Figure C.2: Distribution of Placebo t-values



Note: This graph shows the distributions of the t-values of the tests $\beta_1 = 0$ and $\beta_2 = 0$ obtained when estimating 500 placebo regressions of equation (2). Dependent variable: parental investment index. Sample of firstborns. To obtain placebo values of $EPGS_1$ and $EPGS_1 - EPGS_2$, their actual values are replaced with those from randomly chosen individuals from our sample. See the note to Table 3 for the full list of controls included in the placebo regressions. Number of placebo trials: 500.