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

Anna Sanz-de-Galdeano, Anastasia Terskaya

September 10, 2020

Abstract

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? We use genetic variants that predict educational attainment, which are fixed at conception, as a measure of children's educational endowments. We find that parents of singleton siblings display inequality aversion: given the absolute endowment level of one child, they invest less in him/her if his/her sibling has a lower genetic predisposition to educational attainment. In contrast, parents of twins do not significantly react to endowment differences among their children.

JEL classification: D13, D64, J13, J24

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

*Anna Sanz-de-Galdeano: FAE, Universidad de Alicante, Carretera de San Vicente s/n, 03080 San Vicente – Alicante, Spain. Anastasia Terskaya: Department of Economics, University of Navarra, Campus Universitario, s/n, 31009 Pamplona, 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 Alejandra Abufhele, Tiziano Arduini, Jere Behrman, Pietro Biroli, Antonio Cabrales, David Cesarini, Damian Clarke, Lola Collado, Dalton Conley, Jason Fletcher, Jorge García-Hombrados, Daniel Hamermesh, Daniel Rees, Núria Rodríguez-Planas, Peter Savelyev, and Alexander Young, as well as seminar participants at IFN (Stockholm), University of Alicante, University of Melbourne, University of Padova, Universidad Alberto Hurtado, Universidad Pablo de Olavide, Universidad Autónoma de Madrid, Aarhus University, University of Munich (LMU), CUNEF, Universidad de Navarra, Aalto University, University of Naples Federico II, University of Hamburg, and University of Bologna for their comments. We also thank the participants to the 3rd IZA Workshop on Gender and Family Economics, the 2019 Meeting of the Society of Economics of the Household, the 2019 Russell Sage Foundation Summer Institute in Social-Science Genomics, the 18th IZA/SOLE Transatlantic Meeting for Labor Economists (TAM), and the 2019 Spanish Economic Association Meeting for their feedback.

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 plays a central role 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; Savelyev et al., 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; Behrman et al., 1994). Some authors have also uncovered different patterns of parental behavior across contexts or socio-demographic groups.[3] Interestingly, Yi et al. (2015) provide

[1] Since Behrman et al. (1982) assume that the cost of adding to quality or the price per year of schooling is the same for all children and hence unrelated to their 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.

[3] For instance, Behrman (1988b) found parents are more prone to reinforcing behavior during the lean season in India when food is scarcest, while they follow more closely a compensating strategy in the more abundant season. There is also evidence that parents tend to favor older children Behrman (1988b) and males Behrman (1988a) in India. Other studies have looked into differential patterns by maternal education or socioeconomic status. Some within-family studies have uncovered no significant patterns in the endowment-investment relationship by maternal education (Abufhele et al., 2017; Datar et al., 2010), and others have found that low-educated mothers reinforce sibling differences while high-educated mothers follow a compensating strategy (Hsin, 2012; Restrepo, 2016; Bhalotra and Clarke,

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 educational 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.[4] 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 in their human capital—or the price effect (Becker and Tomes, 1976). We motivate our empirical analysis by means of a general parental preference model that incorporates both mechanisms as previously proposed by Terskaya (2019). 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 in 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

2019a). Breinholt and Conley (2019) find that the link between parental investments and children's genetic propensity towards educational success does not vary by socioeconomic status.

[4]In what follows we will often refer to educational polygenic scores as educational genetic endowments in the sense that genetic variants are fixed at conception and hence beyond individuals' control. However, one should note that genes may influence educational attainment not only through a direct or biological channel but also by evoking environmental responses and by causing individuals to select into environments (Jencks, 1980). Therefore, the association between educational polygenic scores and educational attainment and other outcomes may depend on the environment; this association is not deterministic but probabilistic.

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 score—a summary measure of genome-wide genetic variants that predict educational attainment—as an indicator of children's educational genetic endowments.[5] 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,[6] while endowment indicators measured at birth (*e.g.*, birth weight) for singleton siblings may reflect differences in pre-natal investment decisions (Del Bono et al., 2012; Almond and Currie, 2011; Currie, 2011, among others).[7] 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.[8] 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,[9] most previous studies have focused on the latter.[10] This may be due to the fact that at-birth measures of endowments (other than birth weight), which are less likely to suffer from reverse causality than indicators measured later on in life, are not readily available.[11] Be that as it may, we know much less about parental responses to

[5] In our data up to 4.7% of the variation in educational attainment is explained by the educational polygenic score.

[6] 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.

[7] For an analysis of the relationship between birthweight and schooling, adult physical attributes, and earnings, see Behrman and Rosenzweig (2004) and the references therein.

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

[9] See for instance Behrman et al. (1982), Ayalew (2005), Frijters et al. (2013), and Garcia Hombrados (2017).

[10] See for instance Rosales-Rueda (2014), Yi et al. (2015), Terskaya (2019), Datar et al. (2010), Hsin (2012), Cabrera-Hernández and Orraca-Romano (2016), Restrepo (2016), Abufhele et al. (2017), Bharadwaj et al. (2018), Halla and Zweimueller (2014), Adhvaryu and Nyshadham (2016), Halla and Zweimueller (2014), and Bhalotra and Clarke (2019a).

[11] Most of the studies relying on at-birth indicators focus on birth weight (Datar et al., 2010; Hsin,

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 worse-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).[12] 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; Benjamin et al., 2012; Conley and Fletcher, 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; Ronda et al., 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 separately assessing the relevance of both children's absolute and relative (with respect to their siblings) educational genetic endowments for parental investment decisions.[13] Our results suggest that the effect of individuals' genetic predisposition for education on future outcomes is not

2012; Cabrera-Hernández and Orraca-Romano, 2016; Restrepo, 2016; Abufhele et al., 2017; Bharadwaj et al., 2018; Bhalotra and Clarke, 2019a). Adhvaryu and Nyshadham (2016) use variation in in-utero exposure to a iodine supplementation programme, and Halla and Zweimueller (2014) use in-utero exposure to the radioactive fallout from the Chernobyl accident.

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

[13] Breinholt and Conley (2019) rely on between-family comparisons to study the link bewteen parental investments and children's absolute levels of genetic predisposition towards educational success. Behrman et al. (1994) and Savelyev et al. (2019) exploit differencees between allocations for identical versus nonidentical twins to estimate the effects of relative endowments on parental investment decisions.

only direct, but it may also operate through intra-household allocation decisions.

The remainder of the paper is organized as follows. The next section lays out a model that guides our empirical estimation. In Section 3, 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 4 we describe the Add Health dataset used, we describe our measures of genetic predisposition to educational attainment, and we show that they correlate with several education-related indicators in our sample. Section 5 discusses our estimation results, Section 6 presents robustness checks, and 7 concludes.

2 Theoretical Model

We propose a theoretical framework that builds upon the classical intra-households allocation models of Becker and Tomes (1976) and Behrman et al. (1982) in order to illustrate how parental investment decisions depend on children's endowments. This framework is similar to the one presented in Terskaya (2019) to study the schooling gap between disabled and non-disabled individuals in Mexico.

We assume that parental preferences can be represented by the utility function $U_p = U_p(c, V_1, .., V_n)$, where c denotes parental consumption and V_i is the quality of child i. For simplicity let us assume that there are only two-child families. We also assume that parental preferences are separable in consumption, which allows one to analyse the allocation of resources among children without regard to parental consumption. In particular, we specify parental preferences using a CES utility function as in Behrman et al. (1982):

$$U = \left\{ V_1^{\rho} + V_2^{\rho} \right\}^{\frac{1}{\rho}} \tag{1}$$

The main advantage of this utility form is that ρ measures the degree of parental inequality aversion across children. When $0 < \rho < 1$ parents do not dislike inequality and, instead, care about efficiency. In this case, parents follow a "reinforcement strategy". In the extreme case of linear preferences, parents are indifferent about inequality in the distribution of their children's quality, and they maximize the sum of expected human capital of their children. When $\rho < 0$, parents are more concerned about equality than efficiency —they are inequality averse. The extreme case of inequality aversion is the Rawlsian case, where parents will compensate differences as long as the marginal returns to their inputs positively depend on children's endowments. When $\rho = 0$, parents trade off equality and efficiency.

Following Behrman et al. (1982), we assume that a child's quality function has the following form:

$$V(e_i, PI_i) = e_i^{\alpha_e} PI_i^{\alpha_p} \tag{2}$$

, where e_i denotes the genetic endowment of child i and PI_i denotes parental inputs devoted to child i. Positive and diminishing returns to parental inputs require $0 < \alpha_p < 1$, and positive returns to genetic endowments imply that $\alpha_e > 0$.

Note that, with this function, marginal returns to parental inputs are positively related to genetic endowments. That is, endowments and parental inputs are complements in the production of human capital. The complementarity between parental inputs and children's endowments is important in our context because it introduces an efficiency vs. equality trade-off in parental investment decisions. While some may question this assumption at early ages of childhood, there is strong empirical evidence of complementarities between skills and investments at later stages of childhood (Cameron and Heckman, 2001; Cunha et al., 2006; Cunha and Heckman, 2008; Cunha et al., 2010).[14] In our analysis we focus on parental investment in adolescents, for whom skills and parental inputs are likely to be complements as in (2).[15]

Finally, the parental budget constraint has the following form:

$$p_1 PI_1 + p_2 PI_2 = TPI \tag{3}$$

, where p_i is the cost of parental inputs for child i, whereas by TPI we denote total parental investment in children.[16] Furthermore, following Becker and Tomes (1976), we allow the cost of parental inputs to differ with children's initial endowments e, assuming in addition that $p_i = p(e_i)$ is not increasing in e and that it is homogeneous of degree one.

In Appendix A we solve the utility maximization problem (1) subject to (2) and (3), which yields the following expression for parental investment in child 1:

$$log(PI_1) = log(TPI) + G(e_1) + F\left(\frac{e1}{e2}\right) \tag{4}$$

, where $G(e_1) = -log(p(e_1))$, and $F(\frac{e1}{e2})$ is a function of the parameters of the model

[14] See Heckman and Mosso (2014) for an extensive review.

[15] One may argue that parental investments in teenagers depend on endowments (or the skills stock) in adolescence rather than on genetic endowments (which are fixed at conception). Our conclusions would be the same if we used endowment indicators measured in adolescence instead of genetic endowments under the assumption that differences in endowments due to differences in genetic endowments have not been completely eliminated by the time individuals reach adolescence. To our knowledge, there is no empirical evidence that differences in capability due to differences in genetic endowments decrease throughout development stages. To the contrary, one of the best documented and most replicated findings in behavioral genetics is that the genetic influence on intelligence increases throughout development (Plomin et al., 2016).

[16] Note that both the cost of parental inputs and total investment include monetary as well as non-pecuniary expenditures such as time.

and $\frac{e_1}{e_2}$.

This equation shows that parental inputs for child 1 positively depend on total parental investment in children, while they negatively depend on the price of parental inputs for child 1 ("the price effect") and, since p_1 in turn negatively depends on child 1's endowment, parental inputs for child 1 positively depend on his/her endowment (holding endowment differences across siblings constant). Furthermore, (3) indicates that parental inputs for child 1 depend on the relative endowment of child 1 with respect to his/her sibling ($\frac{e1}{e2}$).

Additionally, it can be shown that the following holds (see Appendix A for derivations):

- $\frac{\partial log(PI_1|e_1)}{\partial\left(\frac{e_1}{e_2}\right)} < 0$ iff $\rho < 0$
- $\frac{\partial log(PI_1|e_1)}{\partial\left(\frac{e_1}{e_2}\right)} > 0$ iff $0 < \rho < 1$
- $\frac{\partial log(PI_1|e_1)}{\partial\left(\frac{e_1}{e_2}\right)} = 0$ iff $\rho = 0$

Hence, an increase in a child's relative genetic endowment with respect to his/her sibling's genetic endowment (holding constant the child's own absolute endowment level) decreases parental investments in this child if and only if parents are inequality averse, while it increases parental investments in this child if and only if parents care more about efficiency than equality. Therefore, to infer whether parents are inequality averse or if they favour efficiency over equality, one should look at how parental investments in child i depend on his/her genetic endowment relative to his/her sibling's endowment while controlling for his/her own genetic endowment level (that is, after taking the "price effect" into account).

3 Empirical Strategy

Our goal is to empirically distinguish between the two mechanisms considered in the previous model that may induce parents to follow a reinforcing strategy (that is, to invest more in better-endowed children than in their lower-endowed siblings) versus a compensating strategy (that is, to invest more in children with a lower relative endowment).

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 schooling differences between their children

if the cost of investing in them is sufficiently lower for non-disabled children than for their disabled siblings, a result also consistent with our model for parental investment allocations among children with different genetic predisposition towards educational attainment. 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.

3.1 Parental Preferences Regarding Equality versus the Price Effect

The theoretical framework laid out in Section 2 guides our empirical strategy, which presents an alternative 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) as in equation (4). In particular, we consider the following empirical specification:

$$PI_{if} = \beta_0 + \beta_1 * EPGS_{if} + \beta_2 * (EPGS_{if} - EPGS_{jf}) + X'_{if}\alpha + S'_{jf}\delta + F'_f\gamma + u_{if} \quad (5)$$

, 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 4. Note that $(EPGS_{if} - EPGS_{jf})$ is our measure of child i's relative genetic predisposition to educational attainment, 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 (5) 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 ($\rho < 0$), 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 ($0 > \rho > 1$), 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 ($\rho = 0$).

As for β_1 in equation (5), 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 because the cost of investing in them is lower. Note that, since $(EPGS_{if} - EPGS_{jf})$ is held constant in (5), a positive value of β_1 cannot not be attributed to parental preferences for efficiency over equality.[17]

In the following sections we highlight the identification challenges involved in the estimation of (5) and we discuss how we address them.

3.2 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 genetic endowments, as individuals' genetic makeup is fixed at conception.

3.3 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

[17] 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. Our equation (5) could be rewritten as equations (4) or (5) in Behrman et al. (1994):

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

Note however that in this case the coefficient for $EPGS_{if}$ measures the combination of the price effect and the effect of parental preferences for equality versus efficiency.

do not observe) affect children's genes as well as (potentially) parental investments (Kong et al., 2018; Wertz et al., 2019).

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 B, indicates that, in the worst case scenario, the true β_2 would be about 60% of our estimated $\hat{\beta}_2$.

In Section 5 we show that our estimated coefficients of interest barely change when controlling for parental characteristics (*e.g.*, parental socioeconomic status, education, *etc.*). In line with this evidence, we also follow Altonji et al. (2005) and Oster (2019) 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 —using the Altonji et al. (2005) ratio— times stronger or 6.3 and 35.7 times stronger —following Oster (2019) methodology— 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 4.3) 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 5 we show that our results are robust to using this relative parental investment measure as an outcome.

3.4 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 (5) 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

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

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 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 (5):

$$PI_{1f} = \beta_0 + \beta_1 * EPGS_{1f} + \beta_2 * (EPGS_{1f} - EPGS_{2f}) + X'_{1f}\alpha + S'_{2f}\delta + F'_f\gamma + u_{1f} \quad (6)$$

, 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 (5) for a sample of twins, and equation (6) 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).

4 Data and Descriptive Statistics

4.1 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 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, our measure of individuals' genetic predisposition for education (described in Section 4.2). This sample 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 (5) using the sample of twins (No. Obs.=274), and we estimate equation (6) using the sample of firstborns (No. Obs.=595).

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 4.3). 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).[18] We describe the construction of this index in Appendix D, while Table E.2 in Appendix E displays summary statistics for our independent variables in the samples of firstborn singletons and twins.

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

4.2 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),[19] 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 (5) and (6) in Section 3) to refer to this indicator.[20]

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 scanning the entire genome in order to identify SNPs that are associated with a phenotype (such as years of schooling) even after strict adjustments for multiple hypothesis testing aimed at avoiding finding false significant results. Educational polygenic scores were constructed for Add Health respondents by computing a weighted sum of these SNPs:

$$EPGS_i = \sum_{j=1}^{k} \hat{\beta}_j SNP_{ij} \tag{7}$$

where $SNP_{ij} \in \{0, 1, 2\}$ is a count of the number of reference alleles for individual i at SNP j, and $\hat{\beta}_j$ is a transformation of the underlying GWAS coefficient estimated for each SNP associated with schooling.[21] 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.[22] Subsequent GWAS based on larger sample sizes such as Okbay et al. (2016) and Lee et al. (2018) have identified more genome-wide significant SNPs, yielding polygenic scores more predictive of educational-related outcomes in general. However, our benchmark

[19] 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.

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

[21] 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.

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

analyses rely on EPGS based on Rietveld et al. (2013) for reasons we discuss in detail below.

Many of the SNPs identified in GWAS of educational attainment 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.

One should keep in mind that the association between genes and educational attainment may vary depending on environment and that genes may influence educational attainment through environmental responses or through selection into environments (Jencks, 1980). We use the term educational genetic endowment to describe educational polygenic score in the sense that in *the given environment* higher educational polygenic score predicts better educational outcomes. We do not state that this association is driven purely by biological processes.

Note also that if parents respond to children's genetic endowments, the weights used to construct polygenic scores ($\hat{\beta}_j$) may reflect the direct effect of genes as well as their indirect effects through parental responses. This could be problematic because in this case the effect of educational polygenic scores may overestimate or underestimate the impact of genetic endowments (net of environmental responses) on parental investments depending on parental behaviour at earlier stages of children's lives. However, in Appendix C we show that: (i) in general, the signs of our estimated effects are correct; and (ii) the magnitude of our estimated effects is correct too because, as we will show, the price effect and the effect of parental inequality aversion cancel each other out and parents parents in our sample follow neither a reinforcing strategy nor a compensating strategy.

Since genetic variants may be associated with the phenotype because of population stratification and genetic differences across ethnic groups, it is standard GWAS practice to include principal components of the full matrix of SNP data as controls to correct for these confounding factors (Price et al., 2006; Benjamin et al., 2012).[23]

Figure E.1 in Appendix E plots the (kernel-smoothed) densities of Add Health respondents' EPGS and those of their nearest younger siblings or twins for firstborns

[23]While population stratification cannot be possibly correlated with sibling differences in EPGS, it might be correlated with individuals' own EPGS. Therefore, omission of population stratification may bias our estimates of the price effect ($\hat{\beta}_1$). To address this issue we control for population stratification using the first 10 principle components of the full SNP matrix, which is a standard practice in genomics.

and twins, respectively.[24] 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) 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),[25] grades, and the probability of dropping out of high school (Columns 3-8 of Table 2). Note that while we (researchers) know the value of Add Health siblings' educational polygenic scores, their parents do not. Instead, parents at least partially observe some of the traits related to their children's polygenic scores and hence their investment decisions can respond to them.

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 (5) and (6) on $EPGS_{1f} - EPGS_{2f}$ for the sample of firstborns (see equation (6)), and on $EPGS_{if} - EPGS_{jf}$ (see equation (5)) 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.

As indicated above, our analyses rely on EPGS constructed based on Rietveld et al. (2013). Subsequent GWAS based on larger sample sizes such as Okbay et al. (2016) and Lee et al. (2018) have identified more genome-wide significant SNPs, and EPGS based on these studies are generally more predictive of completed schooling. However, we rely on EPGS based on Rietveld et al. (2013) for two main reasons. First, the samples of genotyped siblings with available EPGS based on Okbay et al. (2016) and Lee et al. (2018) are smaller in Add Health, where EPGS based on Rietveld et al. (2013) are available for twin and full sibling respondents who provided saliva samples at Wave IV (Domingue et al., 2015). Second, in the sample of Add Health siblings, EPGS based on Rietveld et al. (2013) are more strongly and precisely associated with

[24] 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.

[25] 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.

early manifestations of educational genetic endowments such as PPVT in family fixed effects models than those based on Okbay et al. (2016) and Lee et al. (2018) (see Table E.3 in Appendix E). These three GWAS of educational attainment (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018) have used number of years of schooling completed as a phenotype, while the parental investment indicators we analyze (described in detail in the next section) are measured when children were teenagers. As teenagers have not completed schooling yet, our parental investment measures can only possibly reflect earlier manifestations of EPGS such as PPVT. Additionally, our intra-household allocation framework aims at investigating how parents respond to differences in their children's endowments. Therefore, we choose the EPGS based on Rietveld et al. (2013) because they are the ones most strongly and precisely associated with PPVT *within* families in the Add Health Sibling sample.[26]

4.3 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.[27]

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.

[26] As expected, if we use the number of years of schooling completed as a phenotype, EPGS based on Okbay et al. (2016) and Lee et al. (2018) are more predictive than those based on Rietveld et al. (2013) in the Add Health Sibling sample.

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

Table 2: Educational Polygenic Scores, Years of Schooling and Other Education-Related Indicators

Dependent variable:	(1) Years of education	(2)	(3) PPVT	(4)	(5) Overall GPA	(6)	(7) High-School Drop-Out	(8)
EPGS. Normalized	0.498***	0.291***	4.650***	2.162***	0.278***	0.152***	-0.0414***	-0.0260**
	(0.0717)	(0.0707)	(0.435)	(0.436)	(0.0325)	(0.0354)	(0.00988)	(0.0102)
Controls included		yes		yes		yes		yes
Observations	869	869	833	833	645	645	869	869
R-squared	0.047	0.281	0.105	0.310	0.094	0.306	0.018	0.094

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, the socoeconomic status index, and the first 10 principle components of the full SNP matrix. 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. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

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:[28] *"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.

Table E.5 in Appendix E describes the main outcomes measuring parental investments as well as their components.

5 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 (6) 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 (5) 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

[28] 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 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. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$.

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 child 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 64%, a 48%, and a 78% of the (positive) impacts that a standard deviation increase in families' socioeconomic status has on parental, maternal and

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	0.0871
	(0.0446)	(0.0487)	(0.0534)	(0.0603)
EPGS. Normalized	0.116***	0.120**	0.0574	-0.0757
	(0.0435)	(0.0526)	(0.0740)	(0.0977)
Observations	583	583	272	272
R-squared	0.015	0.087	0.005	0.133
Panel B: Maternal Investment Summary Index. Normalized				
EPGS - Sibling's EPGS. Normalized	-0.0987**	-0.104**	0.00387	0.0498
	(0.0456)	(0.0502)	(0.0574)	(0.0617)
EPGS. Normalized	0.0852*	0.0846	-0.0307	-0.143
	(0.0456)	(0.0546)	(0.0714)	(0.0898)
Observations	568	568	265	265
R-squared	0.009	0.083	0.001	0.149
Panel C: Paternal Investment Summary Index. Normalized				
EPGS - Sibling's EPGS. Normalized	-0.138**	-0.145**	0.0436	0.119
	(0.0542)	(0.0575)	(0.0654)	(0.0731)
EPGS. Normalized	0.142***	0.150**	0.155*	-0.00797
	(0.0517)	(0.0587)	(0.0917)	(0.116)
Observations	411	411	195	195
R-squared	0.020	0.115	0.031	0.189
Panel D: Favouritism Indicator (Relative Parental Investment). Normalized				
EPGS - Sibling's EPGS. Normalized	-0.115**	-0.144***	-0.00359	0.0819
	(0.0512)	(0.0534)	(0.0699)	(0.0675)
EPGS. Normalized	0.128**	0.167***	0.0366	-0.141*
	(0.0511)	(0.0586)	(0.0810)	(0.0794)
Observations	423	423	217	217
R-squared	0.016	0.079	0.001	0.118
Individual and family controls		YES		YES

Note: OLS coefficient estimates and their associated robust standard errors in parentheses. For twins standard errors are clustered at the family level. The regressions in Columns (2) and (4) include the following controls: age, age squared, a female dummy, 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, a socioeconomic status index, and the first 10 principle components of the full SNP matrix. 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. *** $p < 0.01$, $** p < 0.05$, $* p < 0.1$.

paternal investments in their children, respectively.[29]

In section 3.3 we have acknowledged that we do not observe parental genes. Genes are randomly allocated across siblings, so $EPGS_1 - EPGS_2$ is uncorrelated with

[29]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.

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 (6) may bias both $\hat{\beta}_1$ and $\hat{\beta}_2$. However, as shown in Appendix B, 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 37%, a 30%, and a 42% 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. We reach the same conclusion when following the methodology proposed by Oster (2019), which takes into account both coefficient and R-squared movements.[30] 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.[31]

[30] In particular, we compute the relative degree of selection on unobservables versus observables assuming the recommended value of $R_{max} = 1.3 * \tilde{R}^2$ (Oster, 2019), where $\tilde{R}^2$ is the R^2 from from the regression with controls, and R_{max} is the R^2 from a hypothetical regression of the outcome on both observed and unobserved controls. We find that the impact of unobserved factors would have to be at least 32.2 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.

[31] This is consistent with the findings of Agostinelli and Wiswall (2016), who show that maternal

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 B. 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 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) and Oster (2019) 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 should be at least 30 —following Altonji et al. (2005)— or 6.2 —following Oster (2019)— 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.[32] This is a relevant result for the literature on intra-household resource allocation, as it may explain why previous empirical studies relying on family

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).

[32] 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.

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).[33]

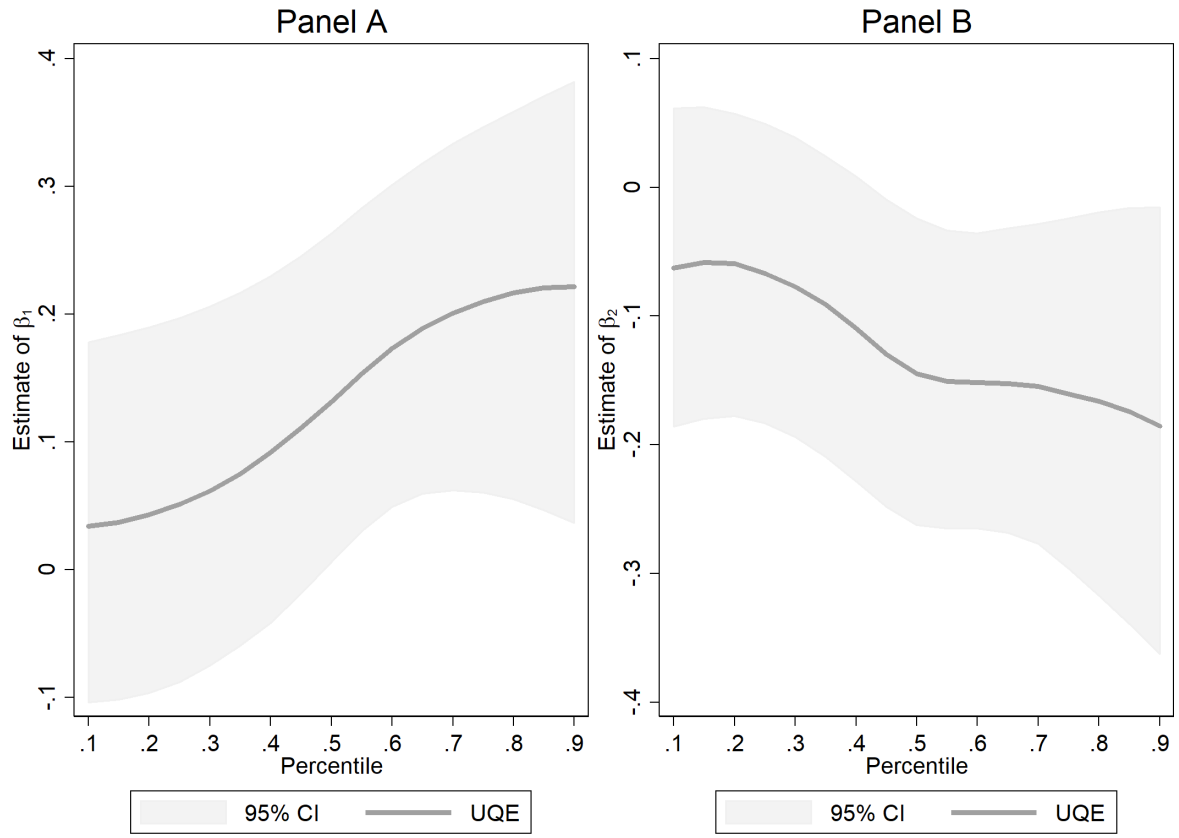
In sum, our evidence for non-twin siblings is clearly supportive of parents displaying inequality aversion such that, for a given level of a child's genetic predisposition for education, they reallocate resources to invest more in him/her if his/her sibling is better-endowed. This finding is very much 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.

Finally, we investigate whether the price effect and the degree of inequality aversion we have previously uncovered for parents of non-twin siblings vary along the parental investment distribution. Table 5 and Figure 1 summarize the results of estimating equation (6) 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

[33]One recent exception is Savelyev et al. (2019), who rely on comparisons between monozygotic and dizygotic twins, as initially proposed by Behrman et al. (1994), and find evidence indicative of compensating behavior with respect to health endowments.

an investment level that allows them to be aware of children's endowments and endowment differences among children.[34]

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.

[34]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) 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	-0.0875	-0.0271	-0.154**	-0.154***	-0.176***	-0.120**	-0.160**	-0.235**
	(0.0543)	(0.0592)	(0.0622)	(0.0572)	(0.0549)	(0.0588)	(0.0755)	(0.0994)
EPGS. Normalized	0.0384	0.0251	0.111	0.143**	0.197***	0.209***	0.223***	0.225**
	(0.0626)	(0.0671)	(0.0677)	(0.0625)	(0.0613)	(0.0648)	(0.0851)	(0.111)
R-squared	0.110	0.103	0.096	0.082	0.082	0.086	0.066	0.077

Note: Unconditional quantile regression estimates (UQUE) 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, 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, a socioeconomic status index, and the first 10 principle components of the full SNP matrix. 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. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

6 Robustness Checks

6.1 Falsification Tests

6.1.1 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 drawn individuals from our sample of firstborns (denoted by $EPGS_1^P$ and $EPGS_1^P - EPGS_2^P$).[35] We then estimate equation (6) 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 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 E.2 in Appendix E summarizes the results of these estimations by displaying the placebo t-value distributions of the tests $\beta_1 = 0$ and $\beta_2 = 0$.

6.1.2 "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 (5) and (6) 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 in the sample of firstborns ($\hat{\beta}_1 = -0.020$ with $SE(\hat{\beta}_1) = 0.028$ $\hat{\beta}_2 = -0.016$ with $SE(\hat{\beta}_2) = 0.024$).

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

6.2 GWAS of Educational Attainment, Educational Polygenic Scores, and Ancestry Groups

To date, all GWAS of educational attainment rely on samples of European-descent individuals, so EPGS may be much less predictive for other groups (Lee et al., 2018). The general recommendation is therefore to perform analyses separately by ancestral groups or, at least, include the principal components of the genetic data to help account for potential bias due to population stratification (Price et al., 2006).

We do the latter, but we do not exclude individuals with non-European ancestry to maximize sample size. Note, however, that that our primary focus is on sibling comparisons, so our main result is not expected to be affected by potential bias due to differences in genetic structure across ancestry groups.

At any rate, in Appendix Table E.4 we assess the robustness of our main findings to the standardization of EPGS within ancestry groups to have a mean 0 and a standard deviation 1 (in Column 1), and to the exclusion of individuals who do not self-identify as non-Hispanic White (in Columns 2-5). These results are remarkably similar to our benchmark findings for firstborns displayed in Panel A of Table 4 (Column 2).

7 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 lower (higher) 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 in-

crease 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 (2019b) 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., J. Behrman, and D. Bravo (2017). Parental preferences and allocations of investments in children's learning and health within families. *Social Science & Medicine 194*, 76–86.

Adhvaryu, A. and A. Nyshadham (2016). Endowments at birth and parents' investments in children. *The Economic Journal 126*(593), 781–820.

Agostinelli, F. and M. Wiswall (2016, July). Estimating the technology of children's skill formation. Working Paper 22442, National Bureau of Economic Research.

Aizer, A. and F. Cunha (2012, September). The production of human capital: Endowments, investments and fertility. Working Paper 18429, National Bureau of Economic Research.

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives 25*(3), 153–72.

Almond, D. and B. Mazumder (2013). Fetal origins and parental responses. *Annu. Rev. Econ. 5*(1), 37–56.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*(1), 151–184.

Ayalew, T. (2005). Parental preference, heterogeneity, and human capital inequality. *Economic Development and Cultural Change 53*(2), 381–407.

Barth, D., N. W. Papageorge, and K. Thom (2019). Genetic endowments and wealth inequality. *Journal of Political Economy (forthcoming)*.

Beauchamp, J. P., D. Cesarini, M. Johannesson, M. J. van der Loos, P. D. Koellinger, P. J. Groenen, J. H. Fowler, J. N. Rosenquist, A. R. Thurik, and N. A. Christakis (2011). Molecular genetics and economics. *Journal of Economic Perspectives 25*(4), 57–82.

Becker, G. S. and N. Tomes (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy 84*(4, Part 2), S143–S162.

Behrman, J. R. (1988a). Intrahousehold allocation of nutrients in rural india: Are boys favored? do parents exhibit inequality aversion? *Oxford Economic Papers 40*(1), 32–54.

Behrman, J. R. (1988b). Nutrition health birth order and seasonality: intrahousehold allocation among children in rural india. *Journal of Development Economics 28*(1), 43–62.

Behrman, J. R., R. A. Pollak, and P. Taubman (1982). Parental preferences and provision for progeny. *Journal of Political Economy 90*(1), 52–73.

Behrman, J. R. and M. R. Rosenzweig (2004). Returns to birthweight. *Review of Economics and Statistics 86*(2), 586–601.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1994). Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of Political Economy 102*(6), 1131–1174.

Belsky, D. W., B. W. Domingue, R. Wedow, L. Arseneault, J. D. Boardman, A. Caspi, D. Conley, J. M. Fletcher, J. Freese, P. Herd, et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences 115*(31), E7275–E7284.

Belsky, D. W. and S. Israel (2014). Integrating genetics and social science: Genetic risk scores. *Biodemography and Social Biology 60*(2), 137–155.

Benjamin, D. J., D. Cesarini, C. F. Chabris, E. L. Glaeser, D. I. Laibson, G. S.-R. S. Age, V. Guðnason, T. B. Harris, L. J. Launer, S. Purcell, et al. (2012). The promises and pitfalls of genoeconomics. *Annu. Rev. Econ. 4*(1), 627–662.

Bhalotra, S. and D. Clarke (2019a, 12). The Twin Instrument: Fertility and Human Capital Investment. *Journal of the European Economic Association*. jvz058.

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

Bharadwaj, P., J. P. Eberhard, and C. A. Neilson (2018). Health at birth, parental investments, and academic outcomes. *Journal of Labor Economics 36*(2), 349–394.

Breinholt, A. and D. Conley (2019). Child-driven parenting: differential early childhood investmnt by offsrping genotype. Mimeo, Available at http://paa2019.populationassociation.org/uploads/191088.

Cabrera-Hernández, F. and P. Orraca-Romano (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.

Cameron, S. V. and J. J. Heckman (2001). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy 109*(3), 455–499.

Cawley, J., E. Han, and E. C. Norton (2011). The validity of genes related to neurotransmitters as instrumental variables. *Health Economics 20*(8), 884–888.

Conley, D. and J. Fletcher (2017). *The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future*. Princeton University Press.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

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

Currie, J. (2011). Inequality at birth: Some causes and consequences. *American Economic Review 101*(3), 1–22.

Datar, A., M. R. Kilburn, and D. S. Loughran (2010). Endowments and parental investments in infancy and early childhood. *Demography 47*(1), 145–162.

Davey Smith, G. and G. Hemani (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics 23*(R1), R89–R98.

Del Bono, E., J. Ermisch, and M. Francesconi (2012). Intrafamily resource allocations: a dynamic structural model of birth weight. *Journal of Labor Economics 30*(3), 657–706.

Domingue, B. W., D. W. Belsky, D. Conley, K. M. Harris, and J. D. Boardman (2015). Polygenic influence on educational attainment: New evidence from the national longitudinal study of adolescent to adult health. *AERA open 1*(3).

Ejrnæs, M. and C. C. Pörtner (2004). Birth order and the intrahousehold allocation of time and education. *Review of Economics and Statistics 86*(4), 1008–1019.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Fletcher, J. M. and S. F. Lehrer (2011). Genetic lotteries within families. *Journal of Health Economics 30*(4), 647–659.

Frijters, P., D. W. Johnston, M. Shah, and M. A. Shields (2013). Intrahousehold resource allocation: do parents reduce or reinforce child ability gaps? *Demography 50*(6), 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.

Halla, M. and M. Zweimueller (2014). Parental response to early human capital shocks. IZA Working Paper 7968.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annu. Rev. Econ. 6*(1), 689–733.

Hsin, A. (2012). Is biology destiny? birth weight and differential parental treatment. *Demography 49*(4), 1385–1405.

Jencks, C. (1980). Heredity, environment, and public policy reconsidered. *American Sociological Review*, 723–736.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Kong, A., G. Thorleifsson, M. L. Frigge, B. J. Vilhjalmsson, A. I. Young, T. E. Thorgeirsson, S. Benonisdottir, A. Oddsson, B. V. Halldorsson, G. Masson, et al. (2018). The nature of nurture: Effects of parental genotypes. *Science 359*(6374), 424–428.

Lee, J. J., R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. K. Linnér, et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics 50*(8), 1112–1121.

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

McQueen, M. B., J. D. Boardman, B. W. Domingue, A. Smolen, J. Tabor, L. Killeya-Jones, C. T. Halpern, E. A. Whitsel, and K. M. Harris (2015). The national longitudinal study of adolescent to adult health (add health) sibling pairs genome-wide data. *Behavior Genetics 45*(1), 12–23.

Okbay, A., J. P. Beauchamp, M. A. Fontana, J. J. Lee, T. H. Pers, C. A. Rietveld, P. Turley, G.-B. Chen, V. Emilsson, S. F. W. Meddens, et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature 533*(7604), 539.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Papageorge, N. W. and K. Thom (2019). Genes, education, and labor market outcomes: Evidence from the health and retirement study. *Journal of the European Economic Association (forthcoming)*.

Pitt, M., M. Rosenzweig, and M. N. Hassan (1990). Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries. *American Economic Review 80*(5), 1139–56.

Plomin, R., J. C. DeFries, V. S. Knopik, and J. M. Neiderhiser (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on psychological science 11*(1), 3–23.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics 38*(8), 904.

Restrepo, B. J. (2016). Parental investment responses to a low birth weight outcome: who compensates and who reinforces? *Journal of Population Economics 29*(4), 969–989.

Rietveld, C. A., D. Conley, N. Eriksson, T. Esko, S. E. Medland, A. A. Vinkhuyzen, J. Yang, J. D. Boardman, C. F. Chabris, C. T. Dawes, et al. (2014). Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological Science 25*(11), 1975–1986.

Rietveld, C. A., S. E. Medland, J. Derringer, J. Yang, T. Esko, N. W. Martin, H.-J. Westra, K. Shakhbazov, A. Abdellaoui, A. Agrawal, E. Albrecht, B. Z. Alizadeh, N. Amin, J. Barnard, Baumeister, et al. (2013). Gwas of 126,559 individuals identifies genetic variants associated with educational attainment. *Science 340*(6139), 1467–1471.

Ronda, V., E. Agerbo, D. Bleses, P. B. Mortensen, and M. Rosholm (2019). Family disadvantage, gender and the returns to genetic human capital. Working paper.

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

Savelyev, P. A., B. Ward, R. F. Krueger, and M. McGue (2019). Health endowments, schooling allocation in the family, and longevity: Evidence from us twins. Working paper, The College of William and Mary, Department of Economics. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3193396.

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.

Wertz, J., T. E. Moffitt, J. Agnew-Blais, L. Arseneault, D. W. Belsky, D. L. Corcoran, R. Houts, T. Matthews, J. A. Prinz, L. S. Richmond-Rakerd, et al. (2019). Using dna from mothers and children to study parental investment in children's educational attainment. *Child development*.

Yi, J., J. J. Heckman, J. Zhang, and G. Conti (2015). Early health shocks, intrahousehold resource allocation and child outcomes. *The Economic Journal 125*(588), F347–F371.

Appendix A Theoretical Appendix (For Online Publication)

Solving the utility maximization problem (1) subject to (2) and (3) yields the following first order conditions:

$$\begin{gathered}\frac{1}{\rho}\left\{V_1(e_1,PI_1)^\rho+V_2(e_2,PI_2)^\rho\right\}^{\frac{1}{\rho}-1}\rho\alpha_p e_1^{\rho\alpha_e}PI_1^{\rho\alpha_p-1}=\lambda p_1\\ \frac{1}{\rho}\left\{V_1(e_1,PI_1)^\rho+V_2(e_2,PI_2)^\rho\right\}^{\frac{1}{\rho}-1}\rho\alpha_p e_2^{\rho\alpha_e}PI_2^{\rho\alpha_p-1}=\lambda p_2\\ PI_2=\frac{I-p_1PI_1}{p_2}\end{gathered} \tag{A.1}$$

, where λ is the Lagrange multiplier.

These conditions yield the following expression:

$$\frac{PI_2}{PI_1}=\left\{\frac{p_1}{p_2}\left(\frac{e_2}{e_1}\right)^{\rho\alpha_e}\right\}^{\frac{1}{1-\rho\alpha_p}}=\frac{I}{p_2PI_1}-\frac{p_1}{p_2} \tag{A.2}$$

Solving (A.2) for PI_1 yields:

$$PI_1^*=\frac{I\gamma}{p_1} \tag{A.3}$$

where $\gamma=\frac{1}{\left\{\left(\frac{p_1}{p_2}\right)^{\alpha_p}\left(\frac{e_2}{e_1}\right)^{\alpha_e}\right\}^{\frac{\rho}{1-\alpha_p\rho}}+1}$.

Taking logs of (A.3) we obtain the following function for parental investments:

$$log(PI_1)=log(I)+G(e_1)+F\left(\frac{e1}{e2}\right) \tag{A.4}$$

, where $G(e1)=-log(p(e_1))$ and $F(\frac{e1}{e2})=log(\gamma)$. Given that $p_i=p(e_i)$ is assumed to be a non-increasing homogeneous of degree one function of e_i, $\frac{p_1}{p_2}$ can be expressed as a function of $\frac{e_1}{e_2}$. Therefore, γ can be expressed as a function of the parameters of the model and of $\frac{e_2}{e_1}$.

Let us specify:

$$\gamma=\frac{1}{f\left(\frac{e_1}{e_2}\right)^{\frac{-\rho}{1-\alpha_p\rho}}+1}\text{, where } f\left(\frac{e_1}{e_2}\right)=\left(\frac{p(e_2)}{p(e_1)}\right)^{\alpha_p}\left(\frac{e_1}{e_2}\right)^{\alpha_e} \tag{A.5}$$

Since e_1,e_2,p_1,p_2 are positive, $f\left(\frac{e_1}{e_2}\right)>0$. Also, given that α_e and α_p are positive, and $\frac{\partial p(e)}{\partial e}\leq 0$, it follows that $\frac{\partial f\left(\frac{e_1}{e_2}\right)}{\partial\frac{e_1}{e_2}}>0$.

Since we are interested in the sign of the effect of children's relative genetic endowments ($\frac{e_1}{e_2}$) on parental inputs in child 1 (PI_1) (holding constant his/her own genetic endowment level and p_1), and on how it depends on parental inequality

aversion (ρ), we can compute the sing of $\frac{\partial log(PI_1)}{\partial f\left(\frac{e_1}{e_2}\right)}$, which is the same as the sign of $\frac{\partial log(PI_1)}{\partial \frac{e_1}{e_2}} = \frac{\partial log(PI_1)}{\partial f\left(\frac{e_1}{e_2}\right)} \underbrace{\frac{\partial f\left(\frac{e_1}{e_2}\right)}{\partial \frac{e_1}{e_2}}}_{>0}$. This yields the following expression:

$$\frac{\partial log(PI_1)}{\partial f\left(\frac{e_1}{e_2}\right)} = \gamma f\left(\frac{e_1}{e_2}\right)^{\frac{-\rho}{1-\alpha_p\rho}-1} \frac{-\rho}{1-\alpha_p\rho} \tag{A.6}$$

Given that γ, $f\left(\frac{e_1}{e_2}\right)$ and $(1-\alpha_p\rho)$ are always positive, the sign of this partial effect depends uniquely on the level of parental inequality aversion ρ. Specifically:

- $\frac{\partial log(PI_1)}{\partial f\left(\frac{e_1}{e_2}\right)} < 0$ iff $\rho < 0$
- $\frac{\partial log(PI_1)}{\partial f\left(\frac{e_1}{e_2}\right)} > 0$ iff $0 < \rho < 1$
- $\frac{\partial log(PI_1)}{\partial f\left(\frac{e_1}{e_2}\right)} = 0$ iff $\rho = 0$

Appendix B Parental Investment Equation with Unobserved Parental Genes

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

$$PI_{1f} = \beta_0 + \beta_1 EPGS_{1f} + \beta_2(EPGS_{1f} - EPGS_{2f}) + \gamma EPGS_{pf} + \epsilon_{1f} \tag{B.7}$$

, 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:

$$EPGS_{1f} = EPGS_{pf} + v_{1f} \tag{B.8}$$

$$EPGS_{2f} = EPGS_{pf} + v_{2f} \tag{B.9}$$

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:

$$L(EPGS_p|EPGS_1, EPGS_2) = \delta_1 EPGS_1 + \delta_2 EPGS_2 \tag{B.10}$$

, 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:

$$\delta_1 = \frac{\mathbb{E}(EPGS_2^2)\mathbb{E}(EPGS_1 EPGS_p) - \mathbb{E}(EPGS_1 EPGS_2)\mathbb{E}(EPGS_2 EPGS_p)}{\mathbb{E}(EPGS_1^2)\mathbb{E}(EPGS_2^2) - \mathbb{E}(EPGS_1 EPGS_2)^2} \tag{B.11}$$

$$\delta_2 = \frac{\mathbb{E}(EPGS_1^2)\mathbb{E}(EPGS_2 EPGS_p) - \mathbb{E}(EPGS_1 EPGS_2)\mathbb{E}(EPGS_1 EPGS_p)}{\mathbb{E}(EPGS_1^2)\mathbb{E}(EPGS_2^2) - \mathbb{E}(EPGS_1 EPGS_2)^2} \tag{B.12}$$

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

$$\mathbb{E}(EPGS_p) = \mathbb{E}(EPGS_1) = \mathbb{E}(EPGS_2) = 0 \tag{B.13}$$

$$\mathbb{E}(EPGS_1^2) = \mathbb{E}(EPGS_2^2) = 1 \tag{B.14}$$

From (B.8) and (B.9) we obtain:

$$\mathbb{E}(EPGS_1 EPGS_p) = \mathbb{E}(EPGS_2 EPGS_p) = \mathbb{E}(EPGS_1 EPGS_2) \tag{B.15}$$

Finally, substituting (B.15) into (B.11) and (B.12) we obtain that:

$$\delta_1 = \delta_2 = \delta = \frac{\mathbb{E}(EPGS_1 EPGS_2) - \mathbb{E}(EPGS_1 EPGS_2)^2}{1 - \mathbb{E}(EPGS_1 EPGS_2)^2} \tag{B.16}$$

Let us rewrite equation (B.10) 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:

$$EPGS_{pf} = 2\delta EPGS_{1f} - \delta(EPGS_{1f} - EPGS_{2f}) + e_{1f} \tag{B.17}$$

,where $\mathbb{E}(e_{1f}|X) = 0$. Substituting (B.17) into (B.7) we obtain :

$$PI_{1f} = \beta_0 + (\beta_1 + 2\delta\gamma)EPGS_{1f} + (\beta_2 - \delta\gamma)(EPGS_{1f} - EPGS_{2f}) + \gamma e_{1f} + \epsilon_{1f} \tag{B.18}$$

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

Therefore, estimating equation (B.7) 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:

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

This implies that:

$$\beta_2 \leq \tilde{\beta}_2 + \frac{\tilde{\beta}_1}{2} \tag{B.19}$$

(B.19) 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 B.1 display $\tilde{\beta}_1$ and $\tilde{\beta}_2$ obtained after re-estimating our main specification without normalizing $EPGS_1 - EPGS_2$ so as to be consistent with the calculations presented in this Appendix.[36] Column 3, in turn, presents the upper bound or the worst-case scenario values of β_2 computed using (B.19). 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 B.1, these worst-case scenario true values amount to sizable shares (between 54% and 62%) of our estimated values displayed in Column 2.

[36]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.

Table B.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%

Appendix C Enviromental Influence on Educational Polygenic Scores

Let us denote years of education of individual i by S_i. PI_i denotes parental investments in child i, and G_i denotes child i's overall genetic educational endowments net of environmental factors. Note that G_i cannot be directly measured, but we have information on individuals' genetic variants that are known to affect schooling outcomes.

Without loss of generality, let us assume that individuals posses only one genetic variant denoted by SNP_i. We can then specify the following model:

$$S_i = \rho PI_i + G_i + \epsilon_i \tag{C.20}$$

where $\rho \geq 0$.

$$G_i = \gamma SNP_i \tag{C.21}$$

Genes can affect education directly through biological mechanisms, and also indirectly through parental investments. Therefore, parents may indirectly respond to these endowments through their investment decisions by following a compensating, neutral or a reinforcing strategy.

Hence, we write:

$$PI_i = \theta G_i + q_i \tag{C.22}$$

, such that $\theta > 0$ implies reinforcement, $\theta < 0$ implies compensation, and $\theta = 0$ if parents follow a neutral strategy.

In our model we use educational polygenic scores as a measure of individuals' educational genetic endowments. Educational polygenic scores $(EPGS_i)$ are obtained from GWAS studies for educational attainment, and therefore may incorporate the direct and the indirect effects of genes on education. The following regression equation is a simplified version of the estimation carried out in GWAS studies, as it only contains one SNP:

$$S_i = \alpha SNP_i + \omega_i \tag{C.23}$$

EPGS are constructed as follows:

$$EPGS_i = \alpha SNP_i \tag{C.24}$$

Substituting (C.27) into (C.20) we obtain:

$$S_i = (1 + \rho\theta)G_i + q_i + \epsilon_i \tag{C.25}$$

And substituting (C.21) into (C.25) and equating it to (C.23), we obtain:

$$S_i = (1 + \rho\theta)\gamma SNP_i + q_i + \epsilon_i = \alpha SNP_i + \omega_i \tag{C.26}$$

, which implies that $\alpha = (1 + \rho\theta)\gamma$.

Genes that affect schooling are positively related to cognitive functions (in other words, genes that are associated with higher S_i are associated with higher G_i). Hence, we only consider the case that α and γ have the same sign, or that $\rho\theta > -1$.[37]

Since we are interested in the estimation of the effect of children's genetic endowments on parental investments, let us specify the model of interest as:

$$PI_i = \delta_1 G_i + \delta_2(G_i - G_j) + u_i \tag{C.27}$$

Our main estimating equation is:

$$PI_i = \beta_1 EPGS_i + \beta_2(EPGS_i - EPGS_j) + z_i \tag{C.28}$$

This is equivalent to:

$$\begin{gathered} PI_i = \beta_1 \alpha SNP_i + \beta_2 \alpha(SNP_i - SNP_j) + z_i = \\ \beta_1(1 + \rho\theta)\gamma SNP_i + \beta_2(1 + \rho\theta)\gamma(SNP_i - SNP_j) + z_i = \\ \beta_1(1 + \rho\theta)G_i + \beta_2(1 + \rho\theta)(G_i - G_j) + z_i \end{gathered} \tag{C.29}$$

Therefore, $\beta_1 = \delta_1/(1 + \rho\theta)$ and $\beta_2 = \delta_2/(1 + \rho\theta)$ and $z_i = u_i$.

Given that $\rho\theta > -1$, the signs of δ_1 and δ_2 are always the same as the signs of β_1 and β_2, respectively.

Note that we can specify:

$$PI_i = (\delta_1 + \delta_2)G_i - \delta_2 G_j + u_i = \theta G_i + q_i \tag{C.30}$$

where $q_i = u_i - \delta_2 G_j$.

[37] Rietveld et al. (2013); Lee et al. (2018) show that genetic variants included in educational polygenic scores are associated with cognitive function and brain-development processes.

Therefore, the sum of δ_1 and δ_2 is equal to θ. Note that we can estimate:

$$\beta_1 + \beta_2 = \frac{\delta_1 + \delta_2}{1 + \rho\theta} = \frac{\theta}{1 + \rho\theta} \tag{C.31}$$

Given that the estimate of $\beta_1 + \beta_2$ that we obtain is not significantly different from zero —since the price effect and the effect of parental inequality aversion cancel out—, the estimate of θ is not significantly different from zero in our sample. Therefore, not only the sign of the estimates is correct, but so is their magnitude. In general however, the magnitude of the estimates might be underestimated or overestimated. Particularly:

- $|\beta_1| > |\delta_1|$ and $|\beta_2| > |\delta_2|$ if $\theta < 0$ (compensation)
- $|\beta_1| < |\delta_1|$ and $|\beta_2| < |\delta_2|$ if $\theta > 0$ (reinforcement)
- $\beta_1 = \delta_1$ and $\beta_2 < \delta_2$ if $\theta = 0$ (neutral strategy, our case)

Also note that the magnitude of the estimates is correct when $\rho = 0$ (when parental investments do not affect schooling). However, in general we do not need to assume that $\rho = 0$.

Appendix D 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.[38] We

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

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."*

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 E Descriptive Statistics

Table E.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 E.3: Associations between Different EPGS and PPVT within Families

	Rietveld et al. (2013)	Okbay et al. (2016)	Lee et al. (2018)
EPGS	1.642**	1.135	0.747
	(0.788)	(0.755)	(0.776)
Observations	1,412	1,300	1,300
R-squared	0.788	0.783	0.782

Note: Dependent variable: PPVT measured at Wave I. OLS coefficient estimates and standard errors clustered at the family level in parentheses. EPGS are educational polygenic scores based on different GWAS –Rietveld et al. (2013) in Column 1, Okbay et al. (2016) in Column 2, and Lee et al. (2018) in Column 3– provided by Add Health (normalized to have mean 0 and standard deviation 1). All regressions include a female dummy, age and age squared, and family fixed effects. *** $p < 0.01$, $**p < 0.05$, $*p < 0.1$.

Table E.4: 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

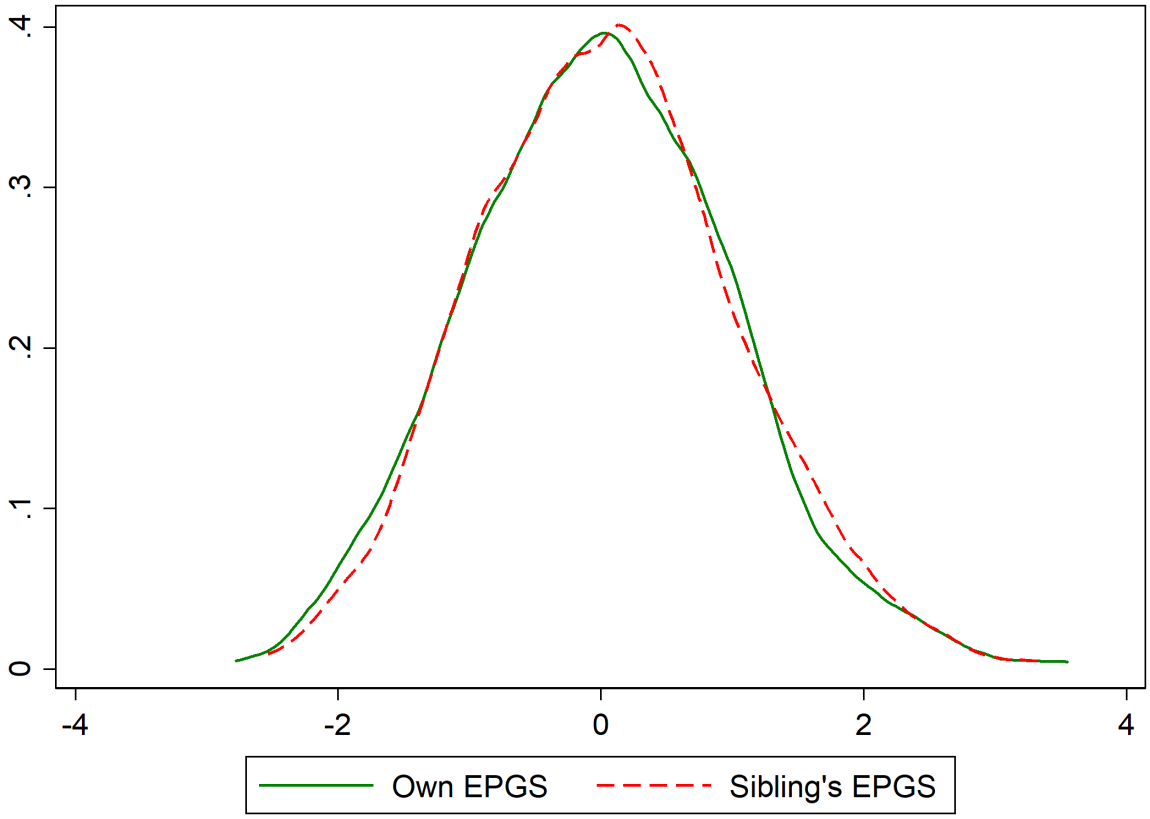
Note: Favouritism is a categorical variable that takes values ranging from 1 (thinks that sibling receives a lot more attention) to 5 (thinks that he/she receives a lot more attention than sibling). It is normalized to have mean 0 and standard deviation 1.

Table E.5: The Effect of Educational Polygenic Scores and Sibling Differences in Educational Polygenic Scores on Parental Investments. Robustness to the Exclusion of Self -Identified Ethnicity Groups

	(1) Standarized by Ethnicity	(2) Non-Hispanic Black excluded	(3) Native-Americans exluded	(4) Asians exluded	(5) Hispanic excluded
EPGS - Sibling's EPGS.	-0.124**	-0.130**	-0.130***	-0.107**	-0.135**
Normalized	(0.0490)	(0.0579)	(0.0488)	(0.0494)	(0.0534)
EPGS. Normalized	0.112**	0.0893	0.116**	0.123**	0.113**
	(0.0474)	(0.0621)	(0.0528)	(0.0540)	(0.0547)
Observations	583	426	579	550	508
R-squared	0.086	0.094	0.084	0.092	0.093

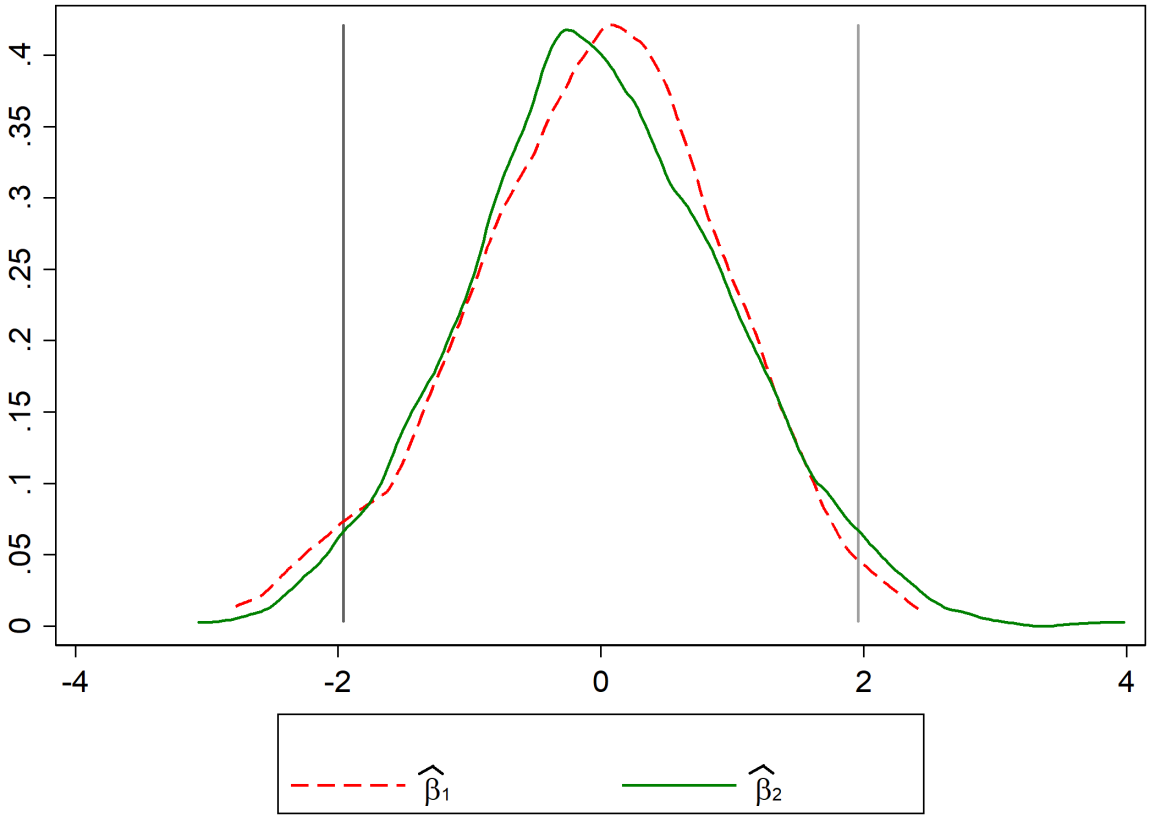
Note: Sample of firstborns. Dependent variable: parental investment index (normalized to have mean 0 and standard deviation 1). OLS coefficient estimates and their associated robust standard errors in parentheses. The following controls are included: age, age squared, a female dummy, race indicators 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, a socioeconomic status index, and the first 10 principle components of the full SNP matrix. 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 also normalized to have mean 0 and standard deviation 1. In column (1) EPGS is standarized to have mean 0 and standard deviation 1 within each ethnic group. *** $p < 0.01$, $** p < 0.05$, $*p < 0.1$.

Figure E.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 E.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 (6). 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.