WILEY

Author Query Form

Journal: Journal of Applied Econometrics

Article: jae3071

Dear Author,

During the copyediting of your manuscript the following queries arose.

Please refer to the query reference call out numbers in the page proofs and respond to each by marking the necessary comments using the PDF annotation tools.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

Query No.	Query	Remarks
Q1	AUTHOR: Please verify that the linked ORCID identifiers are correct for each author.	
Q2	AUTHOR: Please confirm that given names (Blue) and surnames/family names (Vermillion) have been identified correctly.	
Q3	AUTHOR: Please check if authors' affiliations and correspondence details are correct.	
Q4	AUTHOR: Please check if all section headings are correct and presented in their appropriate section levels.	
Q5	AUTHOR: Ref. Attanasio et al., 2020 is cited in text but not provided in the Reference List. Please check.	

Please confirm that the funding sponsor list below was correctly extracted from your article: that it includes all funders and that the text has been matched to the correct FundRef Registry organization names. If a name was not found in the FundRef registry, it may not be the canonical name form, it may be a program name rather than an organization name, or it may be an organization not yet included in FundRef Registry. If you know of another name form or a parent organization name for a “not found” item on this list below, please share that information.

FundRef name	FundRef Organization Name

SPi | Journal Code: J A E | Article ID 3 0 7 1 | Dispatch: 00.00.00 | CE: Marian Gajetos | No. of Pages: 19 | ME:

Received: 12 November 2022 | Revised: 11 March 2024 | Accepted: 1 April 2024

DOI: 10.1002/jae.3071

WILEY Journal of APPLIED ECONOMETRICS

RESEARCH ARTICLE

Quantiles of the gain distribution of an early childhood intervention

Erich Battistin[1] | Carlos Lamarche[2] | Enrico Rettore[3]

[1]University of Maryland, CEPR, IZA and FBK-IRVAPP, College Park, Maryland, USA

[2]University of Kentucky, Lexington, Kentucky, USA

[3]University of Padova and FBK-IRVAPP, Padua, Italy

Correspondence
Enrico Rettore, University of Padova and FBK-IRVAPP.
Email: enrico.rettore@unipd.it

Funding information
Financial support from Fondazione CARIPARO (Progetti di Eccellenza) is gratefully acknowledged.

Summary

We investigate the distribution of gains among participants in the Infant Health and Development Program, an understudied randomized controlled trial that targets infants with low birth weight. Our primary focus is on assessing the effects in cognitive and health outcomes within distinct subgroups, which we define based on the outcomes that would occur in the absence of program participation. We propose a strategy to estimate the distribution of gains from the program by using anthropometrics measurements taken at birth, under the assumption that potential outcomes depend on underlying latent factors explaining neonatal health. Our findings reveal that the enhancements in cognitive and health outcomes at 36 months are not uniformly distributed among program participants. The variability in these effects can be attributed to several factors, including neonatal health, post-natal shocks, and family income.

KEYWORDS

distribution of impacts, factor models, policy evaluation, quantile regression

1 | INTRODUCTION

Randomized assignment is ideal for investigating treatment effect heterogeneity because it identifies the marginal distribution of potential outcomes with and without treatment. Quantile treatment effects (QTE) serve as a valuable tool in quantifying the extent of this heterogeneity, as demonstrated in important studies by Heckman et al. (1997), Bitler et al. (2006), and Powell (2020), among others. Tests for the equality of potential outcome distributions, as in Abadie (2002) and Heckman et al. (2010), show that there might be winners and losers from policy interventions. Heterogeneous treatment effects raise several questions, including how widely gains are distributed across participants and who benefits from interventions. As individual gains cannot be identified, many of such policy-relevant questions cannot be answered in general (Heckman, 2020). For example, knowledge of QTE is not enough to assess what fraction of participants have positive returns from the intervention.

How can one learn about the distribution of gains? We revisit this question using the Infant Health and Development Program (IHDP), a randomized controlled trial aimed at promoting the development of children in poor health (IHDP, 1990). Features of the effect distribution other than its average are particularly relevant in this context, as health status during childhood is an important mechanism for transmission of economic conditions (Currie & Rossin-Slater, 2015). Besides, although IHDP is one of the few early-life interventions implemented as a randomized experiment, it has received relatively scant attention from economists to date (Chaparro & Sojourner, 2019; Chaparro et al., 2020, are recent examples). Differently from the Perry Preschool study and the Abecedarian Project, for which eligibility was income-based, IHDP targeted a demographically heterogeneous population of low birth weight (LBW) children. Moreover, IHDP combined center-based education with home visits and parent group meetings (Gross et al., 1997).

We propose an approach to study the gain from enrolling infants in IHDP at specific values of the baseline outcome, which is a causal parameter defined in Heckman et al. (1997). We depart from their framework by breaking down the status quo outcome, which represents the outcome without any intervention, into two independent components. In the context of IHDP, the first component is influenced by a set of low-dimensional factors denoted as **H**, representing endowment at birth. The second component, denoted as U_0, ranks children based on the intensity of unforeseen shocks they would have experienced after birth in the absence of any intervention. We consider a collection of policy parameters denoting quantiles of the conditional distribution (QCD) of treatment effects from enrolling children at different values of $(\mathbf{H}, U_0)$. Unlike QTE, the QCD parameters provide insights into how the distribution of gains and the proportion of individuals experiencing positive returns vary based on the outcome without intervention. These insights are particularly relevant for targeting interventions and addressing issues related to social justice (see Heckman et al., 1997). Importantly, we show that the analytical framework in which we define the QCD parameters is consistent with an economic model of human capital accumulation during early stages of life.

A virtue of this framework is to introduce flexibility in how latent components shape mobility of participants in potential outcome distributions. Specifically, we assume that the dependence between potential outcomes is generated by the factors **H** (in addition to covariates we can control for). This assumption is used in the treatment effect literature to recover the joint distribution of potential outcomes (Abbring & Heckman, 2007) and lies at the heart of many empirical applications, for example, on earning dynamics (Bonhomme & Robin, 2010) and in education (Ashworth et al., 2021). Compared with this literature, we allow **H** to affect the distribution of gains in a general way by allowing loadings to be quantile-specific, similarly to Chen et al. (2021). This framework implies that the size and sign of post-natal shocks U_0 can affect the gains from participation directly but also indirectly by attenuating or boosting the contribution of neonatal health **H**. As we demonstrate in Section 5, the empirical conclusions from IHDP data are robust to violations of the independence assumptions embedded in our factor model.

We show that QCD can be computed from quantiles of potential outcomes conditional on **H**. This relationship lays the foundation for an analogous estimation method that is computationally convenient and applicable beyond the specific case study. Our approach to estimating the QCD using IHDP data leverages the smooth relationship between quantiles of the outcome conditional on latent neonatal health and quantiles of the outcome conditional on proxies of neonatal health. Specifically, using observed measurements as proxies for latent health can introduce bias in quantile estimates due to measurement errors in these proxies. However, we demonstrate that this bias converges smoothly to zero as the measurement error in the proxies approaches zero. When measurement error in proxies is small, the measurement error variance is the leading term for this convergence (Chesher, 2017). The case of small error variance is relevant in many empirical applications where the choice of proxies is driven by the underlying theory and the measurement tools employed. Indeed, we show that this is true in the IHDP data as well.

Our main methodological contribution revolves around estimating the quantiles of potential outcomes conditional on **H**. This step entails estimating the loadings of a quantile factor model, which are then used to obtain the QCD. Our approach stands out for its computational simplicity and, in contrast to Chen et al. (2021), does not need large panel data, which is often lacking in early childhood interventions such as IHDP. In particular, our approach involves simulating the bias arising from measurement error in proxies, and then estimating quantiles of potential outcomes conditional on **H** by extrapolation to the case of no error (a recent example of extrapolation-driven identification is Arnold et al., 2022). We implement this idea by adapting Cook and Stefanski's (1994) simulation-extrapolation (SIMEX) method and show that it works well to implement our strategy when the measurement error variance is small, which is the case in IHDP data. This approximation at zero error strategy requires knowledge of the variance of the measurement error or an estimate of it, which we show how to obtain in empirical applications. The extrapolation relies on functional form assumptions for the bias introduced by measurement error. The theoretical case for this approach follows from a well-established literature in statistics which, to date, has been relatively overlooked by economists.[1]

Our empirical investigation begins by finding the best proxy of a child's latent health prior to IHDP. As health inherently reflects several prenatal endowments, we consider gestational age, birth weight, length, and head circumference as possible candidates. The latter two measurements are markers of prenatal growth and brain development (Conti et al., 2018). Low-birth weight has long been used as an indicator of poor health among newborns, and its effects on life-long

[1]Carroll et al. (2006) show that SIMEX has a number of attractive features and a high degree of accuracy in parametric and nonparametric settings. Because extrapolation is based on an unknown function, SIMEX provides an approximation rather than a consistent estimator of the quantities of interest. We assess the quality of such approximation using several simulations. We informally use identification at zero and approximation with the same meaning in what follows. We also show, in the supporting information, that SIMEX outperforms other measurement error bias correction approaches.

well-being are documented (Almond & Currie, 2011). We show that birth weight is the measurement of neonatal health with the lowest error.[2] After proxying neonatal health with birth weight, we derive our first empirical result: a child's health endowment at birth affects the whole potential outcome distribution—not only its location—with and without treatment (a fact that a traditional factor model would fail to detect). We find that, without IHDP, better neonatal health yields better and less dispersed outcomes at 36 months. However, participation in IHDP reshuffles how children are ranked in the outcome distribution and, also, changes the rank distribution conditional on neonatal health. This finding serves as a catalyst for investigating treatment effect distributions.

We then present a novel empirical finding—striking effect heterogeneity across participants during the first 3 years of life depending on neonatal health and family income. For example, the share of LBW children above 2000 g with positive returns on cognitive skills is 65% and about 20 points larger than in the LBW group with less than 2000 g. Thus, we document that the difference in average effects between children above and below 2000 g is due to large positive effects among a minority of cases. This could spark further research in the IHDP and potential investigation in other settings. Moreover, we find that within income group variability in treatment effects is at least as important as between income group differences in treatment effects. Past work studied how the effects of IHDP vary with income by looking at differences between group *averages* (e.g., Duncan & Sojourner, 2013). Our results imply that these differences mask an important component of treatment effect heterogeneity.

The results expand the empirical literature concerned with the IHDP that focuses in heterogeneity of effects with respect to observables (e.g., Brooks-Gunn et al., 1992; Duncan & Sojourner, 2013). For instance, Brooks-Gunn et al. (1992) show heterogeneous effects by the level of education of the mother and Duncan and Sojourner (2013) find heterogeneous effects by household income. We contribute to this literature by showing heterogeneous effects in two new dimensions. First, the treatment effects vary across the conditional quantile of the treatment distribution. Also, the effects vary by health conditions at birth and intensity of unanticipated shocks in the nonparticipation status. The results presented in our paper illustrate how policymakers and practitioners can assess the relative importance of latent factors in interventions.

2 | CONTEXT, DATA, AND DESCRIPTIVE STATISTICS

We use data from IHDP, a randomized controlled trial implemented in 1985 to enhance the cognitive, behavioral, and health status of low birth weight (less than 2500 g) and premature (less than 37 weeks) infants born in selected US medical institutions.[3] One unique feature of this program is the population targeted, as in early childhood development interventions eligibility typically depends on the socio-economic status. IHDP consisted of activities to foster child functioning provided in specialized institutions, pediatric follow-ups, frequent home visits, parent support groups, and a systematic educational program (see Elango et al., 2016; Gross et al., 1997; Ramey et al., 1992). The intervention lasted until the age of 36 months.

Two strata were considered at the randomization stage: the "lighter" low birth weight group (LLBW, less than 2000 g) and the "heavier" low birth weight group (HLBW, between 2000 and 2500 g). We use children from both strata as in IHDP (1990). This selection yields a sample of 985 infants, of which 377 belong to the treatment group and 608 to the control group. The size of the IHDP sample is similar to the sample sizes of other model programs in the literature.

We consider three of the outcomes in the original study, all measured at 36 months: one index of cognitive (IQ) development, the Stanford-Binet intelligence score, and two indicators of health. Specifically, we use a morbidity index defined as the total number of hospitalizations, outpatient surgeries, injuries not resulting in hospitalization or outpatient surgery, and different illnesses and conditions over the first 3 years of life (higher values of this index indicate lower health). We also use the General Health Ratings Index, which is constructed using the maternal perception of child health over the first 3 years of life (higher values of this index indicate better health). Due to missing data, our final sample includes 929 infants, of which 343 are in the HLBW group and 586 in the LLBW group. Table 1 shows descriptive statistics for the working sample, along with balancing tests to corroborate the IHDP randomization to treatment and control groups.

T1

[2]We interpret birth weight as error-ridden child's health endowment after netting off demographics that may have affected conditions at birth, as in Chaparro and Sojourner (2019).

[3]Eight institutions were involved in Little Rock, Arkansas; New Haven, Connecticut; Miami, Florida; Cambridge, Massachusetts; Bronx, New York; Philadelphia, Pennsylvania; Dallas, Texas; and Seattle, Washington. Participating sites were selected through a national competitive review. Section S.7 in the supporting information provides background information regarding the activities implemented within IHDP, including details about the recruitment of participants and the measurement of outcomes in our analyses.

TABLE 1 Balancing tests for random assignment of infants.

	All		HLBW (2000–2500 g)		LLBW (Less than 2000 g)	
	Mean	Effect	Mean	Effect	Mean	Effect
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Demographics at birth					
Weight (kg)	1.787	0.035	2.256	0.002	1.520	0.036
	[0.465]	(0.030)	[0.138]	(0.015)	[0.363]	(0.029)
Length (cm)	42.259	0.531**	45.33	0.174	40.511	0.618**
	[3.966]	(0.253)	[2.208]	(0.252)	[3.672]	(0.291)
Body mass index	9.823	−0.003	11.046	−0.055	9.127	−0.024
	[1.602]	(0.109)	[1.310]	(0.157)	[1.311]	(0.108)
Gestational age (weeks)	33.067	0.008	34.961	−0.107	31.989	−0.002
	[2.719]	(0.174)	[1.431]	(0.174)	[2.690]	(0.209)
Neonatal Health Index	99.750	0.923	98.913	0.416	100.227	1.269
	[15.446]	(1.068)	[14.371]	(1.620)	[16.025]	(1.405)
Head circumference (cm)	29.426	0.059	31.437	−0.196	28.282	0.129
	[2.553]	(0.163)	[1.348]	(0.144)	[2.363]	(0.191)
Infant is a boy	0.484	0.014	0.539	−0.042	0.453	0.047
	[0.500]	(0.034)	[0.500]	(0.055)	[0.498]	(0.043)
Infant is a first-born	0.430	0.027	0.393	0.074	0.450	0.001
	[0.495]	(0.033)	[0.490]	(0.055)	[0.498]	(0.042)
	Panel B. Mother demographics					
Age	24.863	−0.306	24.990	−1.019	24.790	0.125
	[6.148]	(0.405)	[6.346]	(0.654)	[6.04]	(0.516)
Married	0.484	−0.066*	0.485	−0.047	0.483	−0.077*
	[0.500]	(0.033)	[0.501]	(0.055)	[0.500]	(0.042)
High school graduate	0.491	−0.054	0.471	−0.091*	0.503	−0.030
	[0.500]	(0.034)	[0.500]	(0.054)	[0.501]	(0.043)
College graduate	0.120	0.004	0.155	−0.039	0.099	0.026
	[0.325]	(0.022)	[0.363]	(0.037)	[0.300]	(0.027)
Black	0.526	0.014	0.500	−0.033	0.541	0.043
	[0.500]	(0.034)	[0.501]	(0.055)	[0.499]	(0.042)
Hispanic	0.116	−0.014	0.126	−0.017	0.110	−0.012
	[0.321]	(0.021)	[0.333]	(0.035)	[0.314]	(0.026)
Observations		929		343		586

Note: Column (1) shows control-group means and standard deviations (in brackets) and column (2) shows treatment-control differences, with robust standard errors in parenthesis. Results in columns (3)–(6) are from regressions stratified for infants with birth weights lower or higher than 2000 g. All regressions control for site and birth weight group effects.*$p < 0.1$.**$p < 0.05$.***$p < 0.01$.

2.1 | Balancing tests and average effects

Treatment and control groups were comparable at initial assignment. This can be seen from Table 1, where no significant differences emerge across a number of demographics. The control group average and standard deviation of each variable at the left are reported in column (1) of the table. Column (2) shows treatment-control differences obtained from regressions of variables at the left on the treatment indicator.[4] Significant differences emerge in a limited number of cases, although the size of these differences is small when compared with standard deviations in column (1). Balancing tests in the remaining columns are for LLBW and HLBW groups and convey a similar message.

IHDP participants showed significantly higher IQ scores, on average, in both the heavier and lighter birth weight groups. This can be seen from columns (2), (6), and (10) in Panel A of Table 2, which presents treatment effect estimates from regressions of outcomes on the participation indicator controlling for mother demographics (shown in panel B of Table 1). Demographics at baseline are used in these specifications to control for residual differences in the treatment and control

T2

[4]Outcomes in regressions are not standardized, and significance is assessed using heteroskedasticity-robust standard errors. For example, the difference in length is 0.531 over a baseline of 42.259, or about 1.2% of this baseline.

TABLE 2 Average treatment effects 3 years from the assignment.

	All infants				HLBW (2000–2500 g)				LLBW (less than 2000 g)			
	Mean	Effect			Mean	Effect			Mean	Effect		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Cognitive development												
Development	84.3	9.348***	9.388***	8.942***	84.8	13.619***	13.996***	13.476***	84.1	7.153***	7.238***	6.750***
Index	[20.1]	(1.158)	(1.150)	(1.120)	[18.9]	(1.713)	(1.667)	(1.681)	[20.7]	(1.520)	(1.512)	(1.482)
Observations		855	855	855		310	310	310		545	545	545
Panel B. Health												
Morbidity	6.9	0.866***	0.885***	0.936***	6.7	0.588*	0.624*	0.639*	6.9	1.027***	1.049***	1.106***
Index	[3.1]	(0.219)	(0.218)	(0.222)	[3.0]	(0.354)	(0.360)	(0.371)	[3.1]	(0.276)	(0.277)	(0.281)
Observations		844	844	844		308	308	308		536	536	536
General health	27.1	0.418	0.376	0.291	27.6	0.945*	0.751	0.644	26.8	0.157	0.131	0.092
Ratings index	[5.1]	(0.368)	(0.367)	(0.368)	[4.7]	(0.570)	(0.573)	(0.591)	[5.2]	(0.471)	(0.476)	(0.476)
Observations		851	851	851		309	309	309		542	542	542

Note: Columns (1), (5), and (9) show control-group means and standard deviations (in square brackets) for variables listed at the left. Coefficients in columns (2), (6), and (10) show treatment-control differences for variables listed at the left, controlling for mother demographics (in Panel B of Table 1). Coefficients in columns (3), (7), and (11) are from regressions with mother demographics and randomization strata effects (site and birth weight group) as additional controls. Coefficients in columns (4), (8), and (12) are from regressions where controls include both mother demographics and infant demographics at birth (in Panel A of Table 1), and randomization strata effects. Results in columns (5)–(12) are from regressions stratified for infants with birth weights lower (LLBW) or higher (HLBW) than 2000 g. Robust standard errors are shown in brackets. Cognitive development is measured using the Stanford-Binet intelligence scale. Morbidity is defined as the presence or absence of health conditions. The index considered here is the sum, over the 3 years, of the number of hospitalizations, outpatient surgeries, injuries not resulting in hospitalization or outpatient surgery, and different illnesses and conditions. Higher scores indicate lower health. The General Health Ratings Index is constructed from the Rand Corporation Health Insurance Study using the maternal perception of child health. Higher scores indicate better perceived health.*$p < 0.1$.**$p < 0.05$.***$p < 0.01$.

groups. We find that IHDP yielded a 9 point increase in mean IQs from a baseline of about 84 points (or about 0.45σ using the standard deviation in the control group), as shown in column (1). Results for IQ were considerably stronger for HLBW children, as shown in column (6), and estimated at about 13.6 points from a baseline of about 85, or about 0.72σ. Columns (3), (7), and (11) report treatment effect estimates obtained from regressions that include randomization strata. The estimates shown in columns (4), (8), and (12) are from regressions with mother and infant demographics (Panel A of Table 1), and randomization strata.

Panel B of Table 2 shows mixed results on health. Specifically, columns (2) to (4) show that, at the age of three, IHDP participants had about one more reported episode of illness (the morbidity index effect is in the 0.87 to 0.94 range depending on the specification adopted). However, there are no statistically significant improvements, on average, using the General Health Ratings Index. In line with other influential studies, we conclude that average effects show relatively more substantive promise for reducing the risk of developmental disability later in life for LLBW premature infants.

2.2 | Treatment effect heterogeneity

F1 Figure 1 offers suggestive evidence that the benefits derived from IHDP enrollment differ among infants. Unconditional QTE estimates are obtained as in Firpo (2007) and Frölich and Melly (2013) by controlling for mother demographics. Each panel plots values generated by a local linear regression (LLR) fit to the estimated QTEs, along with confidence intervals obtained from 100 bootstrap replications. The LLR smoother uses a triangular kernel and optimal bandwidth determined from the procedure in Calonico et al. (2019) We find that QTEs vary across quantiles. For example, the smoothed line for the IQ outcome ranges between 7, at the lowest quantile, and about 11, at the median, with an overall inverted U-shape. However, the limited sample size affects the precision around estimates of QTEs, and the hypothesis of constant gains is only marginally rejected in the data (as we further discuss in Section S.8 in the supporting information).

Effect heterogeneity is likely the result of changes in a child's rank in potential outcome distributions with and without IHDP. We reach this conclusion by looking at the interaction between key neonatal anthropometrics (birth weight, length, and head circumference) and ranks in observed outcome distributions for participants and nonparticipants. The rationale for considering this relationship stems from rank similarity, which is often used as a benchmark in empirical work (see

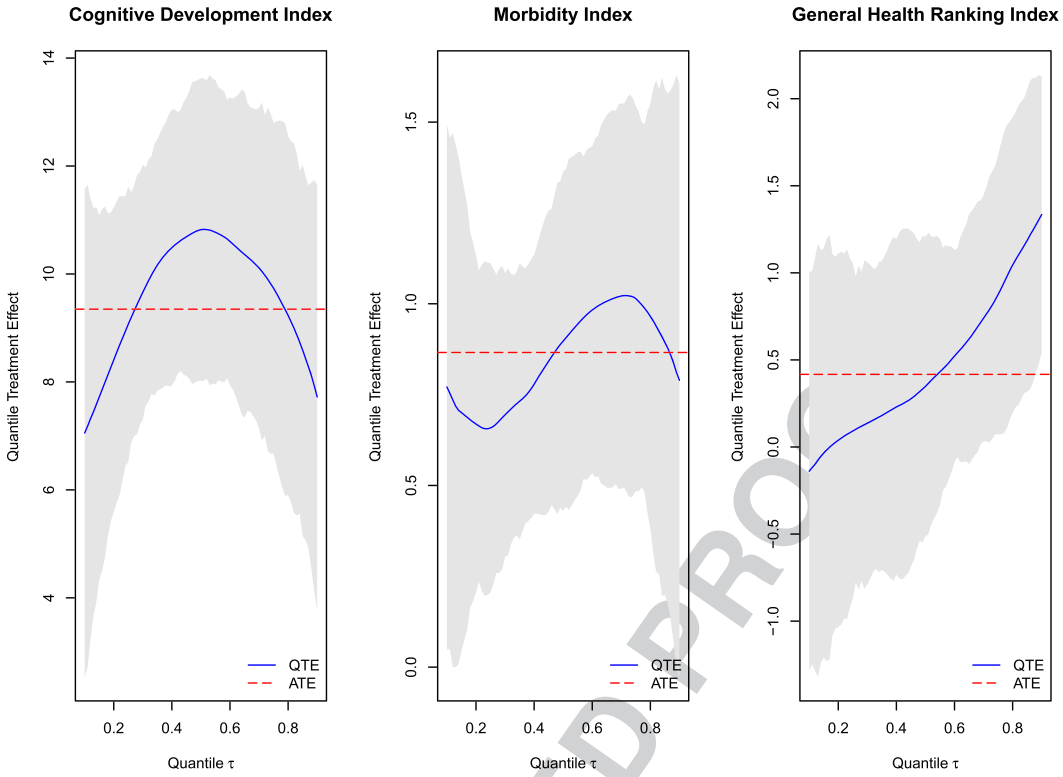

FIGURE 1 Quantile treatment effects. *Note*: The line shows estimates of quantile treatment effects (QTE) for the three outcomes. Mother demographics are included as covariates. The continuous line is generated by a local linear regression fit to estimated QTEs. Shaded areas denote 95% confidence intervals obtained from 100 bootstrap replications.

Dong & Shen, 2018 and ; Frandsen & Lefgren, 2018).[5] Specifically, under rank similarity the distribution of ranks should be identical for participants and nonparticipants conditional on anthropometrics if these variables are "rank shifters." Using the test in Frandsen and Lefgren (2018), we find that rank similarity for the cognitive ability outcome is rejected because birth weight acts as a rank shifter (see Table S.7 in the supporting information). While Figure 1 allows one to conclude that effects are heterogeneous, it does not reveal how this heterogeneity is distributed across participants.

3 | MODEL AND PARAMETER OF INTEREST

3.1 | Factor model for potential outcomes

All results hold conditionally on a vector of covariates $\boldsymbol{X}$, which we suppress for clarity. Let D be the treatment assignment indicator yielding potential outcomes (Y_1, Y_0) for IHDP-treatment and IHDP-control children, respectively. We refer to children in the IHDP-treatment sample as "participants." Children in the IHDP-control sample are called "nonparticipants." Because only $Y = Y_0 + D(Y_1 - Y_0)$ is observed, identification of the distribution of gains $\Delta = Y_1 - Y_0$ is precluded. The randomized assignment to treatment does not eliminate this problem.

We model the dependence between potential outcomes as the result of persistent, long term factors proxying conditions prior to IHDP enrollment. Specifically, for $d \in \{0, 1\}$, we assume

$$Y_d = \lambda_d(U_d)'\boldsymbol{H} + U_d, \tag{1}$$

[5]Under rank invariance, an infant's relative position in the outcome distributions remains the same with or without IHDP. Rank similarity allows for slippages in this position across distributions. For additional details, we refer there reader to Definitions 1 and 2 in Dong and Shen (2018). Rank similarity alone will not be enough to identify treatment effect distributions (or the QCD in our case). Point identification can be achieved under special circumstances, like rank invariance.

where $\boldsymbol{H} = (H_1, \ldots, H_r)'$ is a vector of r continuous random variables (factors), $\lambda_d(\cdot)$ is an r-dimensional and continuous function of the scalar random variable U_d, and the right-hand side of the equation is stricly increasing in U_d.[6] Following the literature, we refer to the heterogeneity term U_d as uniqueness.

We assume to observe $(Y, D, \boldsymbol{W}')$ for a sample of units, where Y is a scalar and continuous outcome and $\boldsymbol{W} = (W_1, \ldots, W_r)'$ is a vector of r continuous random variables. Factors are not observed and related to $\boldsymbol{W}$ through the following measurement equations:

$$\boldsymbol{W} = \boldsymbol{H} + \boldsymbol{V}, \tag{2}$$

where $\boldsymbol{V} = (V_1, \ldots, V_r)'$ is a vector of random variables. The heterogeneity terms (U_0, U_1) are assumed independent of $\boldsymbol{H}$ and $\boldsymbol{V}$, as we state more formally in Assumption 1. In Section 5, we consider Equation (1) and interpret $\boldsymbol{H}$ as a unidimensional factor representing neonatal health. The outcomes we examine are measurements of cognitive and health development at the age of 36 months. We use birth weight as $\boldsymbol{W}$ to proxy for the unobserved neonatal health, and $\boldsymbol{X}$ includes the mother demographics described in Table 1.[7] In some empirical settings, the classical measurement error assumption might be violated. For instance, if $\boldsymbol{W}$ serves as a prediction of $\boldsymbol{H}$, one might argue that the prediction error is uncorrelated with $\boldsymbol{H}$ rather than $\boldsymbol{W}$. In such cases, Berkson-type measurement errors would render equation (2) invalid (Hyslop & Imbens, 2001).

Assumption 1 (**Factor model and measurement error**). In the model defined by equations (1) and (2), for $d = \{0, 1\}$ the variables $(U_d, \boldsymbol{V}, \boldsymbol{H})$ have continuous distributions, with finite moments, and are mutually independent. Moreover, the variables U_d and $\boldsymbol{V}$ have mean zero.

This assumption is standard in the literature. The measurement errors $\mathbf{V}$ have zero mean, so that the location of $\mathbf{H}$ is identified from Equation (2). The classical properties of measurement error are maintained here to connect with models used in most empirical work. However, the extrapolation approach in Section 4 is valid in more general nonclassical models. For example, the variance of measurement error could be allowed to depend on $\boldsymbol{H}$ (Carroll et al., 2006) at the cost of introducing additional parameters whose values would need to be specified in the estimation strategy. Finally, the assumption on the location of U_0 and U_1 is a standard normalization.

We also maintain the assumption that treatment is randomly assigned. Assumption 2 is made for convenience, as identification stems from the factor structure and not from the assignment mechanism. Our analysis extends to regression discontinuity designs using a local randomization argument. Other variants to randomization may be considered to allow for nonrandom selection (such as ignorability conditional on $\boldsymbol{X}$).

Assumption 2 (**Treatment assignment**). The treatment D is randomly assigned.

Finally, the independence condition in Assumption 3 was used in past work to achieve identification of the gain distribution. Together with Assumption 1, it implies that the dependence between potential outcomes is driven by $\boldsymbol{H}$. Battistin and Chesher (2014) show that independence conditional on $\boldsymbol{H}$ does not imply independence conditional on $\boldsymbol{W}$. Thus, the identification results presented below are in general not valid after conditioning on $\boldsymbol{W}$, as shown in the simulations presented in Section S.6 of the supporting information.[8] In Section 5, we show how Assumption 3 can be relaxed to accommodate correlated uniqueness in the context of our empirical investigation.

Assumption 3 (**Independence**). The random variables (U_1, U_0) are independent.

3.2 | Parameter of interest

Equation (1) enables factors $\boldsymbol{H}$ to act as shifters for an individual's rank in the gain distribution and implies

$$\Delta = \left(\lambda_1(U_1)' - \lambda_0(U_0)'\right)\boldsymbol{H} + U_1 - U_0. \tag{3}$$

[6] Factor models have a longstanding tradition in empirical work, including the studies by Carneiro et al. (2003), Aakvik et al. (2005), Cunha et al. (2010), Chaparro and Sojourner (2019), and Attanasio et al. (2020). Factor models are also used to enhance external validity in regression discontinuity designs (Rokkanen, 2015) and in time-varying treatment effect models (Cooley Fruehwirth et al., 2016).

[7] Our conclusions are robust to the choice of control variables, as shown in Section S.8.1 of the supporting information.

[8] When estimands conditional on the latent variable $\boldsymbol{H}$ are of interest, conditioning on $\boldsymbol{W}$ alone will introduce biases. However, average and quantile treatment effect parameters conditional on $\boldsymbol{W}$ are interesting in their own right. Columns (6)–(8) and (10)–(12) in Table 2 report average causal effect estimates given by groups defined using birth weight, which is one of the proxies considered in Section 5.1.

Based on this equation, we define the following parameter: $\Delta(\tau; \boldsymbol{h}, u_0) := Q_{Y_1-Y_0}(\tau|\boldsymbol{H} = \boldsymbol{h}, U_0 = u_0)$, which represents the quantile function of the conditional distribution (QCD) of Δ given $\boldsymbol{H} = \boldsymbol{h}$ and $U_0 = u_0$.[9] Under the assumptions stated above, QCD is monotonically increasing in $\tau \in (0, 1)$ for all $(\boldsymbol{h}, u_0)$. Furthermore, using Equation (3), we have

$$\Delta(\tau; \boldsymbol{h}, u_0) = \left(\lambda_1(\tau)' - \lambda_0(u_0)'\right)\boldsymbol{h} + Q_{U_1}(\tau) - u_0, \tag{4}$$

where we write $\lambda_1(\tau) \equiv \lambda_1(Q_{U_1}(\tau))$ to simplify notation. Equation (4) defines a point-identifying functional conditional on $\boldsymbol{H} = \boldsymbol{h}$ and $U_0 = u_0$ and sets the stage for analog estimation of QCD from knowledge of the quantile functions $Q_{U_d}(\tau)$ and the factor loadings $\lambda_d(\cdot)$. The value u_0 in (4) affects gains from participation directly but also indirectly by attenuating or boosting the contribution of $\boldsymbol{H}$. For example, in Section 5, we find that better neonatal health implies gains in cognitive development, but these gains depend on the value of u_0 that a participant would have experienced without IHDP.

3.3 | Interpretation of the parameter

In relation to our specific application, the factor model (1) and the resulting expression for QCD in (4) can be motivated using an economic model of human capital accumulation during early stages of life. To demonstrate this, we consider the following dynamic equation (see, e.g., Attanasio et al., 2020) for $d = 0, 1$:

$$Y_d = \mu_1 Y_{d,-1} + \mu_2 I_d + \varepsilon_d = \lambda_d' \boldsymbol{H} + (\eta_d + \varepsilon_d). \tag{5}$$

Here, Y_d and $Y_{d,-1}$ represent the infant's outcomes at 36 months and at birth, respectively; I_d represents parental investment that occurs between birth and 36 months; and ε_d represents a zero-mean random shock to infant development in the first 3 years of life. Note that this shock and the inputs of the production function may vary with the random assignment to IHDP. The second equality arises from employing a CEF (conditional expectation function) decomposition, where the CEF of $\mu_1 Y_{d,-1} + \mu_2 I_d$ is $\lambda_d' \boldsymbol{H}$, η_d denotes the corresponding zero-mean projection error, and $\boldsymbol{H}$ represents determinants of the outcome at birth, $Y_{d,-1}$. The expression (5) suggests that the $r \times 1$ vector λ_d is a reduced-form parameter.

The terms within brackets, resulting from projection errors and shocks, may have variance that depends on $\boldsymbol{H}$. For example, heteroskedasticity can arise if health endowment at birth shapes the spread of subsequent shocks ε_d. Heteroskedasticity can also be attributed to errors in approximating the CEF of $\mu_1 Y_{d,-1} + \mu_2 I_d$. Participation in IHDP and initial conditions can influence future parenting styles or preferences, potentially leading to adjustments in the inputs I_d, as discussed in Chaparro and Sojourner (2015) and Chaparro et al. (2020). Consequently, Equation (5) would exhibit treatment-control differences in the loadings λ_d and the residuals η_d. The variability of these residuals may also depend on $\boldsymbol{H}$, across treatment arms, due to the mentioned adjustments.

To provide an example, assume that the errors ε_d and η_d are jointly normally distributed. In that case, their sum in (5) is also normally distributed, and we set its conditional standard deviation to be $1 + \boldsymbol{\gamma}_d' \boldsymbol{H}$, where $\boldsymbol{\gamma}_d$ is a $r \times 1$ vector of parameters. Substituting and rearranging terms, we obtain $Y_d = \left(\lambda_d' + \boldsymbol{\gamma}_d' U_d\right)\boldsymbol{H} + U_d$, where $U_d = (\eta_d + \varepsilon_d)/(1 + \boldsymbol{\gamma}_d' \boldsymbol{H})$ is a standard normal random variable conditional on $\boldsymbol{H}$. The resulting equation for the potential outcome Y_d represents a special case of the factor model in (1) under the following location-shift specification:

$$\lambda_d(U_d) = \lambda_d + \boldsymbol{\gamma}_d U_d. \tag{6}$$

The traditional factor model is obtained by setting $\boldsymbol{\gamma}_d$ to zero in the equation below. In this case, the factors $\boldsymbol{H}$ act solely as a location shifter in the distribution of Δ, as evident from (4).

Although the procedure developed in the next section does not rely on Equation (6), this simplified representation provides examples of potential channels that could account for heterogeneity in the distribution of gains Δ. It also clarifies that, in our specific application, the uniqueness U_d in (1) arises from a combination of shocks and projection errors stemming from the reduced-form representation of a model of human capital accumulation. This clarification allows for a better understanding of the assumption of independent uniqueness between potential outcomes for the same infant (Assumption 3): while potential outcomes may be correlated through the common factors $\boldsymbol{H}$, and covariates $\boldsymbol{X}$, any residual variability in these outcomes must be effectively random.

[9] For variable A having conditional distribution $F_{A|B=b}$, the τth quantile $Q_A(\tau|B = b)$ is defined as the smallest value a such that $F_{A|B}(a) = \tau$.

4 | ESTIMATION

4.1 | Approximation to zero error

Consider the quantile regression of Y_d on $\boldsymbol{H}$ in Equation (1), and let $\boldsymbol{\phi}_d(\tau) := \left(\lambda_d(\tau)', Q_{U_d}(\tau)\right)'$ be the vector of slopes and intercepts at quantile τ for participants ($D = 1$) and nonparticipants ($D = 0$). The theoretical case for considering separate regressions for the two groups when treatment effects are heterogeneous is discussed, for example, in Negi and Wooldridge (2021).

The parameter $\boldsymbol{\phi}_d(\tau)$ cannot be directly estimated from the data due to the measurement errors in $\boldsymbol{W}$, as shown in Equation (2). Let the diagonal matrix $\boldsymbol{\Sigma}$ be the variance of $\boldsymbol{V}$ in Equation (2) and, to fix ideas, assume that this variance is known or has been estimated. Following Cook and Stefanski (1994), we define an additional variable $\tilde{\boldsymbol{W}} = \boldsymbol{W} + \sqrt{\rho}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ is a $r \times 1$ vector of normal and independent errors with mean zero and covariance matrix $\boldsymbol{\Sigma}$ scaled by a known constant $\rho > 0$. The variance of $\tilde{\boldsymbol{W}}$ given $\boldsymbol{H}$ is $(1 + \rho)\boldsymbol{\Sigma}$, and this variance tends to zero as $\rho \to -1$. We denote by $\tilde{\boldsymbol{\phi}}_d(\tau, \rho)$ the vector of parameters at quantile τ retrieved from the quantile model of Y_d on $\tilde{\boldsymbol{W}}$.

The estimation of $\boldsymbol{\phi}_d(\tau)$ involves two steps, as shown in the algorithm detailed in Section S.5 and interpreted graphically in Figure S.1 of the supporting information. In the first step, we obtain via simulation estimates of $\tilde{\boldsymbol{\phi}}_d(\tau, \rho)$, which we denote by $\hat{\boldsymbol{\phi}}_d(\tau, \rho)$, at different ρ's. Specifically, starting with the raw measurements $\boldsymbol{W}$, we draw random values from a normal distribution with mean zero and variance $\rho\boldsymbol{\Sigma}$ to obtain $\tilde{\boldsymbol{W}}$, and estimate the quantile regression of Y_d on $\tilde{\boldsymbol{W}}$. The value of $\hat{\boldsymbol{\phi}}_d(\tau, \rho)$ is then obtained by simulating multiple draws, and averaging across simulations the slopes and intercepts obtained from quantile regressions. In our implementations of the approach, we use 50 draws and let ρ vary in the interval $(0, 2]$. In the second step, we plot $\hat{\boldsymbol{\phi}}_d(\tau, \rho)$ against ρ, fit a second-order polynomial to this plot, and extrapolate back to the "no error" case to obtain the following estimate of $\boldsymbol{\phi}_d(\tau)$:

$$\hat{\boldsymbol{\phi}}_d(\tau) = \hat{\boldsymbol{\phi}}_d(\tau, \rho = -1).$$

Such extrapolation at $\rho = -1$ relies on the parametric model fitted for explaining how $\hat{\boldsymbol{\phi}}_d(\tau, \rho)$ varies with ρ. The term $\tilde{\boldsymbol{\phi}}_d(\tau, \rho)$ can be interpreted as the counterfactual parameter that would be estimated had the measurement error variance been $(1 + \rho)\boldsymbol{\Sigma}$ instead of $\boldsymbol{\Sigma}$. The term $\boldsymbol{\epsilon}$ adds instrumental-like variability to estimate the bias arising from increasingly larger values of the measurement error variance.

4.2 | Properties of the approximation

The combination of these simulation and extrapolation steps is the core of Cook and Stefanski (1994) SIMEX algorithm. The method has been applied to nonlinear models and its asymptotic theory is derived in Carroll et al. (2006) for a general class of estimating equations. In particular, we show in Section S.1 of the supporting information that

$$\tilde{\boldsymbol{\phi}}_d(\tau, \rho) - \boldsymbol{\phi}_d(\tau) \approx -E\left(\left[\sigma(\tau, \boldsymbol{h})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}'\right]\right)^{-1} E\left(\sigma(\tau, \boldsymbol{h})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{V}}'\left(\lambda_d(\tau)'((1 + \rho)\boldsymbol{\Sigma})^{1/2}\right)\right), \tag{7}$$

where $\sigma(\tau, \boldsymbol{h})$ are quantile-specific weights, $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{W}}', 1)'$, $\tilde{\boldsymbol{V}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}$, and the term on the right-hand side of the approximation is zero if $\rho = -1$. This approximation shows that measurement error in our setting is likely to generate an attenuation bias in quantile regressions like in OLS, which is a point also noted by Angrist et al. ((2006), footnote 9).

Section S.6 of the supporting information reports a number of simulation studies to investigate the finite sample properties of our estimator, and to compare it with naive estimates that do not take measurement error into account. We show there that the SIMEX approximation of $\boldsymbol{\phi}_d(\tau)$ works well as long as the measurement error variance $\boldsymbol{\Sigma}$ is not too large (theoretical derivations in Cook and Stefanski, 1994, lead to the same conclusion), which is the case in IHDP data as we demonstrate in Section 5. We also show that using $\boldsymbol{W}$ instead of $\boldsymbol{H}$ (i.e., setting $\rho = 0$) may lead to severely biased conclusions, which is a fact echoing results in Battistin and Chesher (2014) and Chesher (2017) Finally, because of the relationship between $\tilde{\boldsymbol{\phi}}_d(\tau, \rho)$ and $\boldsymbol{\phi}_d(\tau)$ in (7), the literature has shown that a number of parametric models work well for the extrapolation step in most applications. This expectation is borne out in Section S.6, where we show that a quadratic model in ρ is a good extrapolant for the quantities we are interested in IHDP data.

The rationale for our approach is supported by a well-established statistical literature that, to date, has received little attention from economists. There are other alternatives available for estimating the parameter $\boldsymbol{\phi}_d(\tau)$, such as the

corrected-loss function estimator proposed by Wang et al. (2012). However, our approach offers several advantages over their estimator, particularly when the measurement error variance is small, and it does not require the use of numerical integration. We demonstrate this superiority through a simulation study, which is presented in Section S.6. Furthermore, the factor model considered by Chen et al. (2021) offers a broader scope than our own model. In their model, factors can vary by quantile and over time, and loadings are specific to each unit. They discuss an algorithm in their article that effectively locates the stationary points of the objective function used to retrieve these unknown quantities. However, their algorithm relies on large-N, large-T panel data. In contrast, our approach is designed to be applicable to cross-sectional data and accommodates loadings that can vary based on the treatment status of units. Moreover, the estimation procedure in Chen et al. (2021) initiates with latent factor estimates obtained from a principal component analysis and iteratively refines them through the algorithm. In contrast, we start with a proxy for latent factors and enhance the estimates by purging them from measurement error using the SIMEX method.

4.3 | Computation of QCD and inference

The last step of the algorithm described in Section S.5 consists of estimating the QCD parameter in (4) as follows:

$$\hat{\Delta}(\tau; \boldsymbol{h}, u_0) = \left(\widehat{\lambda}_1(\tau)' - \widehat{\lambda}_0(u_0)'\right)\boldsymbol{h} + \hat{Q}_{U_1}(\tau) - u_0, \tag{8}$$

where $\hat{Q}_{U_1}(\tau)$ is the SIMEX quantile intercept at τ from the sample of participants, $\widehat{\lambda}_1(\tau)$ is the SIMEX quantile slope at the value $\hat{Q}_{U_1}(\tau)$ from the sample of participants, and $\widehat{\lambda}_0(u_0)$ is the SIMEX quantile slope at the value u_0 from the sample of nonparticipants. We show $\hat{\Delta}(\tau; \boldsymbol{h}, u_0)$ at specific combinations of $(\boldsymbol{h}, u_0)$. Specifically, we select the value of u_0 by considering specific SIMEX quantile intercepts $\hat{Q}_{U_0}(\tau)$ from the sample of nonparticipants, for example setting τ to the median. Regarding the value of $\boldsymbol{h}$, note that Assumption 1 implies $E(\boldsymbol{H}|Y_0 = y_0) = E(\boldsymbol{W}|Y_0 = y_0)$. Thus, in our application we set $\boldsymbol{h} = \hat{E}(\boldsymbol{H}|Y_0 = y_0)$, where this value is the prediction from a regression of $\boldsymbol{W}$ on Y estimated for nonparticipants.[10]

To obtain standard errors, we employ a pair bootstrap strategy as discussed in Carroll et al. (1996) and Koenker (2005). More precisely, we sample participants and nonparticipants with replacement, and re-estimate the various quantities entering QCD for each bootstrap pseudo-sample. Confidence intervals in Section 5 are obtained from bootstrap distributions of QCD estimates.

5 | QCDS OF LOW BIRTH WEIGHT AND PREMATURE INFANTS

5.1 | Measurement error variance

For IHDP data, we assume that the potential outcomes in Equation (1) are correlated due to their dependence on a common factor ($r = 1$), which represents *neonatal health*. To capture this factor, we consider four anthropometric measures taken at birth as possible candidates for the proxy variable $\boldsymbol{W}$ in Equation (2). These measures are weight, length, head circumference, and gestational age, for which descriptive statistics are reported in Table 1. The first principal component derived from these four measures explains approximately 82.2% of their correlation matrix, bolstering the idea that they can be considered as different manifestations of the same underlying dimension. We also find in the data that the average value of the first principal component for HLBW infants is approximately 1.4 standard deviations larger than that of the LLBW group. Furthermore, the first principal component exhibits substantial associations with all anthropometric measures, indicating that it can be interpreted as a proxy for health endowment at birth. Over-identification tests, reported in this section and in Table S.1 of the supporting information, provide further support for the notion that specifications involving a single latent factor are consistent with the data.

We have assumed thus far a known measurement error variance $\boldsymbol{\Sigma}$ for the variable $\boldsymbol{W}$ in Equation (2). We now present an approach to estimate the measurement error variance in each of the four anthropometric measures. We additionally provide a test for the hypothesis that the correlation among these four measures is spanned by their dependence on one

[10] When the measurement error variance is small, the quantile function of the latent factors, $Q_{\boldsymbol{H}}(\tau)$, could alternatively be computed using the approximation in Chesher (2017) as we show in Section S.2 of the supporting information. Therefore, one could alternatively select a value of $\boldsymbol{h}$ by considering quantiles of the latent factors distribution.

TABLE 3 Measurement error variances.

	Gestational Age (weeks)	Length	Weight	Head circumference	*p*-value
	(1)	(2)	(3)	(4)	(5)
Estimate	0.373***	0.244***	0.067*	0.230***	0.207
	(0.043)	(0.048)	(0.039)	(0.050)	

Note: Columns (1) to (4) in this table present estimates and standard errors of the measurement error variances $E\left(V_j^2\right)$ for the four anthropometric measures of latent health at birth: weight, length, head circumference, and gestational age, denoted by $j = 1, \dots, 4$. These measures are obtained as residuals from separate regressions on mother demographics (refer to Panel B of Table 1). The residuals are standardized to have unit variance. The estimation is performed using the residuals of the measurement equations (9), and by considering the system of equations (10) as implied by Assumption 1. Further details can be found in Section S.3 of the supporting information. The *p*-value in column (5) is for the hypothesis that the assumptions underlying the system (10) explain the correlation structure among the four anthropometric measures.*$p < 0.1$.**$p < 0.05$.***$p < 0.01$.

common factor H. Our objectives are threefold: to provide supporting evidence for the existence of the scalar variable H; to identify the anthropometric measure W in (2) that exhibits the lowest measurement error when used as a proxy for H; to initiate the SIMEX method in Section 4.1 using the estimated measurement error variance for this chosen proxy. The approach in Carroll et al. (1996) accommodates unknown variance, and the attractive features of SIMEX are maintained if a consistent estimate of this variance is available.

With a slight abuse of notation, we consider the following system of measurement equations that generalizes Equation (2):

$$\mathbb{W} = \boldsymbol{a}H + \mathbb{V}, \tag{9}$$

where $\mathbb{W} = (W_1, W_2, W_3, W_4)'$ is a 4×1 vector representing the anthropometric measures, $\mathbb{V} = (V_1, V_2, V_3, V_4)'$ is a 4×1 vector representing the corresponding measurement errors, H is a scalar representing the common factor, and $\boldsymbol{a} = (1, a_2, a_3, a_4)'$ is a 4×1 vector of unknown parameters. This setup follows closely that in Section 3, with the exception that we now have more measures $\mathbb{W}$ (four) than latent factors (one). Our approach, outlined below, leverages this degree of over-identification to estimate the variance of the measurement errors V_j for $j = 1, \dots, 4$. In particular, the scale of H in (9) is normalized to match the scale of the first anthropometric measure. Thus, the anthropometric measure corresponding to the first equation in (9) can be interpreted as the measure used in Equation (2) when $r = 1$.

Equation (9) yields the following moment conditions under Assumption 1:

$$E\left(\mathbb{W}\mathbb{W}'\right) = \boldsymbol{a}\boldsymbol{a}'E\left(H^2\right) + E\left(\mathbb{V}\mathbb{V}'\right), \tag{10}$$

which depend on the unknown scalar $E\left(H^2\right)$, the vector $\boldsymbol{a}$, and the diagonal matrix $E\left(\mathbb{V}\mathbb{V}'\right)$ whose elements are the measurement error variances. We estimate these variances after netting out the observables $\boldsymbol{X}$. Specifically, we compute residuals from regressions of each anthropometric measure on mother demographics (see Panel B of Table 1), separately for treatment and control groups, and standardize residuals to have zero average and unit variance in the sample. These standardized residuals are then used to define $\mathbb{W}$, and the unknown parameters are estimated using the nonlinear system of equations in (10).[11]

We select birth weight as the proxy variable in (2) for the latent factor H due to its lower measurement error compared with other anthropometric measures, as demonstrated in Table 3. In the following sections, we initiate the SIMEX procedure using the value reported in column (3). We show in Section S.3 of the supporting information that the unknowns on the right-hand side of (10) are *over-identified*, meaning that there are more usable empirical moments on the left-hand side than parameters to be estimated. The *p*-value presented in column (5) of Table 3 is obtained from a test assessing the validity of the restrictions imposed in the system (9). We also show in the supporting information that additional equations can be added to (10) by considering the relationship with potential outcomes as implied by Assumption 1 and the assumption

[11] As the residualized anthropometric measures have a zero average in the sample, the equations $E(\mathbb{W}) = \boldsymbol{a}E(H)$ do not provide additional information about the unknowns on the right-hand side. Furthermore, given that the residuals have a unit variance, the equations in (10) imply additional constraints, as explained in Section S.3 of the supporting information, and yield estimates of the "noise" (V_j) to "signal" (W_j) ratio in Equation (9), for $j = 1, \dots, 4$.

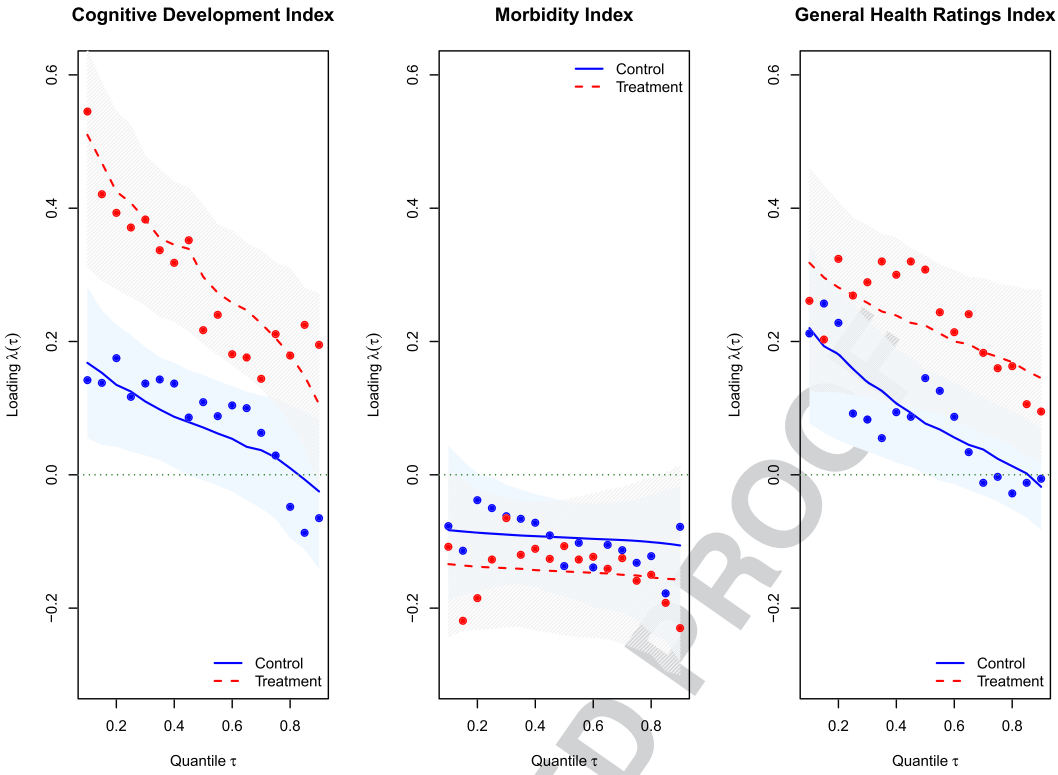

FIGURE 2 Quantile loadings for IHDP participants and nonparticipants. *Note*: Dots in the figure are SIMEX estimates of quantile loadings $\lambda_1(\tau)$ (treatment) and $\lambda_0(\tau)$ (control) for the model in Equation (1). Panels refer to the three outcomes analyzed: the cognitive development index, the morbidity index, and general health ratings index. Definitions for these outcomes are in Table 2. Solid lines in the figure are estimates of quantile loadings obtained under the location-shift specification (6). Shaded areas denote 95% confidence intervals obtained from 200 bootstrap replications.

of a single latent factor. These equations introduce additional degrees of over-identification, which we employ to conduct tests. The *p*-values for these over-identification tests are reported in Table S.1 in the supporting information. In all cases, the *p*-values of the over-identification tests deviate considerably from conventional significance levels, suggesting that the assumptions made are not rejected in IHDP data.[12]

5.2 | SIMEX loadings

Quantile loadings and intercepts are obtained as described in Section 4.1. We consider 50 draws from the distribution of ϵ; the extrapolation step uses a quadratic polynomial over 20 equally spaced values for ρ in the interval $[0.1, 2]$, as in Figure S.1 of the supporting information.

The dots in Figure 2 represent the SIMEX quantile estimates $\hat{\lambda}_1(\tau)$ and $\hat{\lambda}_0(\tau)$ for the slopes in Equation (1). Each panel in the figure corresponds to one of the three outcomes analyzed: cognitive development (Y_d^{CD}), morbidity (Y_d^{MB}), and general health (Y_d^{GH}).[13] While the estimated loadings do not vary much by quantile in the case of morbidity, we observe a sharp decline as we go across τ in the case of cognitive development. Figure 3 displays SIMEX estimates $\hat{Q}_{U_1}(\tau)$ and $\hat{Q}_{U_0}(\tau)$ for the uniqueness terms in Equation (1). We do not find significant differences in the uniqueness distributions of IHDP participants and nonparticipants. Moreover, in Figure 2, we present the estimated slopes that would be obtained under the location-shift specification in Section 3.3—see Equation (6). Specifically, we perform regressions of the SIMEX

[12]Note that independence between V and H aligns with studies that interpret birth weight (W) as a combination of genetic components of the infant (H) and external factors affecting fetal development (V), like oxygen availability. For example, following Rosenzweig and Schultz (1982), V can be seen as biologically exogenous variability, supporting Assumption 1.

[13]In what follows, outcomes are standardized to have zero mean and unit variance in the sample after taking residuals from regressions on mother demographics.

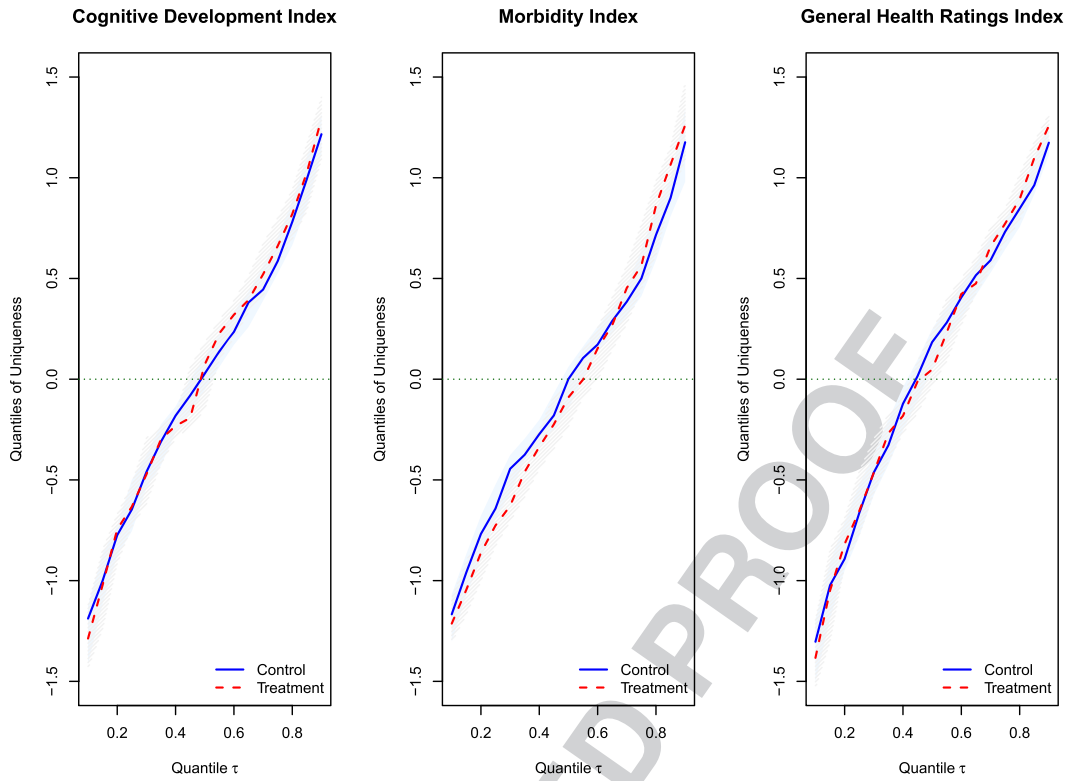

FIGURE 3 Quantile functions of uniqueness for IHDP participants and nonparticipants. *Note*: The figure shows SIMEX estimates of $Q_{U_1}(\tau)$ (treatment) and $Q_{U_0}(\tau)$ (control) for the model in Equation (1). Panels refer to the three outcomes analyzed: the cognitive development index, the morbidity index, and general health ratings index. Definitions for these outcomes are in Table 2. Shaded areas denote 95% confidence intervals obtained from 200 bootstrap replications.

slope estimates $\hat{\lambda}_d(\tau)$ on the SIMEX estimates $\hat{Q}_{U_d}(\tau)$ separately for $d = \{0, 1\}$, and show with solid lines in Figure 2 the predictions derived from these regressions.[14]

We find that the quantile-specific slope and intercept associated with Equation (1) are well approximated by the location-scale shift model. The estimates of the factor loading, corresponding to the intercept of (6), are presented in columns (1) and (2) of Table 4. The estimated slopes in (6) are reported in columns (4) and (5) along with their difference in the last column. Sharper treatment-control differences in slopes emerge for cognitive development.

The results in Table 4 imply that IHDP affects both the location and scale of outcome distributions depending on a child's health at birth. For example, columns (2) and (5) imply that, without IHDP, as one moves across the birth endowment distribution, outcomes become more concentrated around better values. This conclusion can be seen by considering specification (6), upon noting that the scale parameters affect the outcome variance. Treatment-control differences in column (3) of Table 4 imply positive treatment effects on outcome averages for all children. However, IHDP participation has different effects on the variance of the distribution depending on the outcome, as shown in column (6).

5.3 | Distribution of gains

We next estimate the distribution of gains for LLBW and HLBW infants using Equation (8). To exemplify, we compute the QCD at the value $u_0 = Q_{U_0}(0.5) \approx 0$ obtained from SIMEX estimates of the quantiles of U_0 in Figure 3. We consider two values of the latent factor, $h \in \{h_L, h_H\}$. Specifically, h_L is a prediction from a linear regression of birth weight on the outcome in the sample of LLBW nonparticipants (as detailed in Section 4.3). Separate regressions are considered for the three outcomes, and predictions are computed at the outcome average. The value h_H is constructed in a similar manner using HLBW nonparticipants.

[14]To obtain the location and scale parameters in model (6), we estimate $\hat{\lambda}_d(\tau) = \alpha_0 + \alpha_1 \hat{Q}_{U_d}(\tau) + \nu$ by treatment status. The slope coefficient from this regression yields an estimate of γ_d, while the intercept estimate λ_d.

TABLE 4 Location and scale parameters from the location-shift specification.

	Location parameters			Scale parameters		
	Treatment	Control	Difference	Treatment	Control	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Cognitive development index					
Estimate	0.304***	0.069***	0.235***	−0.140***	−0.082***	−0.058***
	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)
	Panel B: Morbidity index					
Estimate	−0.140***	−0.103***	−0.037***	−0.005	−0.027***	0.022***
	(0.004)	(0.003)	(0.006)	(0.004)	(0.004)	(0.005)
	Panel C: General health ratings index					
Estimate	0.224***	0.092***	0.131***	−0.059***	−0.088***	0.028***
	(0.005)	(0.003)	(0.006)	(0.003)	(0.002)	(0.004)

Note: The table shows estimates for the factor loadings obtained under the location-shift specification (6). Columns (1) and (2) show the estimated intercepts (*loadings*) in specification (6) for $d = \{0, 1\}$, while column (3) shows the differences between columns (1) and (2). Columns (4) and (5) show the estimated slopes (*scale parameters*) in specification (6) for $d = \{0, 1\}$, while column (6) shows the differences between columns (4) and (5).*$p < 0.1$.**$p < 0.05$.***$p < 0.01$.

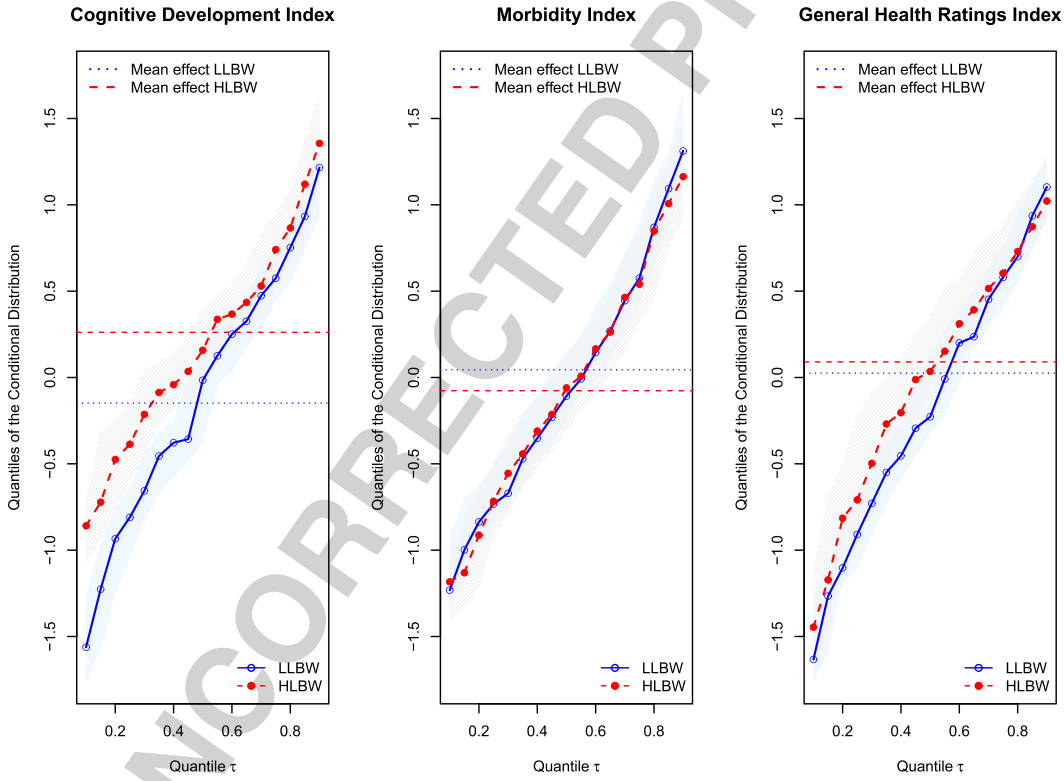

FIGURE 4 Quantiles of the conditional distribution of treatment effects. *Note*: The figure shows estimates of QCD at $u_0 = Q_{U_0}(0.50)$ and $h \in \{h_L, h_H\}$ for LLBW and HLBW infants. The horizontal lines denote estimates of the average effects for LLBW and HLBW groups. Shaded areas denote 95% confidence intervals obtained from 200 bootstrap replications.

We start with IQ gains, as this dimension shows consistently large average effects in many early childhood interventions similar to IHDP. An improvement in neonatal health H influences the location and scale of QCDs. This pattern can be seen from the first panel of Figure 4, where the dashed and continuous lines represent the estimated QCDs, $\Delta\left(\tau; h_H, Q_{U_0}(0.5)\right)$ and $\Delta\left(\tau; h_L, Q_{U_0}(0.5)\right)$, for HLBW and LLBW infants, respectively. The horizontal lines indicate the average effects for these two groups, and the shaded areas represent the 95% point-wise bootstrap confidence intervals. The figure illustrates that, when conditioning on a post-natal shock U_0 value approximately equal to zero, infants with better neonatal health exhibit a gain distribution with the lower tail shifted toward larger values. Differences between QCDs tend to disappear as one moves across quantiles (i.e., 0.70σ at the 0.10 quantile and 0.14σ at the 0.90 quantile). This pattern indicates that

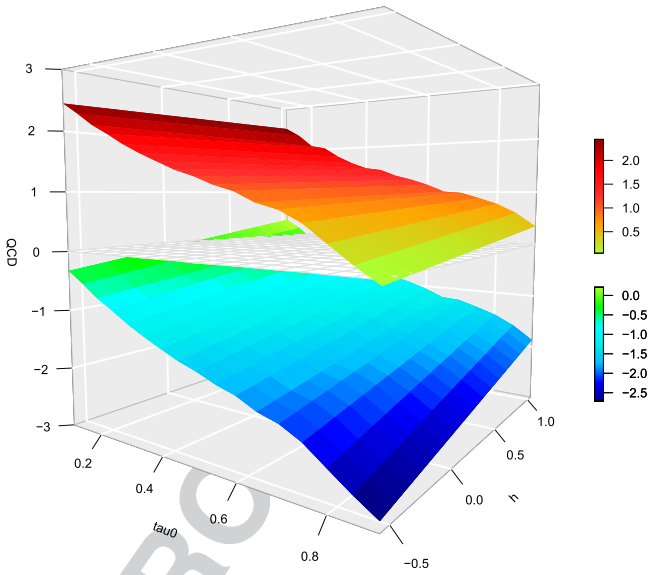

FIGURE 5 Conditional distribution of treatment effects at selected quantiles. *Note*: The figure shows how $\Delta(\tau; h, u_0)$ varies with (h, u_0) at the 0.10 quantile (lower surface) and 0.90 quantile (upper surface) of cognitive development (Stanford–Binet intelligence scale).

the gains from the intervention are more widely distributed among infants with worse health endowment at birth. For example, the interquartile range for LLBW infants is approximately $0.6 - (-0.8) = 1.4\sigma$ compared with a value of $0.7 - (-0.4) = 1.1\sigma$ for HLBW infants. Furthermore, the first panel implies positive gains for about 62% of HLBW participants, whereas this proportion is much lower for LLBW participants, at about 50%.

The distribution of effects in Figure 4 points to negative returns to participation for some children, and the incidence of such negative returns tends to be larger among those participants with comparatively lower health endowments at birth. A closer look at Equation (4) reveals that such negative returns are likely driven by a draw from the U_1 distribution that is below the counterfactual median draw that the same child would have experienced without participating in IHDP (the median of U_0 is approximately zero in our sample). Ramey et al. (1992) highlight significant variation across IHDP sites in terms of how the program was implemented and the extent of participation by children and families. Table 3 in their article provides insights into these variations, indicating that, on average, children attended the development centers for 267 days, but this duration ranged from 225 days to 300 days across different sites. Notably, the standard deviation in the number of days spent at the development centers is approximately 50% to 70% of the mean. This statistic suggests that some children in the treatment group had relatively limited exposure to IHDP. For these children, the value of U_1 is likely lower than the median of U_0 used in the computation of QCD, particularly with low health endowments, and this difference can explain the negative effects documented in Figure 4.

For health-related outcomes, improved neonatal health leads to a shift toward larger values of QCDs, as demonstrated in the remaining panels of Figure 4, where lower values of the morbidity index indicate more favorable outcomes. However, when examining health during the first 3 years of life, the disparities between LLBW and HLBW QCDs are smaller compared with IQ, and they appear to change gradually across quantiles. For example, the difference between curves in the third panel is about 0.19σ at the 0.10 quantile, and 0.08σ at the 0.90 quantile.

F5 The empirical relevance of nonseparability between neonatal health and post-natal shocks is demonstrated in Figure 5, which shows how $\Delta(\tau; h, u_0)$ varies with (h, u_0) at the 0.10 and 0.90 quantiles of cognitive development.[15] The positive gradient in neonatal health suggests that better endowment at birth H leads to an upward shift in the distribution of treatment effects. However, this shift conceals significant heterogeneity based on the intensity of post-natal shocks U_0. To illustrate this point, consider infants in the figure with the same h but different baseline outcomes because of significantly positive (high τ_0) and negative (low τ_0) values u_0. IHDP proves to be particularly effective for infants in the latter group, as they have a higher probability of experiencing positive returns from participation. However, the two surfaces are not parallel. Therefore, IHDP redistributed larger gains toward children who would have encountered relatively worse post-natal conditions in the absence of the intervention, but it had a less pronounced effect on children with a poor endowment at birth.

[15] We consider 17 equally-spaced values for u_0 (using Figure 3) and the 1 to 99 percent range for predictions $\hat{E}(W|Y_0 = y_0)$ of h. Values of τ_0 in the figure refer to quantiles in the U_0 distribution.

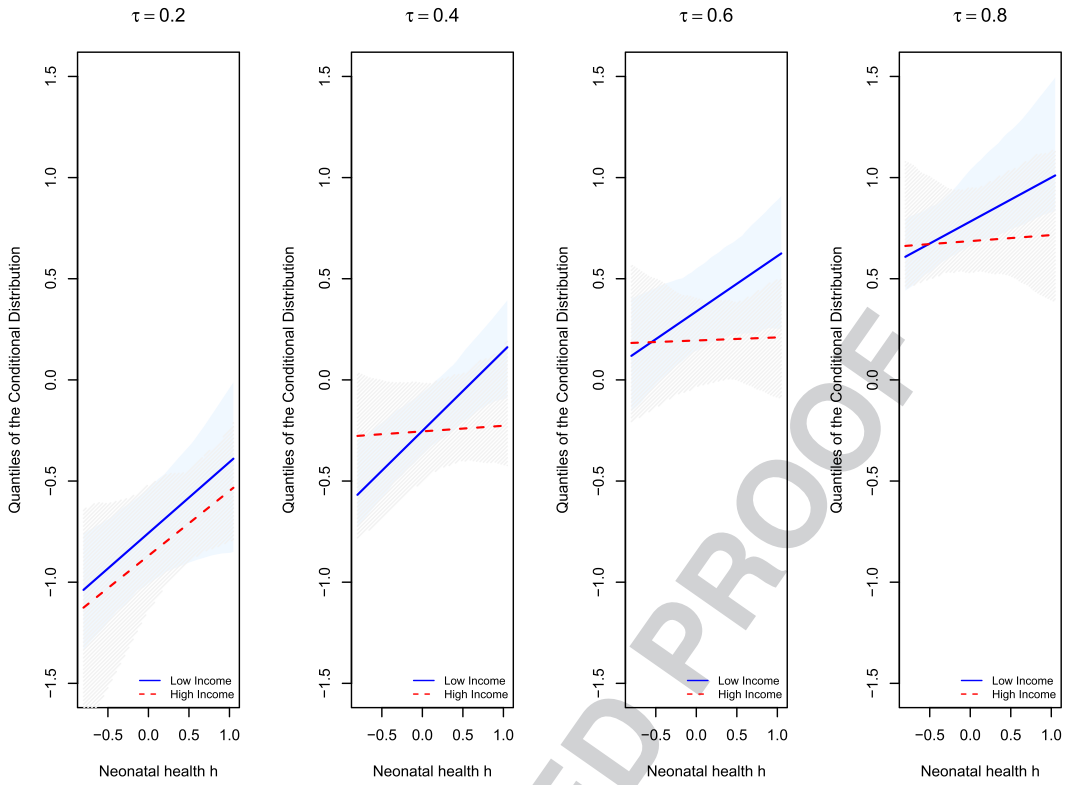

FIGURE 6 Conditional quantiles of cognitive development by neonatal health and family income. *Note*: The figure shows how estimates of QCD at $u_0 = Q_{U_0}(0.50)$ and selected quantiles (in different panels) vary with neonatal health H. Shaded areas denote 90% confidence intervals obtained from 200 bootstrap replications.

5.4 | Interactions with socio-economic status

Are the positive effects of IHDP explained by differences in the socio-economic status of participants? We address this question by defining two *groups* of children in families with income below (low-income) and above (high-income) the IHDP median, and estimating QCDs separately for each group.[16] The four panels in Figure 6 show how estimates of $\Delta\left(\tau; h, Q_{U_0}(0.5)\right)$ at selected quantiles $\tau \in \{0.20, 0.40, 0.60, 0.80\}$ vary with neonatal health h. Only IQ gains are considered, and values for h on the horizontal axis are the 1 to 99 percent range for predictions of latent health $\hat{E}(W|Y_0 = y_0)$ from a regression of birth weight on IQ for nonparticipants. The analysis reveals that treatment effects exhibit an upward trend with h across all quantiles for both income groups. Specifically, the effects at the 0.60 and 0.80 quantiles consistently demonstrate positive outcomes throughout the health spectrum. Moreover, the effects are generally more pronounced for children from lower income backgrounds, particularly at the higher end of the health distribution. However, the precision of our analysis is somewhat affected due to the limited sample size.

Income group differences in treatment effects are comparable in magnitude to within-income group variability. Past research indicates that, on average, low-income families tend to observe larger effects for HLBW children compared with high-income families (e.g., see Duncan & Sojourner, 2013). In our analysis, an OLS regression for HLBW children's IQ based on the IHDP treatment assignment dummy, a lower-income group dummy, and their interaction reveals a coefficient of approximately 0.26σ for the latter variable (results not shown for brevity). In Figure 6, the $0.80 - 0.20$ quantile range for $H = 0.5$ is approximately $0.70 - (-0.71) = 1.41\sigma$ for high-income families and $0.89 - (-0.58) = 1.47\sigma$ for low-income families. These figures suggest that variability in IHDP effects within groups is substantially greater than between groups.

[16] Approximately 10% of observations that lack income information are excluded from the analysis. Duncan and Sojourner (2013) employ an alternative approach to define income groups, which involves multiple imputation of missing values. They estimate average differences based on income by employing weights derived from the Early Childhood Longitudinal Study - Birth Cohort data.

Our point estimates suggest that, while lower-income children generally benefit more from IHDP on average, around 20% of higher-income children achieve gains exceeding the median of their lower-income counterparts. This is evident at $H = 0.5$ in the fourth panel of Figure 6, where the effect for the higher-income group is about 0.6, and the median gain for the lower-income group at $H = 0.5$ is below 0.5, as indicated by the 0.60 quantile in the third panel. Our findings also suggest that the higher average effects observed in the healthiest children within the lower-income group are primarily due to significant gains in a minority of cases. The differences in Figure 6 panels become more pronounced for larger H values, typically between the 0.60 and 0.80 quantiles. Meanwhile, income disparities among participants are less relevant for children with the poorest health conditions, likely LLBW children. Considering IHDP's target population, infants at the higher end of Figure 6 panels resemble normal birth weight infants. This group has well-documented causal links between family income and health and development (e.g., Ko et al., 2020).

5.5 | Robustness of results to correlated uniqueness

We extend our empirical investigation to the case of dependence between the post-natal potential shocks, or uniqueness terms, (U_0, U_1). Positive dependence may arise, for example, due to unobserved factors specific to each infant that are not accounted for by health status at birth or the variables $\boldsymbol{X}$. In what follows, we employ the location-shift specification (6), as we have shown to fit the IHDP data well. Under this specification, the QCD has the following expression:

$$\Delta(\tau; h, u_0) = \left(\left(\lambda_1 + \gamma_1 Q_{U_1|U_0}(\tau|u_0)\right) - (\lambda_0 + \gamma_0 u_0)\right) h + Q_{U_1|U_0}(\tau|u_0) - u_0, \tag{11}$$

where $Q_{U_1|U_0}(\tau|u_0)$ represents the τth conditional quantile of U_1 given $U_0 = u_0$. Note that, even with correlated uniqueness, the location and scale parameters in (11) can be estimated as explained in Section 5.2 and shown in Table 4.

It follows that the computation of (11) rests on the particular expression of $Q_{U_1|U_0}(\tau|u_0)$, which depends on the joint distribution of the uniqueness terms. For example, under Assumption 3, this conditional quantile becomes $Q_{U_1}(\tau)$. We derive in Section S.9 of the supporting information the expression for conditional quantiles considering examples of families of semi-parametric copulas for (U_0, U_1). In particular, we show there that bivariate normality with correlation coefficient $\zeta \in (0, 1)$ implies

$$Q_{U_1|U_0}\left(\tau|Q_{U_0}(\tau_0)\right) = \sqrt{(1-\zeta^2)}\, Q_{U_1}(\tau) + \zeta Q_{U_1}(\tau_0). \tag{12}$$

Thus, without Assumption 3, specific assumptions about the form of the conditional quantile $Q_{U_1|U_0}(\tau|u_0)$ pave the way for the computation of (11). Joint normality and the location-scale shift model (6) are maintained here to flesh out the sensitivity of our results to dependence between the uniqueness.[17] In this case, QCD can be obtained via (12) at given values of the coefficient ζ. We do not use restrictions to point-identify ζ, although we discuss briefly examples of such restrictions in the supporting information. For the sake of brevity, our discussion below focuses on cognitive development, as it is the outcome with the most substantial effects in our analysis.

The difference between lines in Figure 4 is attenuated when $\zeta \neq 0$, but it still decreases as one moves to upper quantiles. This can be seen by noticing that (11) implies $\frac{\partial}{\partial h}\Delta\left(\tau; h, Q_{U_0}(0.5)\right) \approx (\lambda_1 - \lambda_0) + \gamma_1\sqrt{1-\zeta^2}Q_{U_1}(\tau)$, which is decreasing in τ because $\gamma_1 < 0$ (see Table 4) and $Q_{U_0}(0.5) \approx 0$. The takeaways from Figure 5 are qualitatively similar when ζ is not larger than 0.60. This figure demonstrates the relevance of taking into account the nonseparability between neonatal health H and post-natal shocks U_0. When $\zeta = 0$, nonseparability depends on the value of $-\gamma_0$, which is 0.082 in our data and obtained as second derivative of QCD with respect to H and U_0. Nonseparability implies that the effect of improvements in health H on QCD increases in U_0 for all τ's. When $\zeta \neq 0$ this conclusion is maintained if $-\gamma_0 + \gamma_1\zeta > 0$, which represents the second derivative of QCD in (11) with respect to H and U_0 if one assumes (12) and that the uniqueness have the same variance, which is approximately true in IHDP data. Taking at face value the parameter estimates in Table 4, the expression $-\gamma_0 + \gamma_1\zeta > 0$ is valid if $\zeta < 0.60$.[18] Finally, for the low-income group in Figure 6, lines are approximately parallel implying that $\gamma_1 \approx 0$. In this case, the slope of QCD is unaffected by ζ. A similar reasoning implies that $\gamma_1 < 0$ is approximately true for the high-income group. In this case, increasing values of ζ will attenuate the slope of lines in Figure 6.

[17] Figure S.6 in the supporting information suggests that joint normality of the uniqueness is approximately valid for IHDP.

[18] Figures S.7 and S.8 of the supporting information are the analog of Figure 5 considering three scenarios for mild ($\zeta = 0.1$), moderate ($\zeta = 0.2$) and high ($\zeta = 0.4$) dependence between uniqueness.

6 | CONCLUSION

We defined a novel distributional parameter, the QCD of treatment effects, to study the effects of the IHDP program. Identification of this parameter hinges on specifying a quantile factor model for potential outcomes and having proxies for the latent factors. We estimated QCD using a simulation-extrapolation (SIMEX) method, which accounts for the bias introduced by using proxies instead of the unobserved factors. Beyond its methodological contributions, our research shows that the effects of IHDP on cognitive and health outcomes vary across participants and depend on health conditions at birth.

ACKNOWLEDGEMENTS

The manuscript benefited from invaluable comments and suggestions by the co-editor, Frank Vella, and three anonymous referees. We are also grateful to Aaron Sojourner, Chris Bollinger, Matt Harding, Roger Koenker, Jim Ziliak, and seminar participants at the University of Illinois at Urbana-Champaign, Queen Mary University of London, University of Kentucky, University of St. Gallen, the 14th IZA/CEPR European Summer Symposium, the 2020 SEA Meeting, and the 2020 World Congress of the Econometric Society. Financial support from Fondazione CARIPARO (Progetti di Eccellenza) is gratefully acknowledged.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://doi.org/10.15456/jae.2024122.1221495693.

REFERENCES

Aakvik, A., Heckman, J. J., & Vytlacil, E. J. (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to norwegian vocational rehabilitation programs. *Journal of Econometrics*, *125*, 15–51.

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, *97*(457), 284–292.

Abbring, J. H., & Heckman, J. J. (2007). Econometric evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation, *Handbook of econometrics, vol6, part b*: North-Holland, Amsterdam, pp. 5145–5303.

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

Angrist, J., Chernozhukov, V., & Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, *74*(2), 539–563.

Arnold, D., Dobbie, W., & Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, *112*(9), 2992–3038.

Ashworth, J., Hotz, V. J., Maurel, A., & Ransom, T. (2021). Changes across cohorts in wage returns to schooling and early work experiences. *Journal of Labor Economics*, *39*(4), 931–964.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., & Rubio-Codina, M. (2020). Estimating the production function for human capital: Results from a randomized control trial in colombia. *American Economic Review*, *110*(1), 48–85.

Battistin, E., & Chesher, A. (2014). Treatment effect estimation with covariate measurement error. *Journal of Econometrics*, *178*(2), 707–715.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, *96*, 988–1012.

Bonhomme, S., & Robin, J.-M. (2010). Generalized parametric deconvolution with an application to earnings dynamics. *The Review of Economic Studies*, *77*, 491–533.

Brooks-Gunn, J., Gross, R. T., Kraemer, H. C., Spiker, D., & Shapiro, S. (1992). Enhancing the cognitive outcomes of low birth weight, premature infants: For whom is the intervention most effective? *Pediatrics*, *89*(6), 1209–1215.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2019). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192–210.

Carneiro, P. M., Karsten, K. T., & Heckman, J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. *International Economic Review*, *44*(2), 361–422.

Carroll, R. J., Küchenhoff, H., Lombard, F., & Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *Journal of the American Statistical Association*, *91*(433), 242–250.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. (2006). *Measurement error in nonlinear models: A modern perspective* (Second).: CRC Press.

Chaparro, J., & Sojourner, A. (2015). Same program, different outcomes: Understanding differential effects from access to free, high-quality early care. (*9552*): Institute of Labor Economics (IZA).

Chaparro, J., & Sojourner, A. J. (2019). Differential effects from access to high-quality early care. In Reynolds, A. J., & Temple, J. A. (Eds.), *Sustaining early childhood learning gains*. Cambridge.

Chaparro, J., Sojourner, A., & Wiswall, M. J. (2020). Early childhood care and cognitive development. (*26813*), Working Paper Series: National Bureau of Economic Research.
Chen, L., Dolado, J. J., & Gonzalo, J. (2021). Quantile factor models. *Econometrica*, *89*(2), 875–910.
Chesher, A. (2017). Understanding the effect of measurement error on quantile regressions. *Journal of Econometrics*, *200*(2), 223–237.
Conti, G., Hanson, M., Inskip, H. M., Crozier, S., Cooper, C., & Godfrey, K. (2018). Beyond birth weight: the origins of human capital. IFS Working Paper W18/30.
Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328.
Cooley Fruehwirth, J., Navarro, S., & Takahashi, Y. (2016). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics*, *34*(4), 979–1021.
Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skills formation. *Econometrica*, *78*(3), 883–931.
Currie, J., & Rossin-Slater, M. (2015). Early-life origins of life-cycle well-being: Research and policy implications. *Journal of Policy Analysis and Management*, *34*(1), 208–242.
Dong, Y., & Shen, S. (2018). Testing for rank invariance or similarity in program evaluation. *The Review of Economics and Statistics*, *100*(1), 78–85.
Duncan, G. J., & Sojourner, A. J. (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *The Journal of Human Resources*, *48*(4), 945–968.
Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2016). Early childhood education. In Moffitt, R. (Ed.), *Economics of means-tested transfer programs in the united states*, Vol. *2*: University of Chicago Press, pp. 235–297.
Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276.
Frandsen, B. R., & Lefgren, L. J. (2018). Testing rank similarity. *The Review of Economics and Statistics*, *100*(1), 86–91.
Frölich, M., & Melly, B. (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics*, *31*(3), 346–357.
Gross, R. T., Spiker, D., & Haynes, C. W. (1997). *Helping low birth weight, premature babies: The infant health and development program*. Stanford University Press.
Heckman, J. J. (2020). Randomization and social policy evaluation revisited. (*12882*), Working Paper Series: IZA.
Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the highscope perry preschool program. *Quantitative Economics*, *1*(1), 1–46.
Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies*, *64*, 487–535.
Hyslop, D. R., & Imbens, G. W. (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics*, *19*(4), 475–481.
IHDP (1990). Enhancing the outcomes of low-birth-weight, premature infants: A multisite, randomized trial. *JAMA*, *263*(22), 3035–3042.
Ko, H., Howland, R. E., & Glied, S. A. (2020). The effects of income on children's health: Evidence from supplemental security income eligibility under new york state medicaid. (*26639*), Working Paper Series: National Bureau of Economic Research.
Koenker, R. (2005). *Quantile regression*. New York, New York: Cambridge University Press.
Negi, A., & Wooldridge, J. M. (2021). Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, *40*(5), 504–534.
Powell, D. (2020). Quantile treatment effects in the presence of covariates. *The Review of Economics and Statistics*, *102*(5), 994–1005.
Ramey, C. T., Bryant, D. M., Wasik, B. H., Sparling, J. J., Fendt, K. H., & La Vange, L. M. (1992). Infant health and development program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics*, *89*(3), 454–465.
Rokkanen, M. (2015). Exam schools, ability, and the effects of affirmative action: Latent factor extrapolation in the regression discontinuity design. Discussion Paper 1415-03, Department of Economics, Columbia University.
Rosenzweig, M., & Schultz, T. (1982). The behavior of mothers as inputs to child health: The determinants of birth weight, gestation, and rate of fetal growth, *Economic aspects of health*: National Bureau of Economic Research, Inc, pp. 53–92.
Wang, H. J., Stefanski, L. A., & Zhu, Z. (2012). Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika*, *99*(2), 405–421.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of the article.

How to cite this article: Battistin, E., Lamarche, C., Rettore, E. (2024). Quantiles of the gain distribution of an early childhood intervention. *Journal of Applied Econometrics*, 1–19. https://doi.org/10.1002/jae.3071