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

DISCUSSION PAPER SERIES

IZA DP No. 13101

Quantiles of the Gain Distribution of an Early Childhood Intervention

Erich Battistin
Carlos Lamarche
Enrico Rettore

MARCH 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13101

Quantiles of the Gain Distribution of an Early Childhood Intervention

Erich Battistin
University of Maryland and IZA

Carlos Lamarche
University of Kentucky and IZA

Enrico Rettore
University of Padova and IZA

MARCH 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Quantiles of the Gain Distribution of an Early Childhood Intervention*

We offer a new strategy to identify the distribution of treatment effects using data from the Infant Health and Development Program (IHDP), a relatively understudied early-childhood intervention for low birth-weight infants. We introduce a new policy parameter, QCD, which denotes quantiles of the effect distribution conditional on latent neonatal health. The dependence between potential outcomes originates from a new class of factor models where latent health can affect the location and shape of distributions. We first show that QCD depends on quantiles of marginal outcome distributions given latent health. We then achieve identification of these marginal distributions and QCD by proxying latent health with neonatal anthropometrics and accounting for measurement error in these proxies. The effects of enrolling in IHDP are widely distributed across children and depend on neonatal health. Moreover, the large average effects documented in past work for close to normal birth weight children from low-income families are driven by a minority of children in this group.

JEL Classification: C13, C21, I14, J18

Keywords: early childhood, factor models, policy evaluation, quantile regression, treatment effect distributions

Corresponding author:
Carlos Lamarche
Department of Economics
University of Kentucky
223G Gatton College of Business & Economics
Lexington, KY 40506
USA

E-mail: clamarche@uky.edu

* This version: March 26, 2020. The authors would like to thank Aaron Sojourner, Roger Koenker and seminar participants at the University of Illinois at Urbana-Champaign, Queen Mary University of London, University of Kentucky, University of St. Gallen, and the 14th IZA/CEPR European Summer Symposium for comments on previous versions. Financial support from Fondazione CARIPARO (Progetti di Eccellenza) is gratefully acknowledged.

1. Introduction

Randomized assignment is ideal for investigating treatment effect heterogeneity, as distributions of potential outcomes with and without treatment are identified by design. If the distance between these distributions varies over the support, treatment effects must vary across participants. Quantile treatment effects (QTE) are often used to assess the extent of heterogeneity, as in influential studies by Heckman et al. (1997), Bitler et al. (2006), and Andrews et al. (2016), 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).

How can one learn about the distribution of gains? We revisit this question using data from the Infant Health and Development Program (IHDP), a randomized clinical trial which provided comprehensive early intervention services for children in poor health (IHDP, 1990). Learning about features of the effect distribution other than its average is particularly relevant in this context, as poor health during childhood is an important mechanism for transmission of economic conditions (Currie and Rossin-Slater, 2015). Besides, although IHDP is one of the few early-life interventions implemented as randomized experiment, it has received relatively scant attention from economists to date (Chaparro and Sojourner, 2019, is a recent example). 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 served both parents and children at the same time by combining center-based education with regular home visits and parent group meetings, as explained in Gross et al. (1997) among others.

We adopt a framework where potential outcomes with and without treatment are independent conditional on latent factors which affect location and scale of the outcome distribution in a flexible manner. This setting encompasses the traditional factor model as a special case and allows for treatment effect distributions which are more general than in past work. Our first contribution is to consider a collection of new policy parameters denoting quantiles of the conditional distribution (QCD) of treatment effects given the latent factors. In particular, we show that QTE and all distributional parameters discussed in Heckman et al. (1997) among others can be derived from knowledge of QCD. We also show that QCD can be written as a functional of quantiles of *marginal* outcome distributions conditional on latent factors. This relationship paves the way for identification and analogue estimation,

which is computationally convenient and can be applied to a number of empirical contexts beyond the specific case study.

We take this framework to IHDP data by letting potential outcomes depend on unobservable determinants of neonatal health. QCD in this setting allows us to study, for example, differences in the gain distribution of children who otherwise would score low on indicators of health and cognitive development. We additionally use features of the IHDP design to study how QCD varies with family income. As eligibility for early childhood interventions is often means tested, we know little about how distributional gains from early childhood interventions change with family resources. Specifically, past work has considered how the effects of IHDP vary with income or demographics by looking at differences between group *averages* (see, among others, Gross et al., 1997, and Duncan and Sojourner, 2013). The policy conclusions drawn from this approach rely entirely on the between-group variance of treatment effects. However, if gains are distributed unevenly *within* groups, the comparison between groups can be inadequate to fully address heterogeneity (Bitler et al., 2017). For example, this comparison could signal no heterogeneity in average effects despite having large disparities in how gains are distributed within groups. Our empirical investigation of IHDP data based on QCD offers policy recommendations which are more general than those obtained from a comparison of averages or from QTEs.

We identify QCD by exploiting smoothness in the relationship between (unobserved) quantiles of the outcome conditional on latent neonatal health and (observed) quantiles of the outcome conditional on *proxies* of neonatal health. Specifically, proxying latent health with observed measurements poses a problem because of measurement error in the proxies. By adapting results in Angrist et al. (2006), we show that outcome quantiles retrieved using proxies converge to true quantiles smoothly as measurement error vanishes at zero. If measurement error in these proxies is small, the measurement error variance is the leading term for this convergence (as shown in Chesher, 2017). We argue that the case of small error variance is the most relevant in many empirical applications on child development, as the choice of accurate proxies is often driven by the underlying theory and the measurement tools employed. Indeed, we show that this is true for IHDP data as well.

We discuss a strategy to simulate the bias arising from measurement error in proxies of neonatal health at different values of the error variance, and identify quantiles of the outcome conditional on latent neonatal health by extrapolation to the case of no error. Our 'identification at zero' strategy (Lewbel, 2019) requires cross-sectional data, and does not use instrumental variation or repeated measurements when the researcher has information on the measurement error variance. We also discuss a more general case when multiple auxiliary

proxies in a system of equations enable us to identify the measurement error variance. Montecarlo simulations demonstrate that Cook and Stefanski's (1994) simulation-extrapolation method works well and can be applied to IHDP data to obtain quantiles of potential outcomes conditional on latent health.

Our empirical investigation begins by finding the best proxy of a child's latent health prior to IHDP enrollment. 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 development throughout childhood and life-long well-being are documented (Almond and Currie, 2011). We show that an exploratory factor analysis of all neonatal anthropometrics supports a single-factor model for latent neonatal health. Moreover, we find that birth weight is the measurement of neonatal health with the lowest error variance (with a noise-to-signal ratio at about 5%).[1]

After proxying neonatal health with birth weight, we derive our second result - that a child's health endowment at birth affects the location *and* scale of potential outcome distributions with and without treatment (a fact that a standard factor model would fail to detect). Specifically, our outcome variables are indicators of cognitive development and health from conception to age three in the original study (IHDP, 1990). We find that better health endowment yields better and less dispersed outcomes in the baseline case of no participation. However, we also find that participation in IHDP orders children in the outcome distribution depending on neonatal health. This finding invalidates the assumption of rank similarity conditional on neonatal health (Dong and Shen, 2018, and Frandsen and Lefgren, 2018b), and serves as a catalyst for investigating treatment effect distributions.

We then present our third finding - that data reveals striking effect heterogeneity across IHDP participants during the first three years of life, as well as differences in how widely gains are distributed depending on neonatal health and family income. We estimate treatment effect distributions for IQ which are skewed towards, and more concentrated around, large positive values for those children endowed with better health at birth. As a result of this finding, the share of HLBW children (LBW and above 2000 grams) with positive returns from participation is 65% and about 20 points larger than in the LLBW group (LBW and less than 2000 grams). Health-related outcomes show smaller differences between LLBW and HLBW QCDs compared to IQ.

[1]We interpret birth weight as the child's health endowment after netting off demographics that correlate with other factors during pregnancy that may have affected condition at birth, as discussed for example in Chaparro and Sojourner (2019).

We conclude by showing that within income-group variability in treatment effects is at least as important as between income-group differences in treatment effects. Specifically, although the literature demonstrated that less well-off children benefited more from IHDP on *average*, we find that about 20% of children in the high-income group had gains larger than the median gain of children in the low-income group. We also show that the larger average effects for the low-income group are driven by large returns for a minority of children in this group. Finally, we find that the income gradient in treatment effects is much stronger for close to normal birth-weight infants, a group for which the causal effects of family income on health and development are well-documented (e.g., Ko et al., 2020). IHDP findings for this group are suggestive of the gain distribution from early childhood interventions targeting low-income families with normal birth-weight children.

Our results connect with a more general literature that aims at estimating heterogeneous treatment effect distributions. Classical Frechet inequalities yield sharp bounds on the joint distribution of potential outcomes which are consistent with their marginal distributions. Heckman et al. (1997) use Frechet inequalities to bound the effect distribution. The resulting bounds are, however, not sharp, which is a result following from Makarov (1981). Fan and Park (2010) and Firpo and Ridder (2019) obtain sharp bounds on the distribution of the treatment effect. Several studies have proposed alternative bounds that make use of additional assumptions, among which are Manski (1997), Fan et al. (2017), Frandsen and Lefgren (2018a), and Callaway (2019). The effect distribution is point-identified if each unit maintains the same rank across potential outcome distributions (rank invariance). Point-identification is also achieved if gains are independent of the non-participation outcome. In our case study, this condition implies that IHDP would spread gains equally across children with high and low outcomes in the *status quo*. This can be an overly restrictive condition in ours and many other contexts. The approach developed in this paper yields point-identification of gain distributions and a computationally convenient estimation strategy that works well when one has good proxies of the latent factors causing the dependence between potential outcomes.

The remainder of the paper is organized as follows. Section 2 provides the institutional details underlying IHDP implementation and describes the working sample. It also shows average and quantile treatment effects, setting the stage for investigating treatment effect distributions in Section 3. QCD is defined in Section 4, where the assumptions maintained throughout the paper are also presented. Section 5 presents identification results, Section 6 discusses estimation, and Section 7 offers a simulation study on the performance of the proposed approach. Our empirical investigation of the IHDP is presented in Section 8 and

Section 9 concludes. Derivations of mathematical results are provided in Appendices A, B, and C, and additional empirical results in the Online Appendix.

2. Context, Data, and Descriptive Statistics

Working Sample. We use data from IHDP, a randomized clinical trial implemented starting in 1985 to enhance the cognitive, behavioral, and health status of low birth weight (less than 2500 grams) and premature (less than 37 weeks) infants born in catchment areas of selected US medical institutions.[2] Its most unique feature is perhaps the population targeted, as in early childhood development interventions eligibility typically depends on the socio-economic status of parents. IHDP provided services following a center-based and home-based curriculum. The program 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 Gross et al., 1997, and Elango et al., 2016). The intervention lasted until the age of 3, and data were collected on treatment and control children until the age of 18.

The size of the IHDP sample parallels that of many other model programs in the literature. Two strata were considered at the randomization stage: the "lighter" low birth-weight group (LLBW, less than 2000 grams), and the "heavier" low birth-weight group (HLBW, between 2000 and 2500 grams). We use children from both strata in the primary analysis dataset, which is the sample considered 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.

Our analysis is restricted to three outcomes among those considered in the original study, all measured at 36 months. We consider an 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 three years of life (higher values of this index indicate lower health). We also use the General Health Ratings Index, from the Rand Corporation Health Insurance Study, which is constructed using the maternal perception of child health over the first three years of life (higher values of this index indicate better health). Due to missing data, our final sample includes 929 infants, with 343 infants in the HLBW group and 586 infants in the LLBW group (Table 1).

[2]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.

Balancing Tests and Average Effects. The baseline characteristics of 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 on the treatment indicator controlling for randomization strata (medical site and birth-weight group). Outcomes in these regressions are not standardized, and significance is assessed using heteroskedasticity-robust standard errors. Significant differences emerge in a limited number of cases, although the size of these differences is negligible when compared to standard deviations in column (1). This finding is consistent with the fact that randomization did not yield any particular problem (see IHDP, 1990). Balancing tests shown in the remaining columns of the table 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 randomization strata. 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) in Panel A report treatment effect estimates obtained from regressions that include mother demographics (shown in Panel B of Table 1). The estimates shown in columns (4), (8), and (12) are from regressions with both mother demographics and infant demographics at birth (Panel A of Table 1). Demographics at baseline are used in these specifications to control for residual differences in the treatment and control groups, and yield similar conclusions.

The average gains on health were smaller than for IQ and driven by LLBW children, as shown in Panel B of Table 2. 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 0.84 to 0.94 range depending on the specification adopted). However, there are no statistically significant improvements, on average, using the General Health Ratings Index. Estimates in columns (10) to (12) demonstrate that the effects on health are concentrated in the lighter birth-weight group. 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.

Treatment Effect Heterogeneity. Figure 1 weighs against the possibility of constant gains from IHDP enrollment. Dots reported here are QTE estimates for the three outcomes in our analysis. Treatment-control differences at each quantile are obtained from quantile regressions on the treatment indicator controlling for randomization strata, mother demographics and infant demographics at birth. Each panel in the figure also plots values generated by a local linear regression (LLR) fit to 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 appear to vary across quantiles.[3] For example, the figure shows a difference across outcomes of at least 26% from the 0.1 quantile to the 0.9 quantile.

Treatment 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 non-participants. The rationale for considering this relationship stems from the idea of rank similarity, which is often used as a benchmark in empirical work (see Chernozhukov and Hansen, 2005, Dong and Shen, 2018, and Frandsen and Lefgren, 2018b). Specifically, under rank similarity the distribution of ranks should be identical for participants and non-participants conditional on anthropometrics if these variables are "rank shifters". This assumption is grounded on discussions in IHDP documentation, early evaluations of the program (IHDP, 1990) and other studies on child development (for example, Conti et al., 2018). Following the procedure in Frandsen and Lefgren (2018b), we find that the hypothesis of rank similarity is rejected by the data used in our study (results are available on request).

3. General Formulation of the Problem

The notation employed in the potential outcome approach to causal inference is used throughout. All results that follow hold conditional on a p-dimensional vector of covariates $\boldsymbol{X} \in R^p$, 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. Since only $Y = Y_0 + D(Y_1 - Y_0)$ is observed, identification of the distribution of gains $\Delta = Y_1 - Y_0$ is precluded because of a missing data problem. Randomized assignment to the

[3]However, the limited size of treatment and control samples affects the precision around estimates of QTEs, and the hypothesis of constant gains is only marginally rejected in the data. Figure A.1 in the Online Appendix reports LLR estimates of QTE first derivatives along with point-wise 95% confidence intervals at each quantile. Point estimates suggest that QTEs have different slopes across quantiles, weighing against the hypothesis of constant gains. Confidence intervals, however, are large and include zero in all cases unless one considers confidence intervals with coverage between 85% and 90% – results are available on request.

intervention does not eliminate this problem: as randomization reveals only one potential outcome, a child's gain is not identified nor is the gain distribution across children.

We model the dependence between potential outcomes as the result of persistent, long term *factors* proxying conditions prior to IHDP enrollment. This approach has become a workhorse in empirical work, and uses a factor model to take the idea to the data (see, among others, Carneiro et al., 2003, Aakvik et al., 2005, Cunha et al., 2010, and Attanasio et al., 2020). In this setting, the dependence between latent factors and outcomes is mediated by the effects of the intervention. Specifically, factor loadings vary with potential outcomes, which are allowed to have different variances. The model is completed by assuming that the dependence between potential outcomes Y_0 and Y_1 is driven entirely by their common latent factors, as we explain in the next section.[4]

Our approach yields an expression for the gain distribution which is more general than the one implied by the traditional factor model. We consider the following equations for $d \in \{0, 1\}$:

$$Y_d = \boldsymbol{\lambda}_d(U_d)'\boldsymbol{\Theta} + U_d, \tag{1}$$

where $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_r)'$ is a vector of r continuous random variables (factors) and U_d is a scalar random component for unexplained variability.[5] This model allows factors to be location-scale shifters for an individual's rank in the gain distribution, and it also implies:

$$\Delta = \left(\boldsymbol{\lambda}_1(U_1)' - \boldsymbol{\lambda}_0(U_0)'\right)\boldsymbol{\Theta} + U_1 - U_0. \tag{2}$$

For example, in the case of a location-scale shift model equation (1) becomes:

$$Y_d = \boldsymbol{\lambda}_d'\boldsymbol{\Theta} + \left(1 + \boldsymbol{\gamma}_d'\boldsymbol{\Theta}\right)U_d, \tag{3}$$

where $\boldsymbol{\lambda}_d = (\lambda_{1d}, \ldots, \lambda_{rd})'$ are *factor loadings*, $\boldsymbol{\gamma}_d = (\gamma_{1d}, \ldots, \gamma_{rd})'$ are *scale parameters* and:

$$\boldsymbol{\lambda}_d(U_d) = \boldsymbol{\lambda}_d + \boldsymbol{\gamma}_d U_d. \tag{4}$$

Equation (3) defines a factor model for potential outcomes, while allowing for heteroskedasticity in the distribution of Y_d conditional on $\boldsymbol{\Theta}$ if $\boldsymbol{\gamma}_d$ is different from zero. The traditional

[4]Latent factor models have been used in the program evaluation literature also to gain external validity in regression discontinuity designs (Rokkanen, 2015) and in time-varying treatment effect models (Cooley Fruehwirth et al., 2016). Chaparro and Sojourner (2019) look at heterogeneity in average effects of IHDP with respect to variables defined from latent factors.

[5]The model in (1) is similar to the one in Chen et al. (2019). Specifically, they consider estimation of a class of factor models for high dimensional data using panel data.

factor model is obtained by setting $\boldsymbol{\gamma}_d$ to zero, in which case $\boldsymbol{\Theta}$ becomes a location shifter in the distribution of Δ.[6]

Our model also allows for more general violations of rank similarity than in the traditional factor model. For example, rank similarity conditional on $\boldsymbol{\Theta}$ in equation (3) requires identical distributions for errors of potential outcome equations:

$$(1 + \boldsymbol{\gamma}_0'\boldsymbol{\theta})\, U_0 | \boldsymbol{\Theta} = \boldsymbol{\theta} \sim (1 + \boldsymbol{\gamma}_1'\boldsymbol{\theta})\, U_1 | \boldsymbol{\Theta} = \boldsymbol{\theta},$$

which is equivalent to:

$$(1 + \boldsymbol{\gamma}_0'\boldsymbol{\theta})\, U_0 \sim (1 + \boldsymbol{\gamma}_1'\boldsymbol{\theta})\, U_1,$$

because the vector of latent factor $\boldsymbol{\Theta}$ and U_d for $d \in \{0, 1\}$ are assumed independent (see Assumption 1 below). It follows that rank similarity can be violated here because of differences in scale parameters, $\boldsymbol{\gamma}_0 \neq \boldsymbol{\gamma}_1$, or in the distributions of U_0 and U_1. Our investigation of IHDP data yields identical distributions of U_0 and U_1, but different scale coefficients.[7]

4. Parameter of Interest and Assumptions

Consider that the variables $(Y, D, \boldsymbol{W}')$ are observed for a sample of units randomly drawn from the relevant population, where Y is a scalar continuous outcome and $\boldsymbol{W} = (W_1, \ldots, W_r)'$ is a vector of r continuous random variables. Factors are not observed, but proxied by $\boldsymbol{W} \in R^r$. The following measurement equation is defined:

$$\boldsymbol{W} = \boldsymbol{\Theta} + \boldsymbol{V}, \tag{5}$$

where $\boldsymbol{V} = (V_1, \ldots, V_r)'$ is $r \times 1$. Components of the vector $(U_d, \mathbf{V}') \in R^{r+1}$, for $d \in \{0, 1\}$, are assumed mutually independent and independent of latent factors $\boldsymbol{\Theta}$. Measurement errors $\mathbf{V}$ are assumed to have zero mean. The classical properties of measurement error are maintained here to connect with factor models used in most empirical work. However, the approach we take in Section 5 is also valid in more general non-classical models. The independence restriction could be relaxed, for example by allowing the variance of measurement error to

[6]More general forms of heteroskedasticity than in equation (3) could be considered without affecting the results that follow. For example, our identification strategy extends to settings where the loading $\boldsymbol{\lambda}_d(U_d)$ is monotone in U_d. Equation (3) is supported by IHDP data, as we discuss below.

[7]Simple calculations show that our approach can be used to point-identify the density of scoring at the τ-th quantile of the Y_1 outcome conditional on scoring at the same quantile of the baseline outcome Y_0. Using this result, one can define test statistics for the hypothesis of rank invariance in addition to those discussed in the literature (Dong and Shen, 2018, and Frandsen and Lefgren, 2018b).

depend on $\boldsymbol{\Theta}$ (Carroll et al., 2006). This would introduce additional parameters whose values would need to be specified in the estimation strategy below.[8]

Assumption 1. (**Factor Model Representation**). In the model defined by equations (1) and (5) for $d \in \{0, 1\}$, the uniqueness U_d, errors in measurement equations $\boldsymbol{V}$ and latent factors $\boldsymbol{\Theta}$ have continuous distributions, with zero mean and finite moments, and are mutually independent:

$$F_{U_d \boldsymbol{V}' \boldsymbol{\Theta}}(u_d, \mathbf{v}, \boldsymbol{\theta}) = F_{U_d}(u_d) \left[\prod_{k=1}^{r} F_{V_k}(v_k) \right] F_{\boldsymbol{\Theta}}(\boldsymbol{\theta}).$$

We maintain the assumption that treatment is randomly allocated across units, as in the case of IHDP. Randomization ensures that treatment and control units are representative of the same population, so that the conditioning on D is irrelevant. Assumption 2 is made here for convenience: identification stems from the independence condition in Assumption 1 which is implied by the factor structure, rather than from the assignment mechanism. Our analysis extends to regression discontinuity designs if units closest to the cutoff are viewed as being part of a local randomized experiment. Variants to randomization may be considered to allow for non-random selection into treatment (such as ignorability conditional on some observables $\boldsymbol{X}$; see Abbring and Heckman, 2007).

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

Finally, the conditional independence condition stated in Assumption 3 has been used in past work to achieve identification of the gain distribution and is standard in the literature on factor models (see, e.g., Abbring and Heckman, 2007). Together with Assumption 1, it implies that the dependence between potential outcomes is driven by $\boldsymbol{\Theta}$. Battistin and Chesher (2014) show that independence conditional on $\boldsymbol{\Theta}$ does not imply independence conditional on $\boldsymbol{W}$. Thus, the identification results presented below are in general not valid by just conditioning on $\boldsymbol{W}$.

Assumption 3. (**Conditional Independence**) The random variables (U_1, U_0) are independent: $F_{U_0 U_1}(u_0, u_1) = F_{U_0}(u_0) F_{U_1}(u_1)$.

We consider the following parameter:

$$\Delta(\tau; \boldsymbol{\theta}, u_0) := Q_{Y_1 - Y_0} \left(\tau | \boldsymbol{\Theta} = \boldsymbol{\theta}, U_0 = u_0 \right),$$

[8]The notation $F_A(a|B = b)$ indicates the distribution of random variable A calculated at a conditional on random variable B taking value b. A similar notation is employed for the conditional τ-quantile function $Q_A(\tau|B = b) \equiv F_A^{-1}(\tau|B = b)$. Joint distributions $F_{AB}(a, b)$ are defined similarly.

which represents the quantile function of the conditional distribution (QCD) of treatment effects given $\mathbf{\Theta} = \boldsymbol{\theta}$ and $U_0 = u_0$. Specifically, we are interested in the collection of QCDs at different quantiles of the uniqueness U_0 and factors $\mathbf{\Theta}$. Knowledge of $\Delta(\tau; \boldsymbol{\theta}, u)$ allows to answer policy questions regarding, for instance, how widely gains are distributed in the IHDP-treatment group, or to study gains for children of this group at specific values of the base state distribution. Using equation (2), Assumption 1 together with Assumption 3 imply:

$$\Delta(\tau; \boldsymbol{\theta}, u_0) = \left(\boldsymbol{\lambda}_1 \left(Q_{U_1}(\tau)\right)' - \boldsymbol{\lambda}_0(u_0)'\right) \boldsymbol{\theta} + Q_{U_1}(\tau) - u_0, \tag{6}$$

which defines a point-identifying functional conditional on $\mathbf{\Theta} = \boldsymbol{\theta}$ and $U_0 = u_0$. Equation (6) sets the stage for analogue estimation of QCD from knowledge of the quantile functions $Q_{U_d}(\tau)$ and $Q_{\mathbf{\Theta}}(\tau)$, and factor loadings $\boldsymbol{\lambda}_d(U_d)$.

All distributional parameters used in the evaluation literature can be written as weighted versions of QCDs. For example, QCD is more general than QTE because the latter parameter can be obtained from knowledge of the former parameter. We discuss some examples next.

Example 1. QCDs are sufficient to retrieve the conditional density of the gain distribution at values $\mathbf{\Theta} = \boldsymbol{\theta}$ and $U_0 = u_0$ using the relationship:

$$f_{Y_1 - Y_0}\left(\Delta\left(\bar{\tau}; \boldsymbol{\theta}, u_0\right) | \mathbf{\Theta} = \boldsymbol{\theta}, U_0 = u_0\right) = \left[\frac{\partial}{\partial \tau} \Delta\left(\tau; \boldsymbol{\theta}, u_0\right)\Big|_{\tau = \bar{\tau}}\right]^{-1},$$

which follows from differentiation of the quantile function. The conditional density of gains at $\mathbf{\Theta} = \boldsymbol{\theta}$ is obtained by integrating the last expression with respect to:

$$f_{U_0}\left(Q_{U_0}\left(\bar{\tau}\right)\right) = \left[\frac{\partial}{\partial \tau} Q_{U_0}\left(\tau\right)\Big|_{\tau = \bar{\tau}}\right]^{-1}.$$

These expressions can be simplified considerably because of the factor model representation in (3). Specifically, since:

$$\frac{\partial}{\partial \tau} \Delta\left(\tau; \boldsymbol{\theta}, u_0\right)\Big|_{\tau = \bar{\tau}} = \left(1 + \boldsymbol{\gamma}_1' \boldsymbol{\theta}\right) \frac{\partial}{\partial \tau} Q_{U_1}\left(\tau\right)\Big|_{\tau = \bar{\tau}},$$

we have that:

$$f_{Y_1 - Y_0}\left(\eta | \mathbf{\Theta} = \boldsymbol{\theta}\right) = \left(1 + \boldsymbol{\gamma}_1' \boldsymbol{\theta}\right)^{-1} \int_0^1 \left[\frac{\partial}{\partial \tau} Q_{U_1}\left(\tau\right)\Big|_{\tau = \tau_\eta} \frac{\partial}{\partial \tau} Q_{U_0}\left(\tau\right)\Big|_{\tau = \tau_0}\right]^{-1} d\tau_0,$$

where $\tau_\eta \equiv F_{Y_1 - Y_0}\left(\eta | \mathbf{\Theta} = \boldsymbol{\theta}, U_0 = Q_{U_0}\left(\tau_0\right)\right)$ is the cumulative distribution at the point η in the gain distribution conditional on $\mathbf{\Theta} = \boldsymbol{\theta}$ and $U_0 = Q_{U_0}\left(\tau_0\right)$.[9]

[9]The factor structure in (3) also implies simple expressions for the average relationship between Y_1 and the base state value Y_0. For example, $E\left[Y_1 | Y_0 = y_0\right] = \boldsymbol{\lambda}_1' E\left[\mathbf{\Theta} | Y_0 = y_0\right] = \boldsymbol{\lambda}_1' E\left[\boldsymbol{W} | Y = y_0, D = 0\right]$.

Example 2. Integrating out latent factors in equation (6) at the value $u = Q_{U_0}(\tau)$ yields:

$$QTE = E\left[\Delta\left(\tau; \boldsymbol{\theta}, Q_{U_0}(\tau)\right)\right].$$

It follows that QTE can be written as the average of QCDs because Y_0 and Y_1 are independent conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$.

5. Identification

The Effects of Measurement Error. There are two sources of error in the model represented by equations (1) and (5). The first error U_d is in the outcome equation (1). Following the literature on factor models, we refer to this quantity as *uniqueness*. The second error $\boldsymbol{V}$ arises because $\boldsymbol{W}$ are error-ridden measurements of latent factors $\boldsymbol{\Theta}$. We consider a third error here, $\boldsymbol{\epsilon}$, which is instrumental to attaining identification in the way we discuss below. Let $\boldsymbol{\Sigma}$ be the variance of $\boldsymbol{V}$, which we assume known at this stage. We define:

$$\tilde{\boldsymbol{W}} = \boldsymbol{W} + \sqrt{\rho}\boldsymbol{\epsilon},$$

for a known constant ρ by adding $\boldsymbol{\epsilon}$ to the measurement $\boldsymbol{W}$. We assume that $\boldsymbol{\epsilon}$ is distributed as a (multivariate) normal random variable with zero mean and variance $\boldsymbol{\Sigma}$, and independent of all other variables. Basic manipulations yield:

$$\boldsymbol{\Theta} = \tilde{\boldsymbol{W}} - ((1+\rho)\boldsymbol{\Sigma})^{1/2}\tilde{\boldsymbol{V}}, \tag{7}$$

where $\tilde{\boldsymbol{V}} = (V_1, \ldots, V_r)'$ and the V_l are standardized random variables.

We use the relationship in (1) to write:

$$Q_{Y_d}(\tau|\boldsymbol{B} = \boldsymbol{b}) = \boldsymbol{b}'\boldsymbol{\phi}_d(\tau), \tag{8}$$

where $\boldsymbol{\phi}_d(\tau) := \left(\boldsymbol{\lambda}_d\left(Q_{U_d}(\tau)\right), Q_{U_d}(\tau)\right)'$ is the $(r+1)\times 1$ vector of slopes and intercepts at quantile τ, and $\boldsymbol{B} = (\boldsymbol{\Theta}', 1)'$. After replacing (7) in (1), we obtain an alternative quantile function:

$$Q_{Y_d}(\tau|\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{V}} = \tilde{\boldsymbol{v}}) = \tilde{\boldsymbol{b}}'\boldsymbol{\phi}_d(\tau) + \tilde{\boldsymbol{v}}'\boldsymbol{\pi}_d(\tau),$$

where $\tilde{\boldsymbol{B}} = (\tilde{\boldsymbol{W}}', 1)'$ and $\boldsymbol{\pi}_d(\tau) = -\boldsymbol{\lambda}_d'(\tau)((1+\rho)\boldsymbol{\Sigma})^{1/2}$. The solution to the following unfeasible problem:

$$(\boldsymbol{\phi}_d(\tau)', \boldsymbol{\pi}_d(\tau)') = \operatorname{argmin} E\left[\varrho_\tau\left(Y_d - \tilde{\boldsymbol{B}}'\boldsymbol{\phi}_d - \tilde{\boldsymbol{V}}'\boldsymbol{\pi}_d\right)\right], \tag{9}$$

yields $\boldsymbol{\phi}_d(\tau)$, where $\varrho_\tau(u) = u(\tau - I(u<0))$ is the quantile loss function (see Koenker, 2005).

As the omitted variables formula extends to quantile regression (see the "short" versus "long" derivations in Angrist et al., 2006), we can determine the bias in the feasible regression

counterpart of (9) when $\tilde{\boldsymbol{V}}$ is the omitted variable. Specifically, let $\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau)$ be the $(r+1)\times 1$ vector retrieved from feasible quantile regressions of Y on $\tilde{\boldsymbol{B}}$ for participants ($D=1$) and non-participants ($D=0$):

$$\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau) = \text{argmin}\, E\left[\varrho_\tau\left(Y_d - \tilde{\boldsymbol{B}}'\boldsymbol{\phi}_d)\right)\right]. \tag{10}$$

For any $\rho \geq 0$, the quantity $\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau)$ is the counterfactual parameter that would be estimated had the measurement error variance been $(1+\rho)\boldsymbol{\Sigma}$ instead of $\boldsymbol{\Sigma}$. The parameter retrieved from raw data is obtained by setting $\rho = 0$. We show in Appendix A that:

$$\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau) - \boldsymbol{\phi}_d(\tau) \approx -E\left(\left[\sigma(\tau,\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}'\right]\right)^{-1} E\left[\sigma(\tau,\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{V}}'\left(\boldsymbol{\lambda}_d(\tau)'((1+\rho)\boldsymbol{\Sigma})^{1/2}\right)\right], \tag{11}$$

where $\sigma(\tau,\boldsymbol{\theta})$ are quantile specific weights defined as the conditional standard deviation times the sparsity parameter associated with the quantile function $Q_{U_d}(\tau)$. It follows that the error $\boldsymbol{V}$ hampers identification by attenuating estimates from unadjusted quantile regression.

Extrapolation to Zero Error. We achieve identification by exploiting smoothness in the relationship between $\boldsymbol{\phi}_d(\tau)$ and the error-ridden values of slopes and intercepts $\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau)$ at different ρ's. One can take stock of this 'identification at zero error' idea by noting that:

$$\lim_{\rho\to -1} \tilde{\boldsymbol{\phi}}_d^{\rho}(\tau) = \boldsymbol{\phi}_d(\tau).$$

The term $\boldsymbol{\epsilon}$ is fundamental for identification, in that it adds instrumental variability to estimate the counterfactual bias arising from increasingly larger values of the measurement error variance. However, $\boldsymbol{V}$ and the variable $\tilde{\boldsymbol{V}}$ generated through $\boldsymbol{\epsilon}$ need not have the same distribution. Because of this, smoothness alone might not be sufficient for identification as values $\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau)$ at different ρ's will result from changes in the measurement error variance and a measurement error distribution different from that in raw data.[10]

The problem is solved if measurement errors $\boldsymbol{V}$ are *normally* distributed. This result follows because, in this case, $\tilde{\boldsymbol{V}}$ are also normal. This case may not be uncommon in empirical applications. For example, when measurements $\boldsymbol{W}$ are obtained as group means, we can expect close to normally distributed measurement errors with modest group sizes. Also, many random variables have a distribution which is approximately normal. This is the case if the components of $\boldsymbol{V}$ (re-centered at zero) are Beta random variables with

[10]This problem arises in quantile regression. To see why, it is convenient to consider the alternative case of linear regression, where the bias from using $\tilde{\boldsymbol{W}}$ instead of $\boldsymbol{\Theta}$ depends only on the measurement error variance. In particular, this bias is independent of the specific distributions of U_d and $\boldsymbol{V}$. Contamination of measurement error will affect bias only by inflating the measurement error variance. Importantly, this result is valid for any mean-zero distribution of $\boldsymbol{\epsilon}$. Like in any non-linear model, measurement error in quantile regression will result in bias in a more complicated manner. Bias will depend on the distribution of $\boldsymbol{V}$, $\boldsymbol{\epsilon}$ and U_d, as we show in (11).

shape parameters equal and large, or Gamma random variables with large shape parameters relative to their scale parameters. More in general, any contamination of $\boldsymbol{V}$ which is closed or approximately closed under convolution with $\boldsymbol{\epsilon}$ would also solve the problem.

More in general, it is possible to approximate $\boldsymbol{\phi}_d(\tau)$ sufficiently well under weaker distributional assumptions as long as the measurement error variance is small, which is the relevant case in IHDP data. To see this, note that the solution to (10) coincides with the solution to the unfeasible problem (9) when $\boldsymbol{\pi}_d = \mathbf{0}$, for $d \in \{0, 1\}$. It follows that, for small measurement error variance, conditional quantiles given $\tilde{\boldsymbol{W}}$ must be approximately equal to conditional quantiles given $\boldsymbol{\Theta}$. Using this "small variance" idea, Chesher (2017) derives an approximation to the bias when the measurement error in $\tilde{\boldsymbol{W}}$ is small. Exact calculations in his work suggest that the approximation is still accurate with more measurement error than in our empirical analysis below. By adapting Chesher's idea to the case of the quantile function in (8), we show in Appendix B that bias depends on the distribution of U_d, derivatives of the quantile function, and the distribution of $\boldsymbol{\Theta}$. Importantly, this bias will not depend on the specific distribution of the measurement error $\tilde{\boldsymbol{V}}$, but only its variance $(1 + \rho)\boldsymbol{\Sigma}$.[11] Using this result, one can rely on the fact that the bias in (11) is approximately independent of the error distribution generated through $\boldsymbol{\epsilon}$.

One can expect that the performance of the extrapolation to the case of $\rho \to -1$ depends on the particular form of the distribution of U_d. An inspection of Chesher's approximation adapted to the case at hand also reveals that heavily skewed distributions for U_d will in general yield larger bias across all quantiles τ. For distributions symmetric around zero, some terms in this approximation will go to zero, yielding smaller bias across all quantiles. The relevance of this problem is ultimately an empirical matter, which we investigate in the simulation section below.

6. Estimation

Error-Free Quantile Functions. We start by considering estimation of $\boldsymbol{\lambda}_d\left(Q_{U_d}(\tau)\right)$ and $Q_{U_d}(\tau)$, for $d \in \{0, 1\}$, when the measurement error variance $\boldsymbol{\Sigma} = diag(\sigma_1^2, \ldots, \sigma_r^2)$ is *known*. We proceed in two steps. We first simulate how the quantity $\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau)$ varies with ρ using multiple draws from the distribution of $\boldsymbol{\epsilon}$. This step requires repeated computation of quantile regressions from error-ridden data, as we explain below. The quantity $\boldsymbol{\phi}_d(\tau)$ is then obtained by extrapolating this profile to the case of no error, $\rho = -1$. Because of the relationship between $\tilde{\boldsymbol{\phi}}_d^{\rho}(\tau)$ and $\boldsymbol{\phi}_d(\tau)$ in (11), we expect any smooth extrapolant to work well in most

[11] To offer a result that is consistent with the small variance approximation framework used by Chesher (2017), equation (B.2) in Appendix B is obtained for the case where $\rho = 0$.

realistic applications. The combination of simulation and extrapolation steps extends Cook and Stefanski's (1994) SIMEX to quantile regression. The logic for estimating $\boldsymbol{\phi}_d(\tau)$ corresponds to the algorithm below, and the idea is most simply understood using its graphical interpretation in Figure 2.[12] As we discussed, extrapolation to no-error is more appropriate when the proxies used for $\boldsymbol{\Theta}$ have small measurement error variance (as in IHDP data). Because of this result, we will use SV-SIMEX – small variance SIMEX – to refer to this algorithm in what follows.

(i) For any given value $\rho > 0$, generate $\tilde{\boldsymbol{W}}_b^\rho = (\tilde{W}_{1,b}^\rho, ..., \tilde{W}_{r,b}^\rho)'$ as follows:

$$\tilde{W}_{l,b}^\rho = W_l + \sqrt{\rho\sigma_l^2}\epsilon_{l,b},$$

for $l = 1, \ldots, r$, where the $\epsilon_{l,b}$'s are independently distributed standard normals. This step adds to W_l additional error with mean zero and variance $\rho\sigma_l^2$, independently across all measurements in $\boldsymbol{W}$. The last equation implies:

$$\tilde{W}_{l,b}^\rho = \Theta_l + V_l + \sqrt{\rho\sigma_l^2}\epsilon_{l,b},$$

for $l = 1, \ldots, r$, so that $\tilde{W}_{l,b}^\rho$ can be interpreted as an error-ridden proxy of the latent factor Θ_l with measurement error variance equal to $(1+\rho)\sigma_l^2$.

(ii) Use quantile regression to estimate slopes and intercepts of Y on $\tilde{\boldsymbol{W}}_b^\rho$, separately for participants ($D = 1$) or non-participants ($D = 0$). The estimated coefficients are denoted by $\hat{\boldsymbol{\phi}}_{d,b}^\rho(\tau)$, for $d \in \{0, 1\}$.

(iii) For B large and $b = 1, \ldots, B$, the same procedure is iterated to obtain:

$$\hat{\boldsymbol{\phi}}_d^\rho(\tau) := \frac{1}{B}\sum_{b=1}^{B} \hat{\boldsymbol{\phi}}_{d,b}^\rho(\tau),$$

for $d \in \{0, 1\}$.

(iv) Estimate a parametric model of $\hat{\boldsymbol{\phi}}_d^\rho(\tau)$ on ρ, and extrapolate at $\rho = -1$ (no error).

Two features of SV-SIMEX deserve further discussion. First, it is possible to extrapolate the measurement error variance back to zero using alternative parametric models. One might expect that the estimator performance depends on the specific model of choice. The simulation results in Section 7 show that a quadratic model in ρ is a good choice in a number of settings that we consider. Second, perturbation of $\boldsymbol{W}$ with increasingly higher measurement error requires knowledge of the measurement error variances σ_l^2, for $l = 1, \ldots, r$.

[12] To the best of our knowledge, with the exceptions of Shang (2012) and Torres-Saavedra (2013), this is the first application of SIMEX to quantile regression.

Often these variances are derived from auxiliary data or knowledge of the context under investigation. We discuss below how to estimate the measurement error variance when external information is not available.

Although our recommendation is to use SV-SIMEX whenever possible, any alternative approach to consistent estimation of (8) can be used to obtain identification of QCD via (6). This choice depends on the application and the data available. For example, error-free quantile functions could be obtained from the corrected-loss function estimator proposed by Wang et al. (2012). Although their estimator is consistent in a general class of quantile regression models, it requires smoothing the objective function to correct for bias and the correction depends on the distributional assumptions of the measurement error variable. Our approach does not require the use of numerical integration and outperforms the corrected-loss function estimator when the measurement error variance is small, as shown in a simulation study in the Online Appendix.[13]

Latent Factors Distributions. Standard SIMEX can be applied to $f_{W_l}(\tau)$, so that $f_{\Theta_l}(\tau)$ for $l = 1, \ldots, r$ can be retrieved by extrapolation. Wang et al. (2010), in extending Stefanski and Bay's (1996) idea, show that SIMEX outperforms Fourier-type methods for density estimation with small sample size and large measurement error variances. They also document good performance of SIMEX without normal measurement error. The estimated densities $f_{\Theta_l}(\tau)$, for $l = 1, \ldots, r$, can be used to condition on specific values $\boldsymbol{\theta}$ in the support of $\boldsymbol{\Theta}$ in the computation of QCD.

When the measurement error variance is small, we show in Appendix B that the small-error approximation can be used for approximating the quantiles $Q_{\Theta_l}(\tau)$ for $l = 1, \ldots, r$. There are alternatives to the approximation with small variance. For example, when repeated measurements of $\boldsymbol{\Theta}$ are available, Kotlarski's (1967) theorem paves the way for non-parametric estimation of $f_{\Theta_l}(\tau)$ using deconvolution methods (Li and Vuong, 1998).

Alternatively, one can use Assumption 1 to write:

$$E(\boldsymbol{W}|Y_0 = y_0) = E(\boldsymbol{\Theta}|Y_0 = y_0).$$

This quantity represents the average of $\boldsymbol{\Theta}$ for units at different values of the baseline outcome. The QCD in (6) can then be computed at $\boldsymbol{\theta} = E(\boldsymbol{W}|Y_0 = y_0)$, the latter quantity being

[13]The literature offers alternative approaches. For instance, endogeneity arising from measurement error prompts the use of instrumental variables, which are not required in our approach. Instrumental variation is used by Schennach (2008), who considers non-parametric estimation using sieved-based methods. Wei and Carroll (2009) propose a method that makes use of external information. Firpo et al. (2017) show how to correct for measurement error when repeated measurements of the true regressors are available. If the researcher has panel data, the approach by Chen et al. (2019) is also an option.

non-parametrically estimated from the sample of non-participants ($D = 0$). This is the approach we take in our empirical investigation of IHDP.

Unknown Measurement Error Variances. Additional data are required when the measurement error variance is unknown. Influential empirical work on factor models, reviewed above, employs the factor structure to retrieve this unknown variance from multiple proxies of the latent factors. Specifically, assume to have a $r \times 1$ vector $\boldsymbol{Z}$ of additional variables, whose properties are presented in the following assumption.

Assumption 4. (**Instruments**). The variables $\boldsymbol{Z}$ are such that $E(\boldsymbol{V}|\boldsymbol{Z} = \boldsymbol{z}) = E(\boldsymbol{V})$ and $E(Y_d|\boldsymbol{\Theta} = \boldsymbol{\theta}, \boldsymbol{Z} = \boldsymbol{z}) = E(Y_d|\boldsymbol{\Theta} = \boldsymbol{\theta})$ for $d \in \{0, 1\}$.

Assumption 4 states that (i) $\boldsymbol{Z}$ must be mean independent of measurement error $\boldsymbol{V}$, and (ii) the effect of $\boldsymbol{Z}$ on the outcome Y_d must pass only through $\boldsymbol{\Theta}$, for $d \in \{0, 1\}$. In other words, the information on the outcome brought along by $\boldsymbol{Z}$ is coarser than is the information in $\boldsymbol{\Theta}$. This is a standard exclusion restriction in the measurement error literature (see, e.g., Hu and Schennach, 2008, and Chen et al., 2011), and in factor analysis more in general. When additional proxies of $\boldsymbol{\Theta}$ are used, $\boldsymbol{Z}$ brings instrumental variation to identify the measurement error variance which is akin to that from repeated measurements.

Using equation (3), we show in Appendix C that Assumption 1 combined with Assumption 4 imply the following moment conditions for $d \in \{0, 1\}$:

$$\begin{aligned} E(Y_d|\boldsymbol{Z} = \boldsymbol{z}) &= \boldsymbol{\lambda}_d' E(\boldsymbol{W}|\boldsymbol{Z} = \boldsymbol{z}), \\ E\left((\boldsymbol{\lambda}_d'\boldsymbol{W} - Y_d)\frac{W_l}{\lambda_{ld}}\right) &= \sigma_l^2, \qquad l = 1, \ldots, r. \end{aligned} \tag{12}$$

The first equation implies that a consistent estimate of $\boldsymbol{\lambda}_d$ can be retrieved from two-stage least squares (2SLS) using $\boldsymbol{Z}$ as instruments. Assumption 2 ensures that regressions can be estimated using treatment and control samples. Knowledge of factor loadings identifies the variance of the measurement errors $\boldsymbol{V}$ through the moment condition (12). Note that two independent estimates of σ_l^2, $l = 1, \ldots, r$, can be obtained using treatment and control samples. The availability of multiple outcomes, as in our application, yields additional degrees of over-identification for estimating the measurement error variances.[14]

Factor Loadings and Scale Parameters. Finally, we discuss options for estimating factor loadings and scale parameters in equation (3). First, one can consider SV-SIMEX estimates

[14]If the instrumental variability induced by $\boldsymbol{Z}$ satisfies the conditions in Schennach (2008), the quantile function $Q_{Y_d}(\tau|\boldsymbol{\Theta} = \boldsymbol{\theta})$, which we estimate using SV-SIMEX, is non-parametrically identified for $d \in \{0, 1\}$.

of $\boldsymbol{\lambda}_d\left(Q_{U_d}(\tau)\right)$ and $Q_{U_d}(\tau)$ separately for participants ($D=1$) and non-participants ($D=0$), and estimate the following regression:

$$\hat{\lambda}_{ld}\left(Q_{U_d}(\tau)\right) = \alpha_{0,ld} + \alpha_{1,ld}\hat{Q}_{U_d}(\tau) + \nu_{ld}, \tag{13}$$

where ν_{ld} is an error term, $\hat{\lambda}_{ld}\left(Q_{U_d}(\tau)\right)$ is the SV-SIMEX estimate associated with the l-th factor and $l = 1, \ldots, r$. Slope coefficients from these regressions will yield an estimate of $\boldsymbol{\gamma}_d$, while intercepts will estimate $\boldsymbol{\lambda}_d$.

When additional variables $\boldsymbol{Z}$ are available, estimates of $\boldsymbol{\lambda}_d$ can also be obtained using 2SLS regressions under Assumption 4. This case may be particularly relevant in empirical applications. As SV-SIMEX is valid for a general class of factor models, finding that intercepts retrieved from (13) are similar to factor loadings estimated by 2SLS methods suggests that a linear factor structure is a good fit to the data. We show in the Online Appendix that $\boldsymbol{\gamma}_d$ can also be identified using higher order moments under Assumption 1, Assumption 2 and Assumption 4.

7. Simulation Study

We consider the case of potential outcomes generated from equation (3) with one latent factor: $Y_d = \lambda_d\Theta + (1+\gamma_d\Theta)U_d$. Factor loadings in the potential outcome equations are $\lambda_0 = 0.75$ and $\lambda_1 = 1$. The values of γ_0 and γ_1 were chosen to define alternative scenarios for the excess of variance induced by heteroskedasticity. For $d \in \{0,1\}$, this quantity is equal to $1+\gamma_d^2$ and is defined as the ratio of variance of $(1+\gamma_d\Theta)U_d$ to variance of U_d. We consider two scenarios for small (2.25%, $\gamma_0 = 0.15$) and moderate (9%, $\gamma_1 = 0.30$) excess of variance. The variables U_0 and U_1 are distributed as χ^2_3 standardized to satisfy Assumption 1. Moreover, we consider $Z = 0.75\Theta + (1+0.15\Theta)V_2$, where V_2 is distributed as χ^2_3, and standardized to have mean zero and unit variance.

Benchmark results are obtained for the measurement $W = \Theta + V_1$, where Θ is a $\mathcal{N}(0,1)$ random variable and the measurement error V_1 is distributed as $\mathcal{N}(0,\sigma^2)$. The parameter σ^2 is determined by the noise to signal ratio (variance of V_1 over variance of W). We consider two scenarios corresponding to values 5% and 10% of this ratio. We additionally investigate the case of V_1 distributed as Skew Normal, a generalization of the Normal that allows for non-zero skewness. We parameterize this distribution to be heavily right-skewed, with zero mean and variance σ^2 (we set the shape parameter of this distribution to 8, which corresponds to the value 0.934 for the skewness coefficient). Results presented in what follows are from 400 simulated samples of size $2,500$. Simulations considering samples of 500 observations are presented in the Online Appendix, and convey conclusions similar to the ones shown here.

SV-SIMEX estimation of quantile specific intercepts and slopes performs extremely well. Its bias is well below 2% at most quantiles for all variants considered. The attenuation effect of measurement error conjectured in (11) is also evident. Different distributions for measurement error V_1 do not yield significantly different results although, as expected, performance deteriorates when the noise increases. Specifically, Table 3 presents simulation results for SV-SIMEX estimates of $Q_{U_0}(\tau)$ (Panel A) and $\lambda_0(\tau)$ (Panel B) at different quantiles ranging from 0.1 to 0.9. Simulations results for $Q_{U_1}(\tau)$ and $\lambda_1(\tau)$ convey a similar message, and are omitted for brevity. The results for Normal measurement error are presented in columns (2) to (5), and the results for Skew Normal errors are in columns (6) to (9). The estimator bias is defined as the difference between the average of SV-SIMEX estimates across the 400 samples and the true value of the quantity being estimated, which is reported at the left. Standard errors are defined as the standard deviation of SV-SIMEX estimates across the 400 samples. We set $B = 50$ at each value of ρ, starting from the true variance of V_1. The extrapolation step is obtained from a quadratic polynomial in ρ, considering 20 equally spaced values in the interval $[0.1, 2]$ (as in Figure 2).

The performance of the QCD estimator is also satisfactory, although distributional assumptions here become marginally more important. This can be seen from Panel C of Table 3, where we compute the quantity $Q_{Y_1-Y_0}(\tau|\Theta = \theta, U_0 = u_0)$ at values $\theta = \Phi^{-1}(0.25)$ and $u_0 = Q_{U_0}(0.50)$ using the estimated quantile intercept. The largest bias in this panel is 2.5%, which is observed for extreme quantiles and in the case with more measurement error as seen in column (8). Values of QCD are precisely estimated across quantiles. We therefore conclude that SV-SIMEX has very good properties at the values of the measurement error variance considered, which are in line with the amount of error in the empirical application with IHDP data below.

We also investigate the performance of our method in a two-factor model, and in models with U_d distributed as t_5 or a centered χ^2_3. Results in the Online Appendix confirm the good properties of SV-SIMEX documented here. We additionally compare the performance of our estimator with the corrected-loss estimator for Laplace measurement error (CLL) in Wang et al. (2012). Results in the Online Appendix show that SV-SIMEX outperforms CLL in terms of root mean squared error. This is expected because the measurement error variance is small and our approach does not require correction of the quantile loss function by nonparametric methods. We conclude that, while CLL can be applied to a larger class of models, SV-SIMEX might work better in the case of a factor model and for small measurement error variances.

8. QCDs of Low Birth Weight and Premature Infants

Factor Loadings and Measurement Errors. We assume that potential outcomes are correlated because of their dependence on *one* common factor. We consider four determinants of neonatal health to proxy this factor. Specifically, we use measurements on weight, length, head circumference, and gestational age for which descriptive statistics were reported in Table 1. The first principal component constructed from these measurements explains 85% of their correlation matrix, bolstering the idea that they can be considered manifestations of the same latent dimension. We find that the average of the first principal component for HLBW infants is about 1.4σ larger than in the LLBW group. Moreover, the first principal component has large positive associations with all anthropometrics, suggesting that the latent factor can be interpreted primarily as a proxy for health endowment at birth.

We adapt the strategy in Section 6 to estimate factor loadings to the case of IHDP. In particular, we consider each measurement W_j in turn on the left hand side of equation (5):

$$W_j = \Theta + V_j,$$

where the index $j = 1, \ldots, 4$ varies across the four measurements and V_j is a measurement-specific error. We additionally define the following potential outcome equations for cognitive development (Y_d^{CD}), morbidity (Y_d^{MB}) and general health (Y_d^{GH}):

$$Y_d^{CD} = \lambda_d^{CD}\Theta + \left(1 + \gamma_d^{CD}\Theta\right) U_d^{CD},$$
$$Y_d^{MB} = \lambda_d^{MB}\Theta + \left(1 + \gamma_d^{MB}\Theta\right) U_d^{MB},$$
$$Y_d^{GH} = \lambda_d^{GH}\Theta + \left(1 + \gamma_d^{GH}\Theta\right) U_d^{GH},$$

where $d \in \{0, 1\}$. Factor loadings are estimated from outcome-specific regressions using the measurement W_j to proxy for Θ. Specifically, the parameters λ_1^{CD} and λ_0^{CD} are obtained from 2SLS regressions in the IHDP-treatment sample ($D = 1$) and the IHDP-control sample ($D = 0$), respectively, of cognitive development on W_j instrumented with the remaining three measurements W_s, $s \neq j$.[15] Factor loadings for morbidity and the general health index are obtained in the same way. This procedure yields 2 (treatment-control) times 3 (the outcomes considered) estimates of factor loadings, which are presented in columns (1) to (6) of Table 4. Panels in this table are organized by possible choices for the measurement W_j.

[15] For brevity, we will often refer to children in the IHDP-treatment sample as 'participants'. Similarly, children in the IHDP-control sample will be called 'non-participants'. 2SLS estimation is carried out after taking residuals from regressions of all outcomes and measurements on randomization strata and mother demographics (see Panel B of Table 1), separately for treatment and control groups. Here and in what follows, residualized outcomes and measurements are standardized to have zero mean and unit variance in the sample.

For example, Panel A reports estimates of factor loadings using gestational age for W_j which is instrumented with weight, length, and head circumference.

Estimates of factor loadings are robust to the measurement chosen to proxy for the latent factor. This can be seen by looking at the values in the same column across panels of Table 4. The over-identification p-values associated with the estimates reported in each panel are, in general, far from conventional significance levels. Our results show that randomization to IHDP generally affected outcomes by changing dependence with health at birth. More precisely, we find that the estimated λ_1^{CD} is at least two times larger than estimated λ_0^{CD}; a similar comment applies to the relationship between λ_1^{GH} and λ_0^{GH}. On the other hand, we do not detect important treatment-control differences using the morbidity index.

SV-SIMEX Loadings and Scale Parameters. We use the measurement error variance estimated from the factor model to initialize SV-SIMEX. Specifically, we use birth weight to proxy for Θ because this measurement carries the lowest error among the measurements considered. This can be seen from column (7) of Table 4, where reported are GMM estimates of the noise-to-signal ratio $\sigma^2_{V_j}/\sigma^2_{W_j}$, where $\sigma^2_{V_j}$ is the variance of V_j and $\sigma^2_{W_j}$ is the variance of W_j for $j = 1, \ldots, 4$.[16] The simulation step of SV-SIMEX is obtained from naive QR regressions of outcomes on birth weight using the error variance in Panel C of Table 4 and $B = 50$. The extrapolation step uses a quadratic polynomial over 20 equally spaced values for ρ in the interval $[0.1, 2]$, as in Figure 2.

We find that quantile slopes are well approximated by the factor model (3). Specifically, dots in Figure 3 represent "unconstrained" (SV-SIMEX) estimates of $\lambda_1\left(Q_{U_1}(\tau)\right)$ and $\lambda_0\left(Q_{U_0}(\tau)\right)$ at different quantiles τ. These estimates are obtained by extrapolating τ-th slopes for participants ($D = 1$) and non-participants ($D = 0$), respectively, and do not impose the relationship in equation (4). Using SV-SIMEX slopes and intercepts, we estimate equation (13) to obtain estimates of λ_d and γ_d. The estimates of the factor loading, corresponding to the intercept of (13), are presented in columns (1) and (2) of Table 5. These estimates are remarkably similar to the estimates of factor loadings reported in Table 4.

This finding prompts the use of the estimated slopes in (13) to estimate factor scale parameters, which are reported in columns (4) and (5) of Table 5 along with their difference $\hat{\gamma}_1 - \hat{\gamma}_0$ in the last column. The lines in Figure 3 are obtained as $\hat{\lambda}_d + \hat{\gamma}_d \hat{Q}_{U_d}(\tau)$ for the three outcomes considered. Sharper treatment-control differences in slopes emerge for cognitive development.

[16] This quantity is obtained by staking 2×3 equations like (12) defined from estimates of factor loadings in each panel, and imposing one common value for the measurement error variance. As measurements are standardized to have unit variance in the sample, the noise-to-signal ratio is $\sigma^2_{V_j}$.

IHDP affects both location and scale of outcome distributions depending on a child's health at birth, as seen in Table 5, and this finding invalidates the traditional factor model. Loadings and scale parameters in columns (2) and (5) imply that, without IHDP, as one moves across the birth endowment distribution, outcomes should become more concentrated around better values. Treatment-control differences in column (3) imply positive treatment effects on outcome averages for all children. However, IHDP participation has different effects on the distribution variance depending on the outcome, as shown in column (6). For example, a value of -0.052 for the difference $\gamma_1 - \gamma_0$ for IQ implies that the intervention changed the shape of the marginal distributions, and thus, not all children with the same health endowment at birth benefited from the intervention in the same way. This result also implies that rank similarity conditional on birth endowment is violated.

We do not find significant differences in the uniqueness distributions of IHDP participants and controls. This can be seen in Figure 4, which reports estimates of $Q_{U_1}(\tau)$ and $Q_{U_0}(\tau)$ for each of the outcomes along with bootstrap confidence intervals. These estimates are obtained by extrapolation of τ-th intercepts from regressions in the samples of participants ($D = 1$) and non-participants ($D = 0$), respectively.

Quantiles of the Conditional Distribution of Treatment Effects. QCD is evaluated at the value $u_0 = Q_{U_0}(0.5)$ obtained from SV-SIMEX estimates of U_0. The latter quantity is small and close to zero, as shown in Figure 4. We consider values of the non-participation outcome corresponding to $\theta \in \{\theta_{LLBW}, \theta_{HLBW}\}$. Specifically, θ_{LLBW} is a prediction from a linear regression of birth weight on the outcome in the sample of LLBW non-participants (as explained in Section 6). Separate regressions are considered for the three outcomes, and predictions are computed at the outcome average. The value θ_{HLBW} is constructed in a similar manner using HLBW non-participants, and we find $\theta_{LLBW} < \theta_{HLBW}$. It follows that HLBW participants would have scored better outcomes than LLBW participants had IHDP not been implemented.

We start by considering IQ gains, as this dimension shows consistently large average effects in many early childhood interventions similar to IHDP (see Elango et al., 2016, and Table 2, above).[17] We find that a positive change in neonatal health affects location and scale of QCDs. This can be seen from the first panel of Figure 5 where dashed and continuous lines represent estimates for HLBW and LLBW infants, respectively. Horizontal lines here denote average effects, and shaded areas are 95% point-wise bootstrap confidence intervals. Infants with better neonatal health present a gain distribution with a lower tail shifted towards

[17]However, we note that the literature has documented fading average effects on IQ starting from kindergarten years.

larger values. Differences between QCDs tend to disappear as we go across quantiles (i.e., 0.68σ at the 0.10 quantile and 0.07σ at the 0.90 quantile). Gains are, therefore, more widely distributed for participants who would have been worse off without IHDP. For example, the inter-quantile range for LLBW infants is about $0.6 + 1 = 1.6\sigma$ compared to a value of $0.8+0.3 = 1.1\sigma$ for HLBW infants. Moreover, estimates in the first panel imply positive gains for about 65% of HLBW participants. This quantity is much lower for LLBW participants, at about 45%.

Better neonatal health also shifts towards larger values QCDs for health-related outcomes, as shown in the remaining panels of Figure 5 (recall that lower values of the morbidity index are a desirable outcome). However, health in the first three years of life shows smaller differences between LLBW and HLBW QCDs compared to IQ, which appear to change slowly across quantiles. For example, the difference between curves in the third panel is about 0.37σ at the 0.10 quantile, and 0.12σ at the 0.90 quantile. These differences are marginally significant for the general health index in the third panel, but not significantly different from zero in the second panel.

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.[18] The four panels in Figure 6 show how QCD estimates at selected quantiles (0.20, 0.40, 0.60, and 0.80) vary with neonatal health Θ. Only IQ gains are considered in this figure, where values on the horizontal axis are the 1 to 99 percent range for predictions of latent health from a regression of birth weight on IQ for non-participants. IHDP treatment effects increase with neonatal health at all quantiles and for both income groups. The estimated effects at the 0.60 and 0.80 quantiles are positive, uniformly over the health spectrum. We also find that treatment effects tend to be larger for lower income children in all panels. However, it is also clear that the limited size of treatment and control samples somewhat affects the precision of our analysis.

Treatment effect differences between income groups grow increasingly larger at the top end of the neonatal health distribution. For example, the last panel of Figure 6 shows a value of about 0.6σ for the 0.80 quantile of QCD if one considers the healthiest children of the higher income group (e.g., Θ above 0.5). At the same quantile, QCD is approximately twice as large for comparable children of the lower income group. The interaction between

[18]About 10% of observations with missing income information are discarded from the analysis. Duncan and Sojourner (2013) consider a different definition of income groups, which also follows from the multiple imputation of missing values. Duncan and Sojourner (2013) estimate average differences by income using weights defined from the Early Childhood Longitudinal Study - Birth Cohort data.

income and health endowment at birth is also evident in the remaining panels, although the health gradient fades away in the bottom two quintiles (as shown in the first two panels of the figure).

Within income-group variability in treatment effects is at least as large in magnitude as between income-group differences in treatment effects. Past work on IHDP data has documented larger effects, on *average*, for HLBW children from low-income families compared to HLBW children from high-income families (see, for example, Table 3 in Duncan and Sojourner, 2013). In our analysis too, an OLS regression for HLBW children of IQ on the IHDP treatment assignment dummy, a dummy for the lower-income group, and their interaction yields a coefficient of about 0.26σ on the latter variable (estimation results are not presented for brevity). If one considers the 'healthiest' children in Figure 6, the $0.80 - 0.20$ quantile range for $\Theta = 0.5$ is about $0.68 - (-0.75) = 1.43\sigma$ in high-income families and $0.99 - (-0.59) = 1.58\sigma$ in low-income families. These numbers suggest that variability in IHDP effects within groups can be much larger than variability between groups.

Although children in the lower-income group benefit more from IHDP on average, about 20% of children in the higher-income group have gains larger than the median gain in the lower-income group. This can be seen by considering the effect at $\Theta = 0.5$ in the fourth panel of Figure 6, which is about 0.6 for the higher-income group, and by noticing that the median gain for the lower-income group at $\Theta = 0.5$ must be below 0.5, which is the 0.60 quantile for the low-income group in the third panel. Moreover, our results also imply that larger average effects for the healthiest children in the lower-income group are entirely driven by large returns for a minority of children in this group. This can be seen by noticing that the differences in panels of Figure 6 become more important for large values of Θ only between 0.60 and the 0.80 quantile. At the same time, income differences among participants are somewhat irrelevant for children with the worst health conditions (most likely LLBW children here).

Given the particular population targeted by IHPD, infants at the top end of panels in Figure 6 are the most similar to normal birth-weight infants. This fact corroborates the external validity of our conclusions, and the beneficial effects of IHDP-like interventions to close income-based gaps in IQ and health development in young ages.

9. Conclusion

The policy relevance of distributional treatment parameters has been discussed in past work, starting from Heckman et al. (1997). A number of important policy questions arise beyond the simple difference between outcome averages for participants and non-participants.

Quantile treatment effects have been extensively considered in empirical work although, in general, they are not informative about how widely individual gains from participation are distributed across units.

We have revisited this problem by considering the case of IHDP, a small-scale early childhood program implemented in the 1980s which targeted low birth-weight infants. A growing body of evidence demonstrates the beneficial effects of these programs, but IHDP has received scant attention to date. Understanding the number of children with positive gains and characterizing which demographics are the best predictors of these gains, are crucial inputs for cost-effective policy targeting to mitigate the effects of adverse early childhood conditions that exacerbate inequalities over a lifetime.

We have proposed a simple approach to learn about treatment effect distributions. Our approach connects with the literature on factor models, a tool that has become increasingly more common in applied labor and education work. The setting we have considered here is the one faced by researchers in those fields. We have assumed that the dependence between potential outcome distributions is driven by latent factors. This idea in not new and goes back to work by Carneiro et al. (2003), among others. We studied identification of QCD, and we showed that common distributional treatment parameters can be derived from knowledge of this parameter. Moreover, we have proposed a method to estimate this new parameter by offering a simple algorithm for models when the measurement error variance is small. This approach does not require the use of instrumental variables or repeated measurements. We believe that small measurement error is reasonable in many empirical settings, because economic theory or the context under investigation can be informative about relationships between proxies and latent factors. The literature on early childhood interventions is an ideal example because of its important contributions on measurement (see Conti et al., 2018, among others).

Our findings raise a number of additional questions, including consistent estimation of more general models and inference. It is important to investigate a generalization of the proposed approach for cases when the error-measurement variance is large, although practitioners can estimate QCDs by employing the corrected-loss function estimator proposed by Wang et al. (2012). Moreover, statistical inference for the QCD imposes several challenges and it remains to be fully investigated. The investigation of models with a larger number of unknown factors is also important. We hope to address some of these questions in future work.

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (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., and J. J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics, Vol6, Part B*, chap. 72, pp. 5145–5303. North-Holland, Amsterdam.

Almond, D., and J. Currie (2011): "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 25(3), 153–72.

Andrews, R. J., J. Li, and M. F. Lovenheim (2016): "Quantile Treatment Effects of College Quality on Earnings," *Journal of Human Resources*, 51(1), 200–238.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74(2), 539–563.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (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., and A. Chesher (2014): "Treatment effect estimation with covariate measurement error," *Journal of Econometrics*, 178(2), 707–715.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988–1012.

——— (2017): "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment," *The Review of Economics and Statistics*, 99(4), 683–697.

Callaway, B. (2019): "Bounds on Distributional Treatment Effect Parameters using Panel Data with an Application on Job Displacement," `https://bcallaway11.github.io/files/DTE/dte1.pdf`.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2019): "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs1," *The Econometrics Journal,* `10.1093/ectj/utz022`.

Carneiro, P. M., K. Karsten, and J. J. Heckman (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., D. Ruppert, L. A. Stefanski, and C. Crainiceanu (2006): *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC Press, second edn.

Chaparro, J., and A. J. Sojourner (2019): "Differential effects from access to high-quality early care," in *Sustaining Early Childhood Learning Gains*, ed. by A. Reynolds, and J. Temple. Cambridge.

Chen, L., J. J. Dolado, and J. Gonzalo (2019): "Quantile Factor Models," UC3M Working Papers, Universidad Carlos III de Madrid.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49(4), 90137.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

Chesher, A. (2017): "Understanding the effect of measurement error on quantile regressions," *Journal of Econometrics*, 200(2), 223 – 237.

Conti, G., M. Hanson, H. M. Inskip, S. Crozier, C. Cooper, and K. Godfrey (2018): "Beyond birth weight: the origins of human capital," IFS Working Paper W18/30.

Cook, J. R., and L. A. Stefanski (1994): "Simulation-Extrapolation Estimation in Parametric Measurement Error Models," *Journal of the American Statistical Association*, 89(428), 1314–1328.

Cooley Fruehwirth, J., S. Navarro, and Y. Takahashi (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., J. J. Heckman, and S. M. Schennach (2010): "Estimating the technology of cognitive and noncognitive skills formation," *Econometrica*, 78(3), 883931.

Currie, J., and M. Rossin-Slater (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., and S. Shen (2018): "Testing for Rank Invariance or Similarity in Program Evaluation," *The Review of Economics and Statistics*, 100(1), 78–85.

Duncan, G. J., and A. J. Sojourner (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., J. L. García, J. J. Heckman, and A. Hojman (2016): *Early Childhood Education*pp. 235–297. University of Chicago Press.

Fan, Y., E. Guerre, and D. Zhu (2017): "Partial identification of functionals of the joint distribution of potential outcomes," *Journal of Econometrics*, 197, 42–59.

Fan, Y., and S. S. Park (2010): "Sharp Bounds on the Distribution of Treatment Effects and Their Statistical Inference," *Econometric Theory*, 26(3), 931–951.

Firpo, S., A. F. Galvao, and S. Song (2017): "Measurement errors in quantile regression models," *Journal of Econometrics*, 198(1), 146 – 164.

Firpo, S., and G. Ridder (2019): "Partial identification of the treatment effect distribution and its functionals," *Journal of Econometrics*, 213, 210–234.

Frandsen, B. R., and L. J. Lefgren (2018a): "Partial Identification of the Distribution of Treatment Effects with an Application to the Knowledge Is Power Program (KIPP)," Working Paper 24616, National Bureau of Economic Research.

Frandsen, B. R., and L. J. Lefgren (2018b): "Testing Rank Similarity," *The Review of Economics and Statistics*, 100(1), 86–91.

Gross, R. T., D. Spiker, and C. W. Haynes (1997): *Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program*. Stanford University Press.

Heckman, J., S. H. Moon, R. Pinto, P. Savelyev, and A. Yavitz (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. (2020): "Randomization and Social Policy Evaluation Revisited," Discussion Paper 12882, IZA.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64, 487–535.

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Journal of Economic Literature*, 76(1), 195216.

IHDP (1990): "Enhancing the Outcomes of Low-Birth-Weight, Premature Infants: A Multisite, Randomized Trial," *JAMA*, 263(22), 3035–3042.

Ko, H., R. E. Howland, and S. A. Glied (2020): "The Effects of Income on Childrens Health: Evidence from Supplemental Security Income Eligibility under New York State Medicaid," Working Paper 26639, National Bureau of Economic Research.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press, New York, New York.

Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20(1), 69–76.

Lewbel, A. (2019): "The Identification Zoo: Meanings of Identification in Econometrics," *Journal of Economic Literature*, 57(4), 835–903.

Li, T., and Q. Vuong (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65(2), 139 – 165.

Makarov, G. D. (1981): "Estimates for the Distribution Function of a Sum of Two Random Variables When the Marginal Distributions are Fixed," *Theory of Probability & Its Applications*, 26(4), 803–806.

Manski, C. F. (1997): "Monotone Treatment Response," *Econometrica*, 65(6), 1311–1334.

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.

Schennach, S. M. (2008): "Quantile Regression with Mismeasured Covariates," *Econometric Theory*, 24, 1010–1043.

Shang, Y. (2012): "Measurement Error Adjustment Using the SIMEX Method: An Application to Student Growth Percentiles," *Journal of Educational Measurement*, 49(4), 446–465.

Stefanski, L. A., and J. M. Bay (1996): "Simulation Extrapolation Deconvolution of Finite Population Cumulative Distribution Function Estimators," *Biometrika*, 83(2), 407–417.

Torres-Saavedra, P. A. (2013): "Quantile Regression for Repeated Responses Measured with Error," Department of Statistics, North Carolina State University.

Wang, H. J., L. A. Stefanski, and Z. Zhu (2012): "Corrected-loss estimation for quantile regression with covariate measurement errors," *Biometrika*, 99(2), 405–421.

Wang, X.-F., Z. Fan, and B. Wang (2010): "Estimating smooth distribution function in the presence of heteroscedastic measurement errors," *Computational Statistics Data Analysis*, 54(1), 25 – 36.

Wei, Y., and R. J. Carroll (2009): "Quantile Regression With Measurement Error," *Journal of the American Statistical Association*, 104.

Appendix A. Partial Quantile Regression with Measurement Error

The notation in Section 5 is employed here. Let $\boldsymbol{\epsilon}$ be a $r \times 1$ random variable that is jointly distributed as F_V with mean 0 and variance $\boldsymbol{\Sigma}$. Based on equation (5), define:

$$\tilde{\boldsymbol{W}} = \boldsymbol{W} + \sqrt{\rho}\boldsymbol{\epsilon} = \boldsymbol{\Theta} + \boldsymbol{V} + \sqrt{\rho}\boldsymbol{\epsilon},$$

where $\rho > 0$. Under these assumptions we have that:

$$\tilde{\boldsymbol{W}} = \boldsymbol{\Theta} + ((1+\rho)\boldsymbol{\Sigma})^{1/2}\tilde{\boldsymbol{V}}, \tag{A.1}$$

where $\tilde{\boldsymbol{V}} \sim F_V(0,1)$. Solving for $\boldsymbol{\Theta}$ in equation (A.1) and replacing it in the outcome equation we obtain:

$$\begin{aligned} Y_d &= \boldsymbol{\lambda}_d(U_d)[\tilde{\boldsymbol{W}} - ((1+\rho)\boldsymbol{\Sigma})^{1/2}\tilde{\boldsymbol{V}}] + U_d, \\ &= \boldsymbol{\phi}_d(U_d)'\tilde{\boldsymbol{B}} - \boldsymbol{\pi}_d(U_d)'\tilde{\boldsymbol{V}}. \end{aligned}$$

Recall that $\tilde{\boldsymbol{V}}$ is the omitted variable.

Let $\varrho_\tau(u) = u(\tau - I(u<0))$ be the quantile regression loss function. The feasible quantile regression problem solves:

$$\tilde{\boldsymbol{\phi}}_d(\tau) = \operatorname{argmin} E\left[\varrho_\tau\left(Y_d - \boldsymbol{\phi}_d'\tilde{\boldsymbol{B}})\right)\right],$$

while the unfeasible version of the problem finds the following coefficient vector:

$$(\boldsymbol{\phi}_d(\tau)', \boldsymbol{\pi}_d(\tau)') = \operatorname{argmin} E\left[\varrho_\tau\left(Y_d - \boldsymbol{\phi}_d'\tilde{\boldsymbol{B}} - \boldsymbol{\pi}_d'\tilde{\boldsymbol{V}}\right)\right].$$

Theorem 1 in Angrist et al. (2006) shows that the quantile regression solves a weighted least squares problem. The solution depends on weights that can be approximated as:

$$w_\tau(\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{\phi}}_d) = \frac{1}{2} f_{Y_d}\left(Q_{Y_d}(\tau|\boldsymbol{\theta})|\boldsymbol{\theta}\right) + \varrho_\tau(\bar{\boldsymbol{W}}, \tilde{\boldsymbol{\phi}}_d), \tag{A.2}$$

where:

$$\varrho_\tau(\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{\phi}}_d) \leq \frac{1}{6}\left|\tilde{\boldsymbol{W}}'(\tilde{\boldsymbol{\phi}}_d - \boldsymbol{\phi}_d) - \boldsymbol{\pi}_d'\tilde{\boldsymbol{V}}\right| \bar{f}'(\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{\phi}}_d),$$

and $\bar{f}'$ denotes the first derivative with respect to Y_d. The first term on the right hand side of (A.2) is equal to:

$$f_{Y_d}\left(Q_{Y_d}(\tau|\boldsymbol{\theta})|\boldsymbol{\theta}\right) = \left[(1 + \boldsymbol{\gamma}_d'\boldsymbol{\theta})\,\frac{\partial}{\partial\tau}Q_{U_d}(\tau)\right]^{-1},$$

implying that the density is inversely proportional to the standard deviation conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$. We denote the term inside the brackets in the last expression as $\sigma(\boldsymbol{\theta})$. It follows

that:

$$w_\tau(\bar{\boldsymbol{W}}, \tilde{\boldsymbol{\phi}}_d) \approx \frac{1}{2} f_{Y_d}\left(Q_{Y_d}(\tau|\boldsymbol{\theta})|\boldsymbol{\theta}\right) = \frac{1}{2}\sigma(\boldsymbol{\theta})^{-1}.$$

Following closely Angrist et al.'s (2006) derivations for the omitted variable bias in quantile regression we obtain:

$$\begin{aligned}
\tilde{\boldsymbol{\phi}}_d(\tau) &\approx \boldsymbol{\phi}_d(\tau) - \left(E\left[\sigma(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}'\right]\right)^{-1} E\left[\sigma(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{V}}'\boldsymbol{\pi}_d(\tau)\right], \\
&\approx \boldsymbol{\phi}_d(\tau) - \left(E\left[\sigma(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{B}}'\right]\right)^{-1} E\left[\sigma(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{V}}'\left(\boldsymbol{\lambda}_d(\tau)'((1+\rho)\boldsymbol{\Sigma})^{1/2}\right)\right].
\end{aligned}$$

As long as $\rho \to -1$, the remainder term $\varrho_\tau(\tilde{\boldsymbol{W}}, \tilde{\boldsymbol{\phi}}_d) \to 0$ and:

$$E\left[\sigma(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{V}}'\left(\boldsymbol{\lambda}_d(\tau)'((1+\rho)\boldsymbol{\Sigma})^{1/2}\right)\right] \to 0.$$

It follows that $\tilde{\boldsymbol{\phi}}_d(\tau) \to \boldsymbol{\phi}_d(\tau)$ as $\rho \to -1$.

Appendix B. Approximate Effects for Small Error Variance

Equation (6) in Chesher (2017) in the case of one latent factor under our assumptions yields (the index d is omitted throughout for simplicity):

$$\begin{aligned}
Q_Y(\tau|W=w) &= Q_Y(\tau|\Theta=w) + \sigma^2\Big[Q_Y^\theta(\tau|\Theta=w)g_\Theta^\theta(w) + \frac{1}{2}Q_Y^{\theta\theta}(\tau|\Theta=w) \\
&- \frac{Q_Y^{\tau\theta}(\tau|\Theta=w)Q_Y^\theta(\tau|\Theta=w)}{Q_Y^\tau(\tau|\Theta=w)} + \frac{1}{2}\frac{Q_Y^{\tau\tau}(\tau|\Theta=w)(Q_Y^\theta(\tau|\Theta=w))^2}{(Q_Y^\tau(\tau|\Theta=w))^2}\Big] \\
&+ o(\sigma^2), \qquad \text{(B.1)}
\end{aligned}$$

where $W = \Theta + V$, σ^2 is the variance of V, and:

$$\begin{aligned}
Q_Y^\theta(\tau|\Theta=w) &= \frac{\partial}{\partial\theta}Q_Y(\tau|\Theta=w)\Big|_{\theta=w}, \quad g_\Theta^\theta(w) = \frac{\partial}{\partial\theta}\log(f_\Theta(w))\Big|_{\theta=w}, \\
Q_Y^{\theta\theta}(\tau|\Theta=w) &= \frac{\partial}{\partial\theta}Q_Y^\theta(\tau|\Theta=w), \quad Q_Y^\tau(\tau|\Theta=w) = \frac{\partial}{\partial\tau}Q_Y(\tau|\Theta=w)\Big|_{\theta=w}, \\
Q_Y^{\tau\tau}(\tau|\Theta=w) &= \frac{\partial}{\partial\tau}Q_Y^\tau(\tau|\Theta=w), \quad Q_Y^{\tau\theta}(\tau|\Theta=w) = \frac{\partial}{\partial\theta}Q_Y^\tau(\tau|\Theta=w).
\end{aligned}$$

The location-scale shift model in equation (3) yields the following expressions:

$$\begin{aligned}
Q_Y^\theta(\tau|\Theta=w) &= \lambda + \gamma Q_U(\tau) = \lambda(\tau), \\
Q_Y^{\theta\theta}(\tau|\Theta=w) &= 0, \\
Q_Y^\tau(\tau|\Theta=w) &= (1+\gamma w)Q_U^\tau(\tau), \\
Q_Y^{\tau\tau}(\tau|\Theta=w) &= (1+\gamma w)Q_U^{\tau\tau}(\tau), \\
Q_Y^{\tau\theta}(\tau|\Theta=w) &= \gamma Q_U^\tau(\tau),
\end{aligned}$$

where $Q_U(\tau)$ is the quantile of the uniqueness U, $Q_U^\tau(\tau) = \frac{\partial}{\partial\tau}Q_U(\tau)$, and $Q_U^{\tau\tau}(\tau) = \frac{\partial}{\partial\tau}Q_U^\tau(\tau)$. Replacing the derivatives of the quantile function in equation (B.1) we obtain:

$$\begin{aligned} Q_Y(\tau|W=w) &= Q_Y(\tau|\Theta=w) + \sigma^2\Bigg[\lambda(\tau)f_\Theta^\theta(w) - \frac{\gamma Q_U^\tau(\tau)\lambda(\tau)}{(1+\gamma w)Q_U^\tau(\tau)} \\ &\quad + \frac{1}{2}\frac{(1+\gamma w)Q_U^{\tau\tau}(\tau)(\lambda(\tau))^2}{((1+\gamma w)Q_U^\tau(\tau))^2}\Bigg] + o(\sigma^2). \end{aligned}$$

The error contaminated quantile function $Q_Y(\tau|W=w)$ can be approximated as follows:

$$\begin{aligned} Q_Y(\tau|W=w) &= Q_Y(\tau|\Theta=w) \\ &+ \sigma^2\lambda(\tau)\left[g_\Theta^\theta(w) - \frac{\gamma}{(1+\gamma w)} + \frac{1}{2}\frac{Q_U^{\tau\tau}(\tau)\lambda(\tau)}{(1+\gamma w)(Q_U^\tau(\tau))^2}\right] + o(\sigma^2). \quad \text{(B.2)} \end{aligned}$$

The small variance approximation can also be used for the quantiles of the latent factor Θ. Let $Q_V(\tau)$ be the quantile of V, $Q_V^\tau(\tau) = \frac{\partial}{\partial\tau}Q_V(\tau)$, and $Q_V^{\tau\tau}(\tau) = \frac{\partial}{\partial\tau}Q_V^\tau(\tau)$. Because $W = \Theta + V$, we can employ Chesher's (2017) equation (6) considering $\lambda(\tau) = \lambda = 1$ and $\gamma = 0$. Basic manipulations lead to:

$$Q_W(\tau) = Q_\Theta(\tau) + \sigma^2\left[g_\Theta^\theta(w) - \frac{1}{2}g_V^v(\tau)\right] + o(\sigma^2),$$

because,

$$\frac{Q_V^{\tau\tau}(\tau)}{Q_V^\tau(\tau)} = -\frac{\partial}{\partial v}\log(f_V(v))\Big|_{v=Q_V(\tau)} = -g_V^v(\tau).$$

Appendix C. Unknown measurement error variances

The conditioning on D is omitted throughout as the participation status is randomly assigned (Assumption 2). Equation for potential outcomes (3) implies for $d = 0, 1$:

$$E(Y_d|\boldsymbol{Z}=\boldsymbol{z}) = \boldsymbol{\lambda}_d' E(\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}) + \boldsymbol{\gamma}_d' E(\boldsymbol{\Theta}U_d|\boldsymbol{Z}=\boldsymbol{z}) + E(U_d|\boldsymbol{Z}=\boldsymbol{z}). \quad \text{(C.1)}$$

Use Assumption 4. The assumption $E(\boldsymbol{V}|\boldsymbol{Z}=\boldsymbol{z}) = E(\boldsymbol{V})$ implies:

$$E(\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}) = E(\boldsymbol{W}|\boldsymbol{Z}=\boldsymbol{z}),$$

since $E(\boldsymbol{V}) = \boldsymbol{0}$ from Assumption 1. The assumption $E(U_d|\boldsymbol{\Theta}=\boldsymbol{\theta}, \boldsymbol{Z}=\boldsymbol{z}) = E(U_d|\boldsymbol{\Theta}=\boldsymbol{\theta})$ implies for $d = 0, 1$:

$$E(\boldsymbol{\Theta}U_d|\boldsymbol{Z}=\boldsymbol{z}) = 0, \qquad E(U_d|\boldsymbol{Z}=\boldsymbol{z}) = 0,$$

because of Assumption 1. The relationships above substituted in (C.1) imply:

$$E(Y_d|\boldsymbol{Z}=\boldsymbol{z}) = \boldsymbol{\lambda}_d' E(\boldsymbol{W}|\boldsymbol{Z}=\boldsymbol{z}).$$

For $d = 0, 1$ this defines the 2SLS regression of Y_d on $\boldsymbol{W}$ in which the latter is instrumented with $\boldsymbol{Z}$.

Using (5) and Assumption 1, for $d = 0, 1$ we have:

$$\begin{aligned} E\left((\lambda_d'\boldsymbol{W} - Y_d)\frac{W_l}{\lambda_{ld}}\right) &= E\left((\boldsymbol{\lambda}_d'\boldsymbol{V} - \boldsymbol{\gamma}_d'\boldsymbol{\Theta}U_d - U_d)\frac{\Theta_l + V_l}{\lambda_{ld}}\right) \\ &= \sigma_l^2, \qquad l = 1, \ldots, r \end{aligned}$$

which defines two moment conditions (for treatment and control samples) for variances in $\boldsymbol{\Sigma}$.

TABLE 1. Balancing tests for random assignment of infants.

	All		**HLBW** (2,000–2,500 grams)		**LLBW** (Less than 2,000 grams)	
	Mean	Effect	Mean	Effect	Mean	Effect
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Demographics at Birth					
Weight (gm)	1.787	0.025	2.256	-0.001	1.520	0.041
	[0.465]	(0.019)	[0.138]	(0.015)	[0.363]	(0.029)
Length (cm)	42.259	0.485**	45.33	0.165	40.511	0.682**
	[3.966]	(0.206)	[2.208]	(0.247)	[3.672]	(0.293)
Body Mass Index	9.823	-0.034	11.046	-0.070	9.127	-0.017
	[1.602]	(0.088)	[1.310]	(0.149)	[1.311]	(0.107)
Gestational Age (weeks)	33.067	-0.061	34.961	-0.091	31.989	-0.015
	[2.719]	(0.148)	[1.431]	(0.168)	[2.690]	(0.213)
Neonatal Health Index	99.75	0.924	98.913	0.365	100.227	1.243
	[15.446]	(1.077)	[14.371]	(1.641)	[16.025]	(1.425)
Head Circumference (cm)	29.426	0.050	31.437	-0.168	28.282	0.183
	[2.553]	(0.132)	[1.348]	(0.141)	[2.363]	(0.192)
Infant is a boy	0.484	0.013	0.539	-0.052	0.453	0.048
	[0.500]	(0.034)	[0.500]	(0.056)	[0.498]	(0.043)
Infant is a first-born	0.430	0.032	0.393	0.077	0.450	0.008
	[0.495]	(0.033)	[0.490]	(0.054)	[0.498]	(0.042)
	Panel B. Mother Demographics					
Age	24.863	-0.214	24.99	-0.831	24.79	0.154
	[6.148]	(0.388)	[6.346]	(0.645)	[6.04]	(0.489)
Married	0.484	-0.057*	0.485	-0.035	0.483	-0.070*
	[0.500]	(0.032)	[0.501]	(0.054)	[0.500]	(0.040)
High School Graduate	0.491	-0.054	0.471	-0.105*	0.503	-0.028
	[0.500]	(0.034)	[0.500]	(0.055)	[0.501]	(0.042)
College Graduate	0.120	0.004	0.155	-0.034	0.099	0.028
	[0.325]	(0.021)	[0.363]	(0.035)	[0.300]	(0.027)
Black	0.526	-0.008	0.500	-0.074	0.541	0.029
	[0.500]	(0.030)	[0.501]	(0.048)	[0.499]	(0.038)
Hispanic	0.116	0.000	0.126	0.007	0.110	-0.003
	[0.321]	(0.019)	[0.333]	(0.033)	[0.314]	(0.024)
Observations		929		343		586

Note. Columns (1), (3), and (5) show control-group means and standard deviations (in square brackets) for variables listed at the left. Columns (2), (4), and (6) show treatment-control differences, with robust standard errors in parenthesis. Results in columns (3)-(6) are from regressions stratified for infants with birth weights lower (LLBW) or higher (HLBW) than 2,000 grams. All regressions control for randomization strata effects (site and birth weight group). The health index is defined from the residuals of a regression of gestational age on birth weight and site dummies. These residuals were centered at 100 and rescaled to have a standard deviation of 16. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 2. Average treatment effects after three years from the assignment.

	All Infants				**HLBW** (2,000–2,500 grams)				**LLBW** (Less than 2,000 grams)			
	Mean	Effect			Mean	Effect			Mean	Effect		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Cognitive Development												
Development Index	84.3 [20.1]	9.156*** (1.267)	9.388*** (1.150)	8.942*** (1.120)	84.8 [18.9]	13.588*** (1.967)	13.996*** (1.667)	13.476*** (1.681)	84.1 [20.7]	6.804*** (1.631)	7.238*** (1.512)	6.750*** (1.482)
Observations		855	855	855		310	310	310		545	545	545
Panel B. Health												
Morbidity Index	6.9 [3.1]	0.843*** (0.221)	0.885*** (0.218)	0.936*** (0.222)	6.7 [3.0]	0.627* (0.351)	0.624* (0.360)	0.639* (0.371)	6.9 [3.1]	0.963*** (0.284)	1.049*** (0.277)	1.106*** (0.281)
Observations		844	844	844		308	308	308		536	536	536
General Health Ratings Index	27.1 [5.1]	0.326 (0.368)	0.376 (0.367)	0.291 (0.368)	27.6 [4.7]	0.457 (0.578)	0.751 (0.573)	0.644 (0.591)	26.8 [5.2]	0.215 (0.477)	0.131 (0.476)	0.092 (0.476)
Observations		851	851	851		309	309	309		542	542	542

Note. Effects on cognitive development (Panel A) and health conditions (Panel B) after 36 months. 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 randomization strata effects (site and birth weight group). Coefficients in columns (3), (7), and (11) are from regressions with mother demographics (in Panel B of Table 1) 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). Results in columns (5)-(12) are from regressions stratified for infants with birth weights lower (LLBW) or higher (HLBW) than 2,000 grams. 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 three 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 3. Simulation results for intercepts, slopes, and QCDs.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Panel A: Intercept, Q_{U_0}							
		Normal errors				Skew-Normal errors			
		5% noise		10% noise		5% noise		10% noise	
True value		Bias	SE	Bias	SE	Bias	SE	Bias	SE
0.10	-0.986	0.003	0.014	0.004	0.017	0.011	0.014	0.021	0.017
0.25	-0.730	0.002	0.019	0.004	0.019	0.001	0.018	0.005	0.018
0.50	-0.259	-0.000	0.024	0.001	0.025	0.000	0.024	0.002	0.026
0.75	0.452	-0.001	0.040	0.001	0.041	0.000	0.038	0.001	0.041
0.90	1.327	0.000	0.062	0.002	0.065	-0.002	0.061	-0.003	0.062
		Panel B: Slope, $\lambda_0(\tau)$							
		Normal errors				Skew-Normal errors			
		5% noise		10% noise		5% noise		10% noise	
True value		Bias	SE	Bias	SE	Bias	SE	Bias	SE
0.10	0.602	-0.001	0.015	-0.002	0.017	-0.006	0.014	-0.015	0.016
0.25	0.641	-0.002	0.017	-0.005	0.018	-0.003	0.017	-0.007	0.017
0.50	0.711	-0.003	0.025	-0.007	0.026	-0.002	0.024	-0.001	0.025
0.75	0.818	-0.004	0.036	-0.008	0.038	-0.001	0.038	0.002	0.041
0.90	0.949	-0.006	0.062	-0.010	0.063	0.002	0.063	0.006	0.067
		Panel C: QCD, $Q_{Y_1-Y_0}(\tau\|\Theta=\theta,U_0=u_0)$							
		Normal errors				Skew-Normal errors			
		5% noise		10% noise		5% noise		10% noise	
True value		Bias	SE	Bias	SE	Bias	SE	Bias	SE
0.10	-0.723	-0.001	0.028	-0.011	0.030	0.016	0.027	0.025	0.029
0.25	-0.518	-0.004	0.029	-0.006	0.030	0.003	0.029	0.015	0.029
0.50	-0.142	-0.002	0.032	-0.001	0.033	-0.003	0.033	-0.004	0.035
0.75	0.425	-0.000	0.042	0.004	0.043	-0.006	0.040	-0.013	0.042
0.90	1.123	0.006	0.067	0.015	0.073	-0.004	0.063	-0.017	0.066

Note. The table reports bias and standard error of SV-SIMEX estimators for $Q_{U_0}(\tau)$ (Panel A), $\lambda_0(\tau)$ (Panel B), and $QCD(\tau;\theta,Q_{U_0}(0.5))$ at $\theta=\Phi^{-1}(0.25)$ (Panel C). True values at the quantile $\tau\in\{0.1,0.25,0.5,0.75,0.9\}$ are reported in column (1). The distribution of V_1 is Normal in columns (2) to (5), and (right) Skew Normal in columns (6) to (9). The variance of V_1 is either 5% or 10% of the variance of W_1. Group size (treatment and control) is $2,500$. Results from 200 replications are obtained setting $B=50$ and considering a 20-point grid for ρ between 0.10 and 2. See Section 7 for additional details.

TABLE 4. Factor loadings and measurement equations

	Cognitive Development		**Morbidity**		**General Health**		**Error Variance**
	Treatment	Control	Treatment	Control	Treatment	Control	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Gestational Age (weeks)							
Estimate	0.380***	0.111**	-0.180**	-0.123**	0.256***	0.116**	0.344
	(0.077)	(0.054)	(0.081)	(0.052)	(0.079)	(0.053)	
Sargan p-value	0.615	0.181	0.816	0.081	0.288	0.206	
Panel B: Length at Birth (cm)							
Estimate	0.347***	0.0922*	-0.167**	-0.120**	0.228***	0.099**	0.273
	(0.070)	(0.049)	(0.074)	(0.047)	(0.072)	(0.048)	
Sargan p-value	0.624	0.219	0.872	0.092	0.222	0.988	
Panel C: Birth Weight (gm)							
Estimate	0.285***	0.115**	-0.150**	-0.093**	0.185***	0.109**	0.051
	(0.062)	(0.046)	(0.068)	(0.045)	(0.065)	(0.046)	
Sargan p-value	0.905	0.922	0.837	0.138	0.553	0.340	
Panel D: Head Circumference at Birth (cm)							
Estimate	0.334***	0.095*	-0.155**	-0.129***	0.234***	0.105**	0.278
	(0.068)	(0.050)	(0.072)	(0.049)	(0.070)	(0.049)	
Sargan p-value	0.469	0.347	0.975	0.949	0.850	0.176	
Observations	332	523	331	513	330	521	

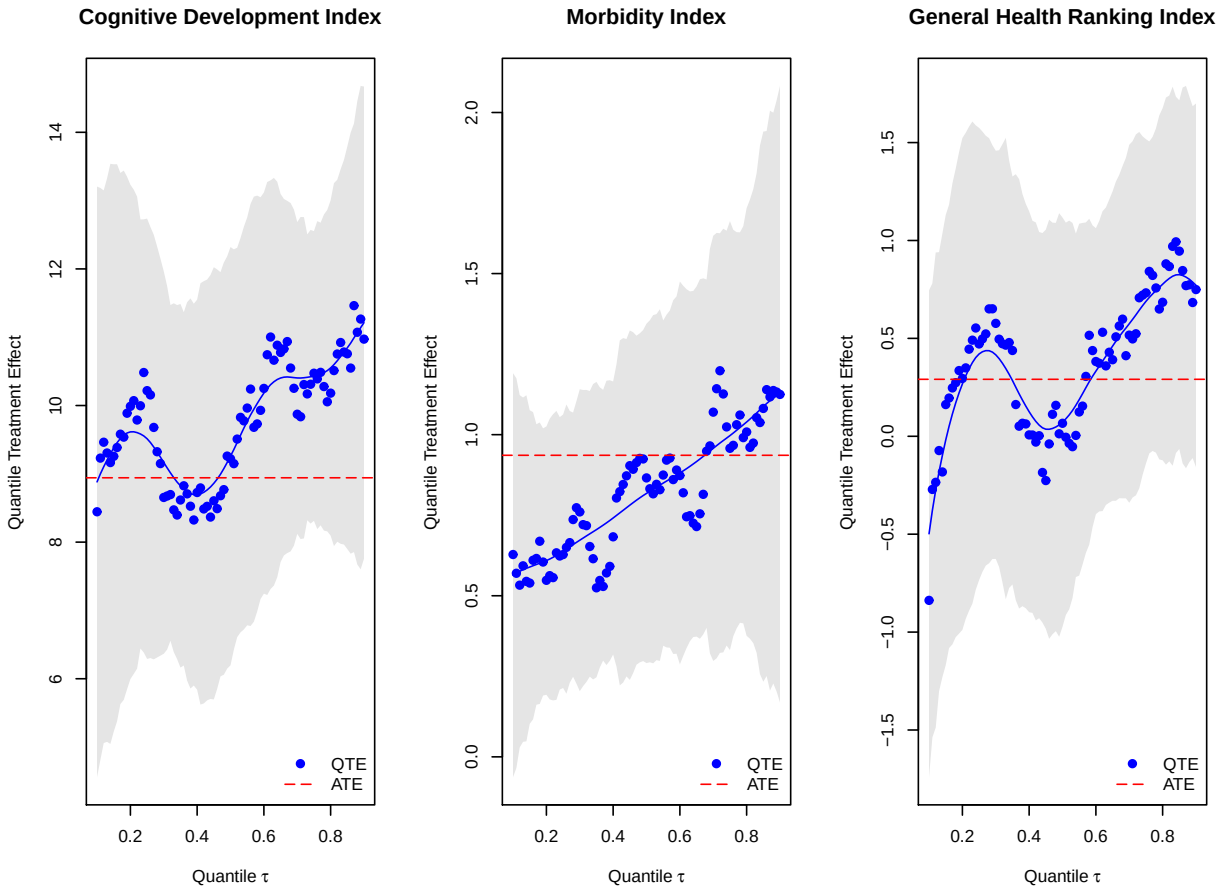
Note. Results from a system of seven equations: six equations describe potential outcomes after three years from the assignment, and one measurement equation is for the latent factor at birth. Outcomes considered: Cognitive Development Index, Morbidity Index, and General Health Ratings Index. Definitions for these outcomes are in Table 2. Estimation results using gestational age are in Panel A. All variables are residuals from separate regressions on randomization strata effects and mother demographics (see Panel B of Table 1). These residuals are standardized to have unit variance. Columns (1) to (6) of this panel show estimates of the factor loadings for the potential outcome equations. The error variance in the measurement equation using gestational age is in column (7), and can be interpreted as the noise to signal ratio because of the standardization. Loadings are estimated using 2SLS equations, where the excluded instruments are length at birth, birth weight, and head circumference at birth. P-values for over-identification tests are presented as well. The remaining panels have the same interpretation. In each panel, the excluded instruments are all the remaining measurements at birth (e.g., gestational age, birth weight, and head circumference at birth in Panel B). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 5. SV-SIMEX estimates of factor loadings (or, location parameters) and scale parameters

	Loadings			**Scale Parameter**		
	Treatment	Control	Difference	Treatment	Control	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Cognitive Development Index					
Estimate	0.297***	0.070***	0.227***	-0.135***	-0.083***	-0.052***
	(0.005)	(0.003)	(0.006)	(0.004)	(0.003)	(0.004)
	Panel B: Morbidity Index					
Estimate	-0.140***	-0.100***	-0.041***	-0.007*	-0.028***	0.021***
	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)
	Panel C: General Health Ratings Index					
Estimate	0.221***	0.086***	0.135***	-0.052***	-0.089***	0.037***
	(0.005)	(0.003)	(0.005)	(0.003)	(0.002)	(0.004)

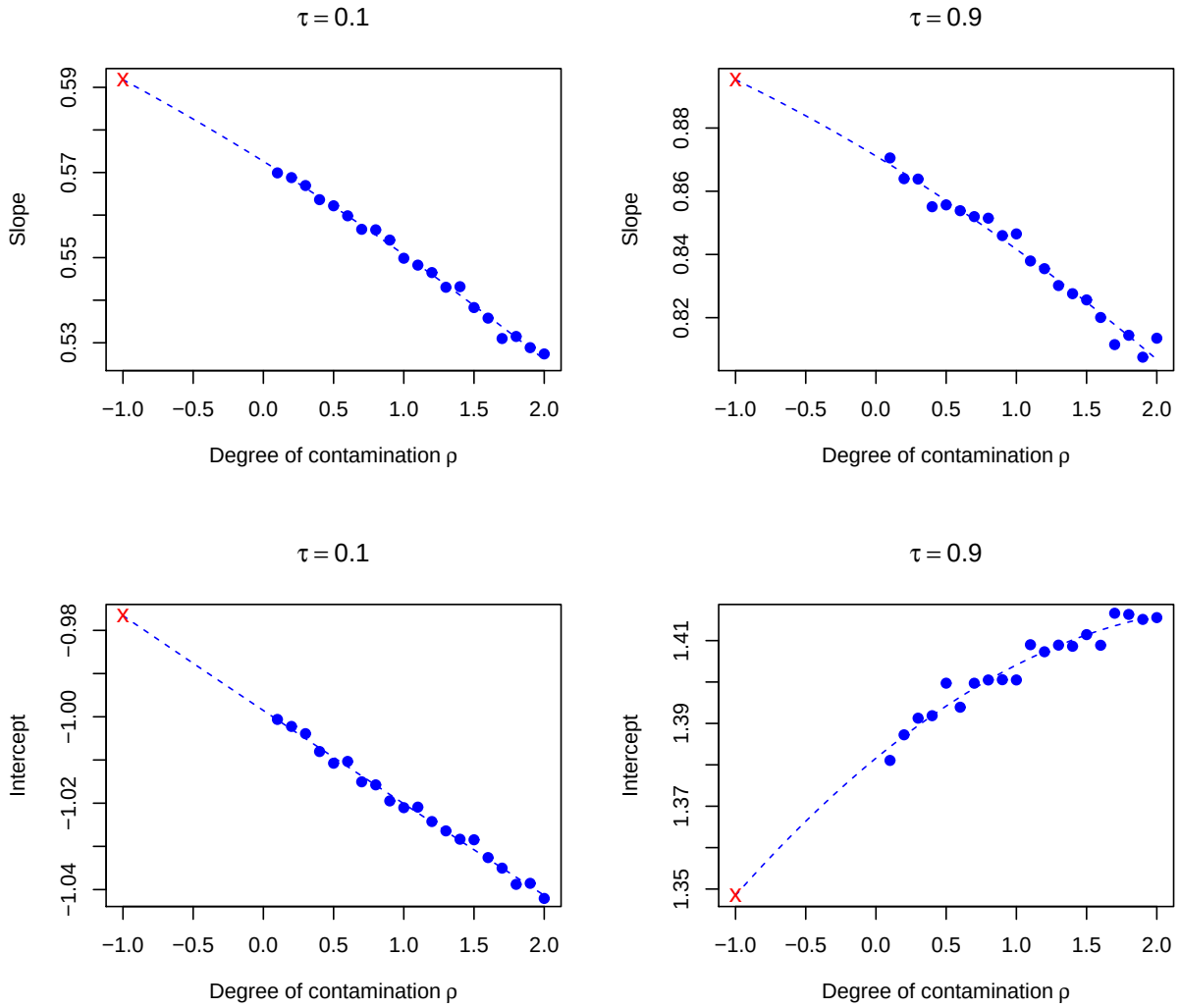
Note. The table shows estimates for the factor loadings and the scale parameters obtained from SV-SIMEX, separately for treatment and control groups. Specifically, columns (1) and (4) are obtained from 200 bootstrap replications of estimates of $\lambda_1(\tau)$ and $Q_{U_1}(\tau)$ (treatment group). For each bootstrap replication, we regress values of $\lambda_1(\tau)$ on values of $Q_{U_1}(\tau)$ using a grid from $\tau = 5\%$ to $\tau = 95\%$. This regression yields one intercept and one slope for each bootstrap replication. Column (1) reports average and standard deviation of the intercept across bootstrap replications. Column (4) reports average and standard deviation of the slope across bootstrap replications. Columns (2) and (5) of the table are obtained in the same way, using 200 bootstrap replications of SV-SIMEX estimates of $\lambda_0(\tau)$ on $Q_{U_0}(\tau)$ (control group). Columns (3) and (6) show treatment-control differences in estimates for intercept and slope, respectively. Panels in the table refer to different outcomes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

FIGURE 1. Quantile Treatment Effects and Infant Health and Development Program



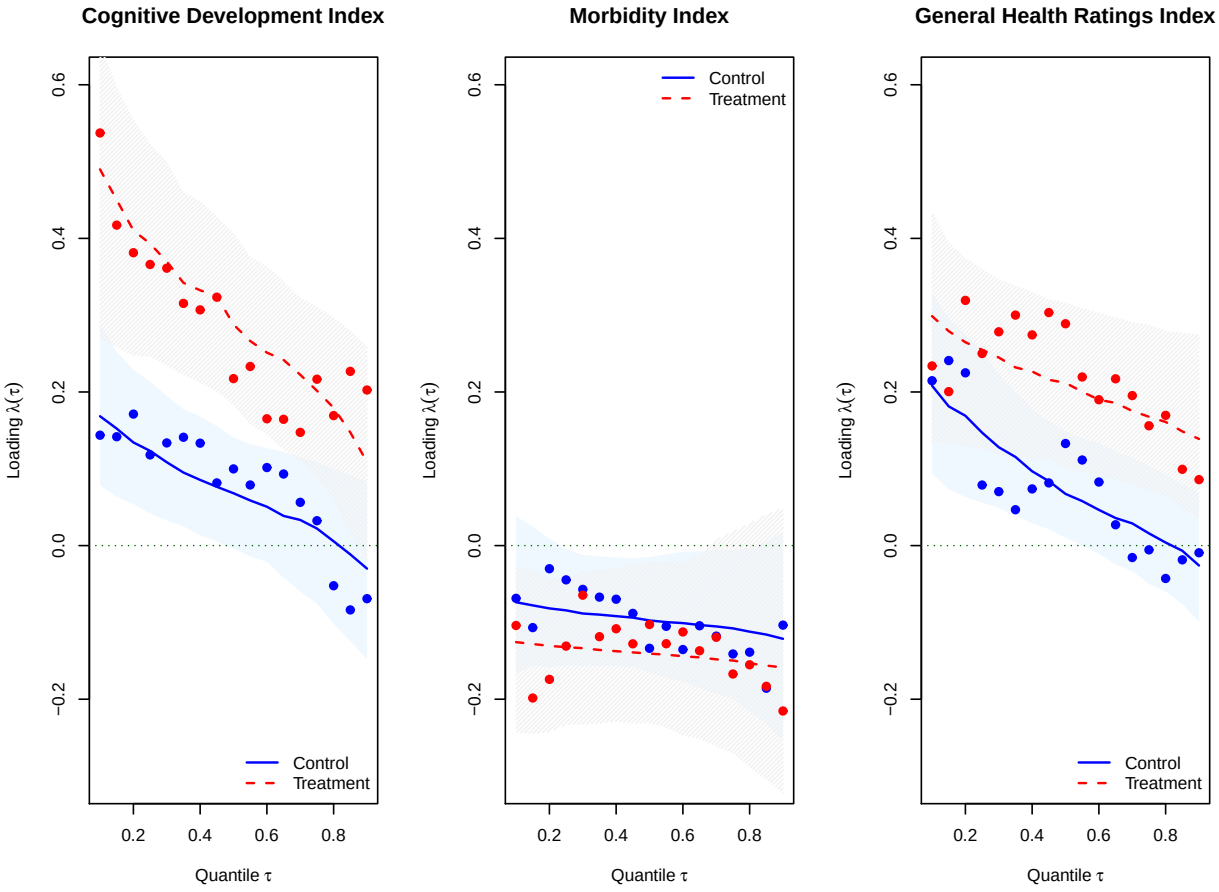
Note. Dots show estimates of quantile treatment effects (QTE) for the three outcomes. Treatment-control differences at each quantile are obtained from quantile regressions on the treatment indicator controlling for randomization strata, mother demographics, and infant demographics at birth. The continuous line is generated by a local linear regression fit to estimated QTEs. Shaded areas denote 95% confidence intervals obtained from 100 bootstrap replications.

FIGURE 2. SV-SIMEX estimation



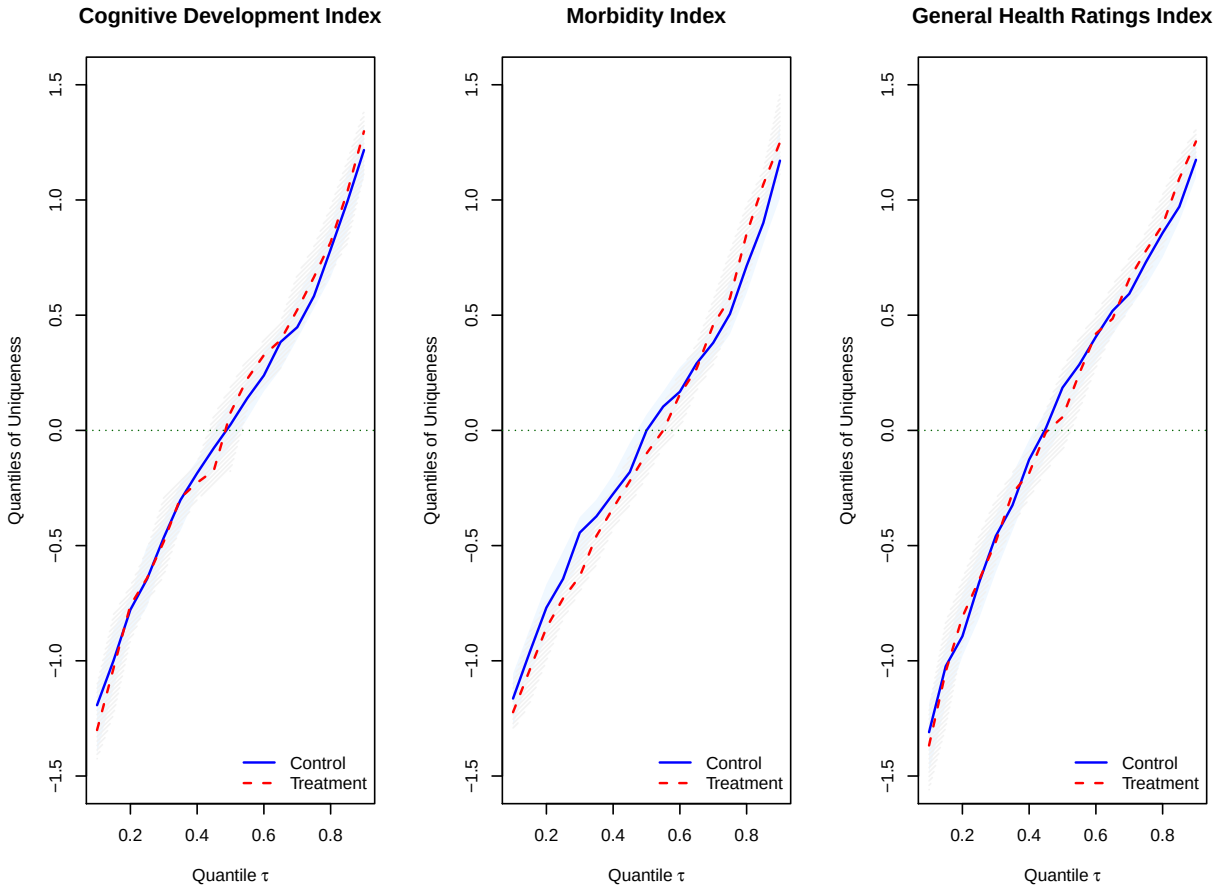
Note. The estimation idea behind SV-SIMEX is presented using one of the Montecarlo experiments in Section 7. Dots are estimates of $\phi_d^{\rho}(\tau)$ at different values of ρ (degree of contamination, on the horizontal axis) obtained setting $B = 50$. The dashed line is a quadratic fit from these dots, and is used to extrapolate to $\rho = -1$ (no measurement error).

FIGURE 3. Estimates of quantile slope of factor for treatment and control children



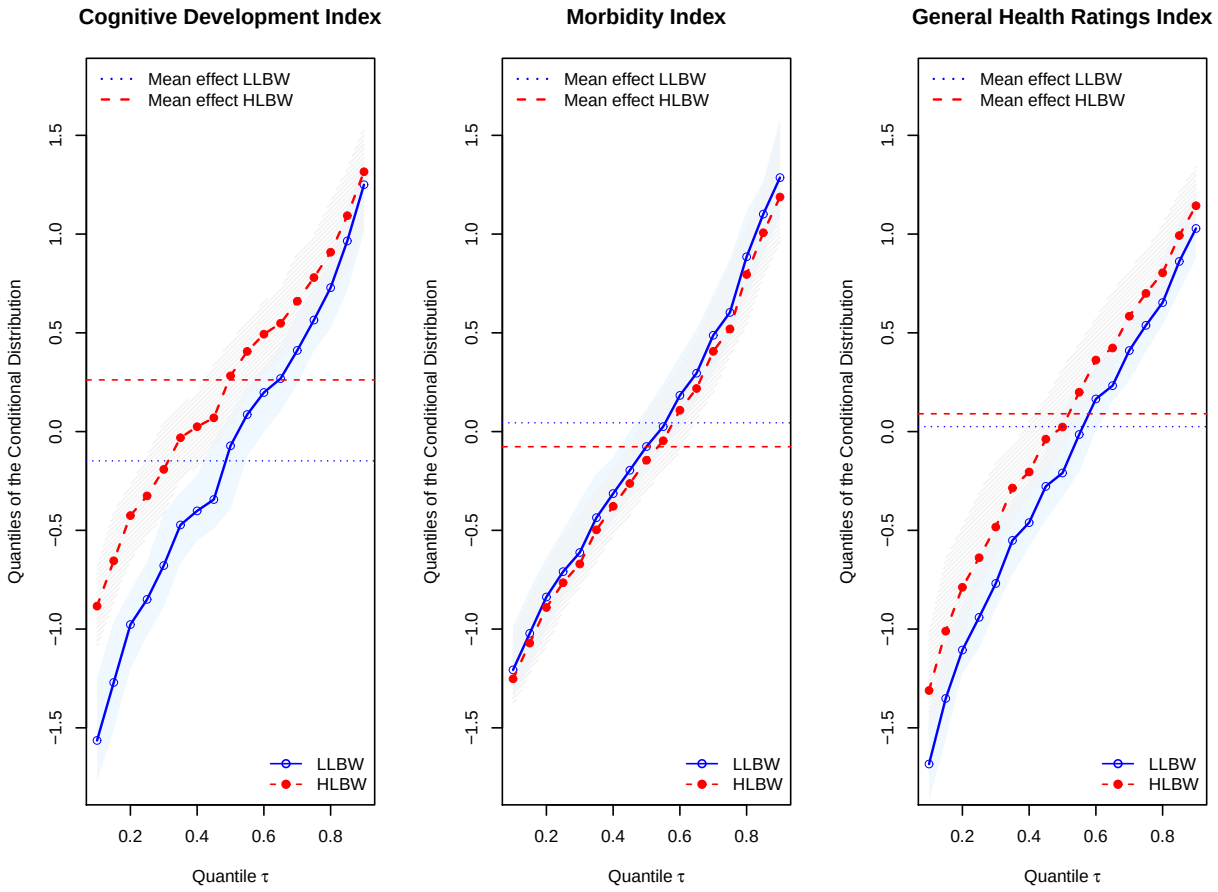
Note. Dots in the figure are SV-SIMEX estimates of $\lambda_1(\tau)$ (treatment) and $\lambda_0(\tau)$ (control). The solid line is the estimate of $\lambda_1(\tau)$ obtained by imposing the factor model. This estimate is obtained from factor loading and scale parameter estimates in columns (1) and (4) of Table 4, and SV-SIMEX estimates of $Q_{U_1}(\tau)$ in Figure 4. The dashed line is the estimate of $\lambda_0(\tau)$ obtained by imposing the factor model. This estimate is obtained from factor loading and scale parameter estimates in columns (2) and (5) of Table 4, and SIMEX estimates of $Q_{U_0}(\tau)$ in Figure 4. Shaded areas denote 95% confidence intervals obtained from 200 bootstrap replications. Panels refer to the three outcomes considered in the analysis.

FIGURE 4. Estimates of quantile intercepts for treatment and control children



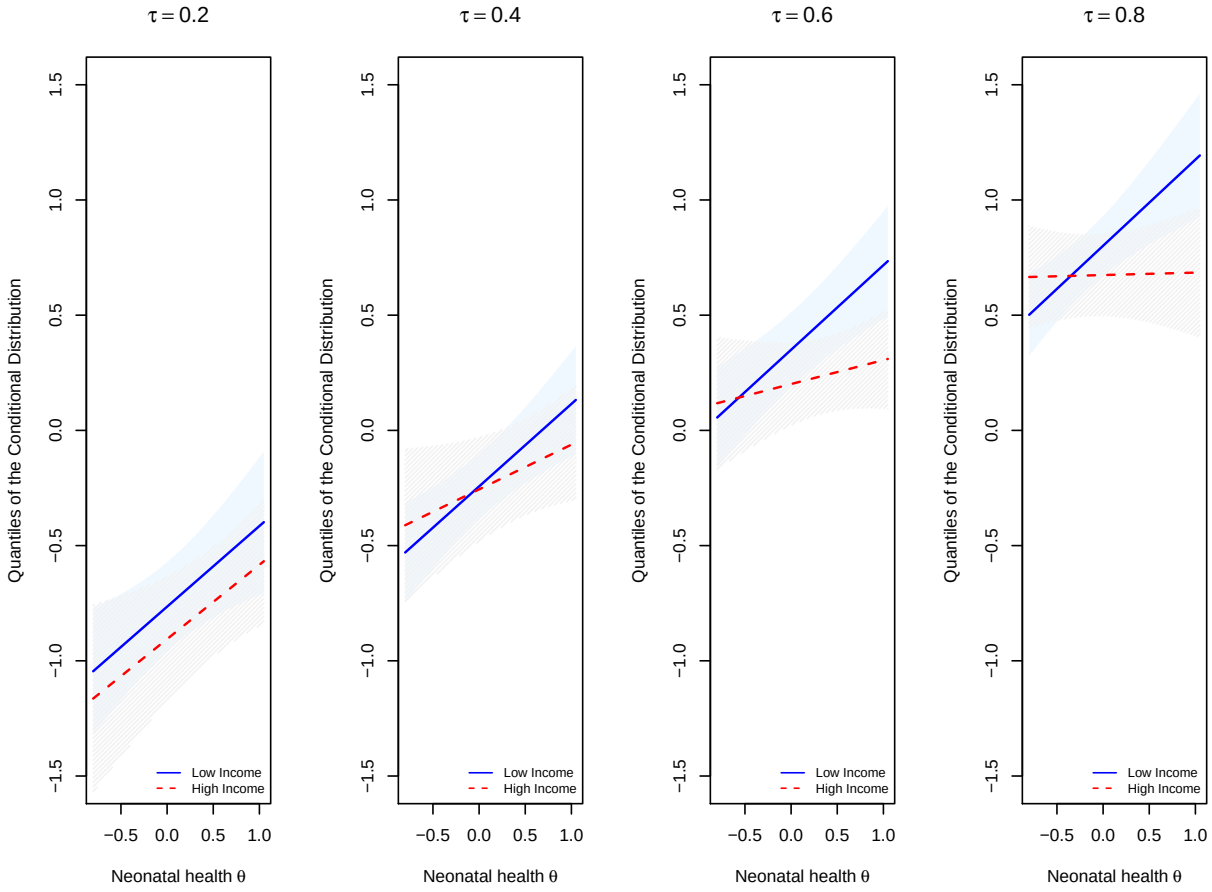
Note. The figure shows SV-SIMEX estimates of $Q_{U_1}(\tau)$ (treatment) and $Q_{U_0}(\tau)$ (control). Shaded areas denote 95% confidence intervals obtained from 200 bootstrap replications. Panels refer to the three outcomes considered in the analysis.

FIGURE 5. Estimates of quantiles of the conditional distribution of treatment effects



Note. The figure shows estimates of QCD at $u_0 = Q_{U_0}(0.50)$ and selected values of θ for the LLBW and HLBW groups (see text for details). The horizontal lines denote estimates of the average effects for LLBW and HLBW groups obtained from OLS regressions of residualized outcomes (see footnote 15) on the treatment indicator. Shaded areas denote 95% confidence intervals obtained from 200 bootstrap replications. Panels refer to the three outcomes considered in the analysis.

FIGURE 6. Conditional quantiles 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 Θ. QCD is estimated separately for children in families with income below (low-income) and above (high-income) the sample median. Only the cognitive development outcome (Stanford-Binet intelligence scale) is considered. Shaded areas denote 90% confidence intervals obtained from 200 bootstrap replications. See text for details.