

BIRTH WEIGHT AND ACADEMIC ACHIEVEMENT IN CHILDHOOD

PINKA CHATTERJI, DOHYUNG KIM* and KAJAL LAHIRI

Economics, University at Albany-SUNY, Albany, NY, USA

SUMMARY

Research has shown that birth weight has a lasting impact on later-life outcomes such as educational attainment and earnings. This paper examines the role of health at birth in determining academic achievement in childhood, which may provide the link between birth weight and adult outcomes. Using three waves of the Child Development Supplement of the Panel Study of Income Dynamics data over 1997–2007, we build on the literature by employing the fetal growth rate as a proxy for net nutritional intake *in utero* and propose a nested error-component two-stage least squares estimator that draws on internal instruments from alternative dimensions of the multilevel panel data set. In particular, this alternative estimator allows us to exploit the information on children with no siblings in the sample, which comprise over 40% of the observations in our sample, as well as to obtain coefficient estimates for the time-invariant variables such as race and maternal education. This would not be feasible with the usual mother fixed effects estimation. We obtain modest but significant effects of both birth weight and the fetal growth rate on math and reading scores, with the effects concentrated in the low birth weight range. Infant health measures appear to explain little of the well-documented racial disparity in test scores. Copyright © 2014 John Wiley & Sons, Ltd.

Received 11 October 2013; Revised 5 May 2014; Accepted 8 May 2014

KEY WORDS: hierarchical panel data; nested error component 2SLS; mother fixed effects; racial disparity; mother's education; PSID-CDS

1. INTRODUCTION

The extensive literature linking health at birth to long-term outcomes shows that higher birth weight infants achieve higher levels of educational attainment and earnings and are healthier as adults as well, compared with those with lower birth weights.¹ Studies using twin samples or natural experiments provide compelling evidence that infant health plays a causally important role in determining adult outcomes.² Much less is known, however, about the mechanism through which low birth weight translates into worse outcomes in adulthood.

Two distinct hypotheses have been advanced in the literature to explain the association between birth weight and adult outcomes. The leading explanation has drawn on the Barker hypothesis that associates low birth weight with adult chronic diseases.³ In this explanation, low birth weight has indirect consequences on adult productivity through adult health. In an alternative hypothesis, intrauterine malnutrition impairs the cognitive development of children, which may persist into their adulthood (Morgane *et al.*, 1993; Almond *et al.*, 2011).

*Correspondence to: Economics, University at Albany-SUNY, Albany, NY, USA. E-mail: dkim5@albany.edu

¹See Barker (1995), Currie and Hyson (1999), Case *et al.* (2005). For a survey on this growing body of the literature, see Almond and Currie (2011), Currie (2011). For a recent survey for developing countries, see Currie and Vogl (2013).

²See Behrman and Rosenzweig (2004), Black *et al.* (2007), Royer (2009), Figlio *et al.* (2013) for twin studies. For natural experiments, see Almond and Mazumder (2011), Almond *et al.* (2011), Van Ewijk (2011), Lindeboom *et al.* (2010), Neelsen and Stratmann (2011).

³Barker (1995, 1998) hypothesized that the intrauterine environment is crucial for adult health in that fetal insults can cause adult chronic diseases such as heart disease or diabetes.

This explanation is consistent with evidence that the effect of health at birth seems to emerge before any adult chronic conditions can develop because of compromised fetal growth.

Several studies in economics examine the test score gap between children born at low versus normal birth weight. Many studies use fixed effects estimation using samples that include twins or siblings to address the potential mother-level omitted variable bias, but the estimated effects are often statistically insignificant after controlling for mother fixed effects.⁴ This may be explained by inadequate statistical power, as the size of sibling or twin samples is typically small and the fixed effects estimation only exacerbates the problem by exploiting only the variation within mothers. Moreover, in samples of singletons, researchers often use birth weight only as a measure of health at birth, failing to account for gestational age, a factor that may have separate effects on child academic achievement. In this case, it is difficult to interpret the estimated coefficient on birth weight in the context of the fetal nutrition argument because low birth weight can reflect either a slow rate of fetal growth due to uterine malnutrition or a preterm birth or both.

In this paper, we investigate the role of health at birth in determining child academic achievement and its implications for the Black-White test score gap in childhood. In addition to birth weight, we use the fetal growth rate as a measure of net nutritional intake *in utero* in order to test the intrauterine nutrition hypothesis more directly. On the basis of a general model that includes unobserved child heterogeneity and unobserved mother heterogeneity, we propose the nested error-component two-stage least squares (hereafter NEC2SLS) estimator that uses internal instruments from alternative dimensions of the multilevel panel data. Unlike the usual mother fixed effects (hereafter MFE) estimation, this alternative estimation method enables us to exploit information from single-child families that comprise more than 40% of observations in the entire sample, as well as to obtain consistent estimates of parameters under identification assumptions that are weaker than those required for generalized least squares (GLS) estimation. Furthermore, our approach allows us to identify coefficients for mother-specific, time-invariant covariates such as race and maternal education, which would not be feasible with the usual fixed effects estimator.

Using the NEC2SLS estimator, we find positive and statistically significant effects of birth weight on math and reading test scores of children. We find, however, that the estimated effects are concentrated over the low birth weight range (less than 2.5 kg) and are modest in magnitude. In particular, it is compromised fetal growth, rather than preterm birth, that leads to lower test scores. In addition, the results indicate that maternal education is an important factor in mediating the effects of the fetal growth rate. Interestingly, we find that the estimated racial gap in test scores changes little after controlling for birth weight or the fetal growth rate.

The rest of the paper is organized as follows. In the next section, we provide a brief overview of the related literature. In Sections 3 and 4, we describe the data set and develop the NEC model. In Section 5, we discuss the inadequacies of NECGLS and MFE estimators in this class of models and suggest an alternative NEC2SLS estimator together with a simple algorithm. In Section 6, we present the empirical results. In Section 7, we allow mother's education to be endogenous. Finally, conclusions are summarized in Section 8.

2. LITERATURE

The interest in birth weight and IQ dates back at least a century.⁵ Observational studies generally find a positive association between birth weight and IQ (Sørensen *et al.*, 1997; Breslau *et al.*, 2001; Hack *et al.*, 2002), but a spurious association has been suspected because unobserved family background or genetic factors may be responsible for both infant health and child cognitive outcomes. For example, in a pioneering study using the 1950–1954 British cohorts, Record *et al.* (1969) found a strong association between birth weight and verbal test scores, but this association cannot be found within sibling pairs.

Within-twin studies can provide compelling evidence on the causal role of fetal nutrition in determining cognitive development of children, but the results are generally mixed. Boomsma *et al.* (2001) report that the

⁴One notable exception is Figlio *et al.* (2013) who found a significant and positive effect of birth weight on test scores among recent cohorts of US twins.

⁵See Asher (1946) for an early history.

effect of birth weight on child IQ can be found among dizygotic twin pairs but not among monozygotic twin pairs, suggesting that genetics may be a mediating factor, whereas Petersen *et al.* (2009) found a significant effect of birth weight among Danish male twins regardless of zygosity but not among female twins. In a study using a sample of Danish twins, Christensen *et al.* (2006) found significant effects of birth weight on test scores, although the magnitude is small. More recently, Figlio *et al.* (2013) used data on twins from a large registry in Florida and found an effect of birth weight on test scores that is remarkably stable across school grades and across socioeconomic backgrounds.

Several studies from economics use sibling samples of recent cohorts to address problems associated with confounding by unobserved factors such as family background or genetic makeup. Using Canadian registry data, Oreopoulos *et al.* (2008) found infant health has positive and significant effects on short-term health outcomes and adult outcomes but not for language arts scores, after controlling for twin or sibling fixed effects. Other within-sibling studies based on US cohorts also found that the estimates become insignificant when MFE estimation is used. In a paper that examines comprehensive life cycle outcomes, Johnson and Schoeni (2011) used MFE estimation and reported a substantial gap in test scores between male siblings having normal birth weight and those born at 1.5 kg. However, their estimates are statistically insignificant at conventional significance levels, and the model contains a birth weight spline that allows a jump at the low birth weight cutoff, which may be implausible. Fletcher (2011) also found some evidence of a positive association between birth weight and test scores, but the association is statistically insignificant in family fixed effects estimation. Moreover, because these studies do not control for gestational age in regressions, it is difficult to interpret what the estimated effects of birth weight actually capture.

We build on the literature by addressing these limitations. First, we draw on the same data source as in Johnson and Schoeni (2011) but add the recent 2007 wave of the Child Development Supplement of the Panel Study of Income Dynamics (PSID-CDS) to the sample and present some evidence on a potential misspecification in the birth weight spline function.⁶ Second, we employ the fetal growth rate as a measure of net nutritional intake *in utero*, whereas many studies use birth weight only without accounting for the gestational age. Third, we begin with a general model that includes child and mother heterogeneity and propose an alternative estimation method that exploits information more efficiently than the usual MFE estimation. Finally, we investigate the implications of birth weight effects on achievement test scores for the racial disparity in test scores, an analysis that would not be feasible with the conventional MFE estimation.

3. DATA

We use the 1997, 2002–2003, and 2007 waves of the PSID-CDS. The CDS provides assessments of academic achievement of children who are born between 1984 and 1997 in PSID households and rich socioeconomic and demographic data. In 1997, the first wave of the CDS interviewed 2394 families on 3563 children aged 12 years or younger, and these children were reinterviewed in 2002–2003 and 2007 if they were 18 years or younger at the time of the interview. Hence, the data set includes multiple observations, at most three, for each child.

We restrict our sample to children whose primary caregiver is the biological mother and the head or wife of a PSID household to access information on maternal and family characteristics from the main PSID files. Table I gives the summary statistics on the variables used. Our final sample used in the analysis includes 2673 children and over 4000 observations, depending on the test scores being analyzed. Over 40% of the observations come from single-child families.⁷

⁶The Panel Study of Income Dynamics (PSID) is a public use data set. It is produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. Details about the PSID and the PSID-CDS are available at <http://psidonline.isr.umich.edu/>.

⁷We define a single-child family as a family that has a single child in the Child Development Supplement survey in a given survey wave. Information from these families is swept out when mother fixed effects estimation is used.

Table I. Summary statistics

Variables	Dimension	N	Mean	Standard deviation
<i>Test scores</i>				
Applied problems	M C T	4609	104.04	16.94
Passage comprehension	M C T	4106	102.67	16.63
<i>Time-varying characteristics</i>				
Family income (2007-constant thousand dollars)	M . T	5734	54.11	58.24
Mother working	M . T	5734	0.68	0.47
Child age in months	M C T	5734	121.2	56.4
<i>Time-invariant child characteristics</i>				
Birth weight (kg)	M C .	2673	3.33	0.63
Low birth weight (<2.5 kg)	M C .	2673	0.09	0.28
Gestational age (weeks)	M C .	2673	39.48	2.19
Fetal growth rate (grams per week)	M C .	2673	83.96	14.33
Female	M C .	2673	0.49	0.50
Maternal age at child birth	M C .	2673	27.20	5.56
Mother age at child birth <20	M C .	2673	0.09	0.28
Mother age at child birth >35	M C .	2673	0.07	0.25
Mother single at child birth	M C .	2673	0.29	0.45
First born	M C .	2673	0.39	0.49
Second born	M C .	2673	0.35	0.48
Third born	M C .	2673	0.17	0.37
Fourth born	M C .	2673	0.06	0.23
Fifth or more	M C .	2673	0.03	0.17
Home environment	M C .	2673	19.19	3.56
<i>Time-invariant maternal characteristics</i>				
Non-Latino White	M . .	1807	0.50	0.50
Non-Latino African American	M . .	1807	0.39	0.49
Latino	M . .	1807	0.06	0.24
Other race	M . .	1807	0.05	0.21
Mother's education (years)	M . .	1807	12.89	2.51

The entries are from 1997, 2002–2003, and 2007 PSID-CDS. The second column indicates which dimensions the variable varies across where M indicates mother, C indicates child, and T indicates survey wave. For example, birth weight varies across mothers (M) and children (C) but does not vary over time.

3.1. Health at birth

The PSID-CDS contains detailed information on health at birth. In particular, the primary caregiver, who is the biological mother in our sample, reports at the time of the first CDS interview the birth weight of children along with the gestational age. The birth weight is reported in pounds and ounces, which we convert into kilograms to allow comparability with other studies.⁸ The gestational age is reported in days before or after the due date, which we convert into weeks and fractions thereof. Table I shows that the sample mean of birth weight is 3326 g, which matches closely to 3369 g reported in Almond *et al.* (2005) based on the US natality files for the singletons born in 1989. The average gestational age of 39.48 weeks in the full sample also matches closely to 39.3 weeks from the same source. These figures suggest that the recall bias in the PSID-CDS may be minimal.

Birth weight is the most widely used measure of health at birth. We use log of birth weight, which provides the best fit and accounts for potential nonlinearity in the effect of birth weight.⁹ In an alternative specification, we use a birth weight spline function to estimate the difference in the magnitude of the birth weight effect over the birth weight distribution.¹⁰

⁸We exclude three children from the analysis whose reported birth weight is over 6 kg. The analysis is not sensitive to this exclusion.

⁹Other studies also find log of birth weight provides the best fit for different outcomes. For example, Black *et al.* (2005) and Figlio *et al.* (2013).

¹⁰See Almond *et al.* (2005), Behrman and Rosenzweig (2004), Currie and Moretti (2007), Royer (2009) for nonlinear effects of birth weight on various outcomes. Among the studies that examine cognitive abilities, Boardman *et al.* (2002) use a sample of US children from the National Longitudinal Survey of Youth and find larger cognitive deficits at the left-hand tail of birth weight distribution. Similarly, Figlio *et al.* (2013) found nonlinear effects of birth weight on test scores among the sample of US twins born in Florida.

Conceptually, birth weight is determined simultaneously by the average growth rate *in utero* and gestational age. When a twin sample is used, the gestational age is controlled for by twin fixed effects estimation. However, when singleton samples are used, the estimated birth weight effect may capture the effect of gestational age in addition to the effect of the intrauterine growth rate.¹¹ In order to distinguish between the two effects, we include measures of both the intrauterine growth rate and gestational age in the same equation. These are, respectively, the fetal growth rate (defined as birth weight divided by gestational age) and a preterm birth indicator (defined as one if the gestational age is less than 37 weeks and zero otherwise). This specification will be useful to test the fetal nutrition hypothesis because the fetal growth rate can better proxy nutritional status *in utero* as compared with birth weight (Behrman and Rosenzweig, 2004). Moreover, we can investigate the effect of preterm birth, which in itself can be a measure of infant health.¹²

Figure 1A displays a clear positive correlation between birth weight and gestational age in our sample, implying that birth weight is partly determined by gestational age. Figure 1B shows that the fetal growth rate is still positively correlated with gestational age, although it is a measure that has already accounted for gestational age. These empirical regularities also are found in the general population as can be seen in the line plots in Figure 1C–F, which are obtained on the basis of all US singleton births over 1989–1997 in the United States.

Indeed, one limitation of using the average growth rate *in utero* is that the momentary growth rate may change over gestational age. The observed momentary fetal growth rate steadily increases till about 40 weeks of gestation and tapers off thereafter as can be seen in Figure 1D and F. For this reason, Torche and Echevarría (2011) suggest standardizing the fetal growth rate using the sample mean and standard deviation of the fetal growth rate for each gestational age. This standardized fetal growth rate is orthogonal to gestational age by construction and measures deviation from the average growth rate given gestational age.¹³ Unfortunately, this adjustment is not feasible using our sample because data points are too sparse for small gestational ages.¹⁴ However, this definitional issue is not critical in our analysis because the vast majority of the observations in our sample come from the children whose gestational ages are greater than 30 weeks, for which the change in the rate of fetal growth is modest, as can be seen in Figure 1D or F. The sample correlation coefficient between the fetal growth rate and gestational age sharply drops from 0.36 to 0.12 as we restrict our sample to term births. Moreover, the inclusion of a preterm birth indicator in the regression will account for the potential confounding arising from a separate effect of preterm birth. Therefore, following Barker (1966) and Behrman and Rosenzweig (2004), we focus on estimates using the fetal growth rate after controlling for a preterm birth indicator.

3.2. Academic achievement

To measure children's academic achievement, we use the scores on the Woodcock–Johnson Psycho-Educational Battery–Revised (WJ-R) academic achievement tests, which are a well-established measure of several dimensions of academic achievement including the degree of mastery in reading and mathematics

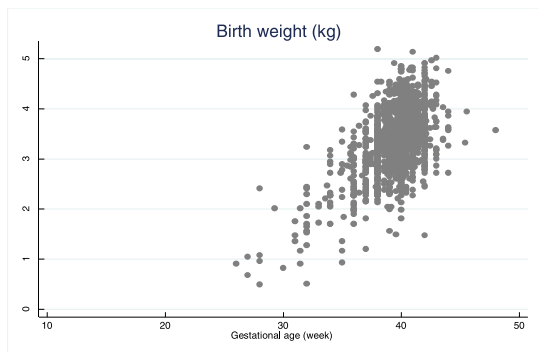
¹¹The literature in epidemiology almost always controls for gestational age when singleton samples are used even if gestational age is not the variable of interest.

¹²There is a line of research in epidemiology that focuses solely on the consequences of preterm birth, see Bhutta *et al.* (2002), for example. However, the causes of preterm birth remain largely unknown.

¹³In the medical literature, an indicator for small-for-gestational age is often used in place of a low birth weight indicator because the fetal growth rate depends on gestational age.

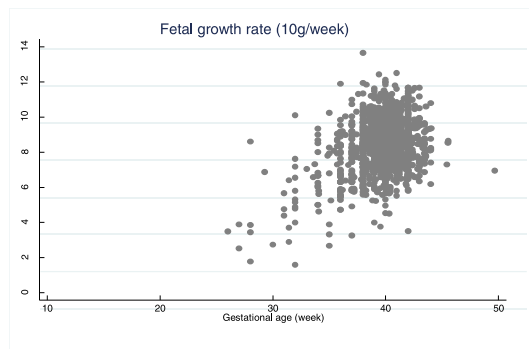
¹⁴In results not shown in the paper, we attempted to standardize fetal growth rate measure. Instead of using sample means and sample standard deviations, we standardized fetal growth rate using the sample means and standard deviations obtained from over 36 million US births from the Vital Statistics Birth Files over 1989–1997 birth cohorts, which closely match to those in our sample. Because fetal growth profile may vary genetically by race and sex, we further standardized the fetal growth rate by race and sex in addition to gestational age. We have estimated our model with this measure together with a preterm birth indicator. The estimates were very similar and can be obtained from the authors.

A. Observations in sample – birthweight



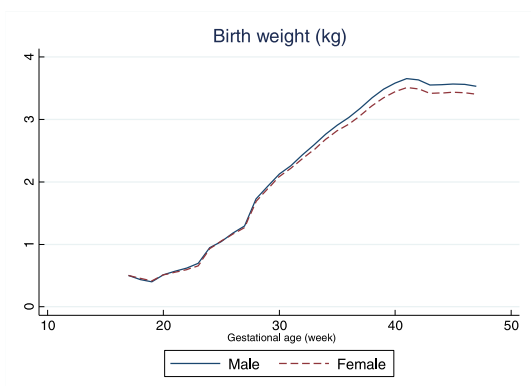
Source: PSID-CDS, N = 4,609.

B. Observations in sample – fetal growth rate



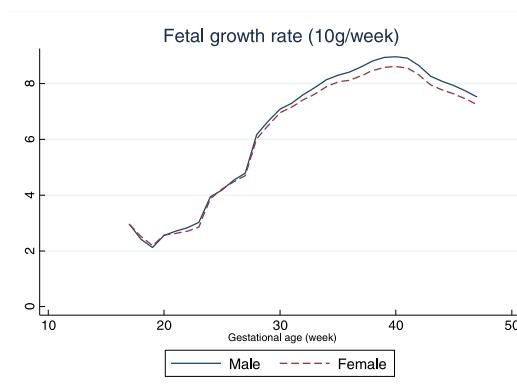
Source: PSID-CDS, N = 4,609.

C. Group mean by sex – birthweight



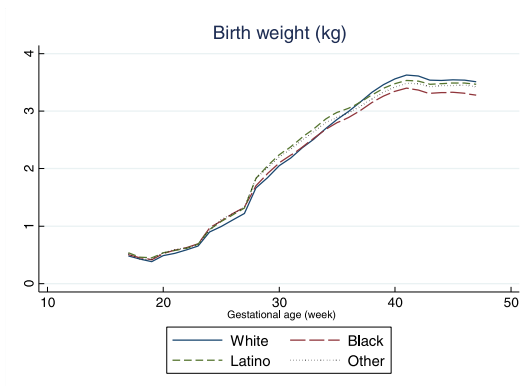
Source: Vital Statistics Birth files 1989-97, N = 34,106,761.

D. Group mean by sex – fetal growth rate



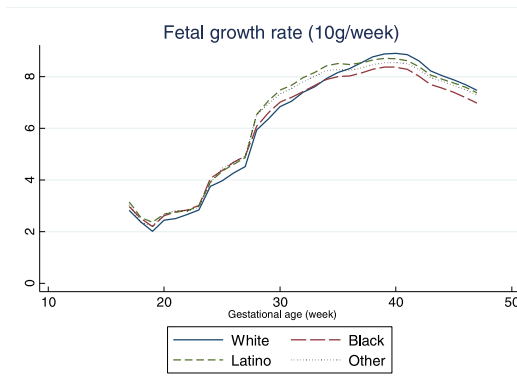
Source: Vital Statistics Birth files 1989-97, N = 34,106,761.

E. Group mean by race – birthweight



Source: Vital Statistics Birth files 1989-97, N = 34,106,761.

F. Group mean by race – fetal growth rate



Source: Vital Statistics Birth files 1989-97, N = 34,106,761.

Figure 1. Birth weight and fetal growth rate by gestational age.

(Woodcock and Johnson, 1989).¹⁵ Scores on three subtests are available for all three waves of the CDS: the applied problems, the letter-word identification, and the passage comprehension test. The applied problems test measures skills in math reasoning, and the task involves analyzing and solving practical mathematical problems. The letter-word identification test measures basic reading skills, and the task requires identifying and pronouncing isolated letters and words. These tests are administered to all children aged 3 years and older. The passage comprehension test measures vocabulary and skills in reading comprehension, and the task requires reading a short passage silently and then supplying a key missing word. This subtest is administered to older children (aged 6 years and older) because it requires reading ability.¹⁶ For math scores, we examine the applied problems test primarily because the calculation test is available only in the first wave of the CDS. For reading scores, we chose the passage comprehension over the letter word, which is no more than a simple test of letter identification and word pronunciation.¹⁷ We present results for reading scores, but our preferred outcome measure is the mathematics score, which is consistently shown to be a stronger predictor of subsequent earnings in the literature (Murnane *et al.*, 1995).

The WJ-R is administered at home to children who answer the questionnaires using a response book under the supervision of trained interviewers. A raw score is the summation of the total number of correct responses, each correct response receiving 1 point. Because the children are assessed at different ages, we use the standard scores given in the CDS that are age-adjusted in reference to the national distribution of raw scores among the children of the same monthly age. These age-equivalent scores are normalized to have mean 100 and standard deviation 15. Table I shows that the mean and standard deviation of test scores in our sample are close to those for the national sample, indicating that the score distributions in our sample are fairly representative. For applied problems, we have total of 4609 scores including at most three repeated observations for each child. The highest score attained by any child is 171 and the median is 103 with the lowest being 2. For passage comprehension, we have total of 4016 scores including repeated observations. The highest score is 187, and the median is 101. Ten children scored 0, the lower bound, and the age for these children ranged from 6 to 7 years (71 to 78 months) at the time of test.

3.3. Other covariates

One advantage of using the CDS is that rich and reliable information on family and maternal characteristics can be obtained by matching the CDS files to the PSID main files. In the regression, we include demographic characteristics such as sex, race, and child age measured in months (non-Latino White and female as the reference groups). The child age at assessment is exogenous by construction because we use the standard test scores that are aged-adjusted. To control for family characteristics that may affect birth weight and academic achievement in childhood, we include log of permanent family income, which is measured by 6-year average of family incomes in terms of 2007-constant dollars. We also control for a binary indicator for the mother being single at the time of the child's birth and indicators for maternal age at child's birth being younger than 20 years and older than 35 years. We include in all regressions a set of indicators for birth order, which has been shown to affect cognitive abilities of children (Black *et al.*, 2005; Sulloway, 2007).

¹⁵The Woodcock–Johnson Psycho-Educational Battery–Revised (WJ-R) test often has been used in the education and developmental psychology literatures, as well as in the economics literature. For example, the WJ-R test score was used as a measure of academic skills in the well-known randomized experiment, the Carolina Abecedarian project (Campbell *et al.*, 2002). In economics, in addition to Johnson and Schoeni (2011)'s study of birth weight and achievement test scores, subscales of the WJ-R have been used to study other topics such as language assimilation among children of immigrants (Akresh and Akresh, 2011), effects of home ownership on children's academic achievement (Mohanty and Raut, 2009), and effects of parental risk aversion on children's academic achievement (Brown *et al.*, 2012).

¹⁶For more information on the Woodcock–Johnson Psycho-Educational Battery–Revised academic achievement tests, see Mather (1991).

¹⁷Broad reading or broad math scores are often used in the literature, each of which is a composite of two separate tests. The Child Development Supplement (CDS) provides broad reading scores for all the CDS waves, and the estimates are very similar to those for a single test score.

Finally, we include in the regressions: the number of years of maternal education and the quality of the child's home learning environment, measured by the Home Observation for Measurement of the Environment–Short Form (HOME-SF) (Caldwell and Bradley, 1984). Both maternal education and the home environment have been shown to be correlated with each other and are important for children's academic achievement (Phillips *et al.*, 1998; Carneiro *et al.*, 2013). For example, Phillips *et al.* (1998) show that mother's education and home environment are significant predictors of test score for 5- and 6-year-olds. The HOME-SF is a continuous measure of cognitive stimulation and emotional support given to children, and it is based on both caregiver reports and interviewer observations.¹⁸ The HOME-SF includes a large number of items that vary by the age of child. Many of the items are intended to capture the mother's parenting practices and how she interacts with, provides stimulation for, and disciplines her child. These items are dichotomized and summed to form a total raw score, with higher scores indicating a better quality home environment. This raw score is included in the models to capture the quality of the home environment. This measure has been extensively used as a determinant of development in early childhood.¹⁹

4. MODEL

4.1. One-way error component model

We begin with the model considered in Johnson and Schoeni (2011). Their specification can be written as follows:

$$y_{ijt} = \alpha + w_{ij}\beta + \delta D_{ij} + \gamma^L D_{ij}(BW_{ij} - 1.5) + \gamma^N (1 - D_{ij})(BW_{ij} - 1.5) + m_i + e_{ijt} \quad (1)$$

where y_{ijt} denotes test scores assessed at the survey wave t of child j of mother i , D_{ij} a binary indicator of low birth weight (less than 2.5 kg), BW_{ij} the birth weight, w_{ij} a vector of child and family characteristics, m_i the unobserved mother heterogeneity, and e_{ijt} the error term.

Two features of equation (1) should be noted. First, this model is essentially an MFE model, which allows for correlated maternal heterogeneity, but precludes implicitly correlated child heterogeneity. We formally test for correlated child heterogeneity as we extend this model in the next subsection. Second, the model accommodates potential nonlinearity in the effect of birth weight on outcomes by specifying two different slopes for birth weight γ^L and γ^N with a knot at the low birth weight cutoff. In particular, the estimate for δ will give the test score gap, which is generated by the greater slope over the low birth weight range as compared with the normal birth weight range, evaluated at a particular point (1.5 kg) in birth weight distribution. One unintended consequence of this specification is that it allows a jump at the 2.5 kg knot as is depicted in panel A of Figure 2, which may be implausible.²⁰ We will use log of birth weight in a baseline specification because it provides the best fit but will also consider a continuous piecewise linear specification in birth weight. More specifically, we use the conventional low birth weight cutoff as a knot for spline, but the results are qualitatively similar when some variations in the values of the knots are allowed for. The semiparametric estimates shown in panel B of Figure 2 suggest that the conventional low birth weight cutoff (dotted line) is fairly reasonable.²¹

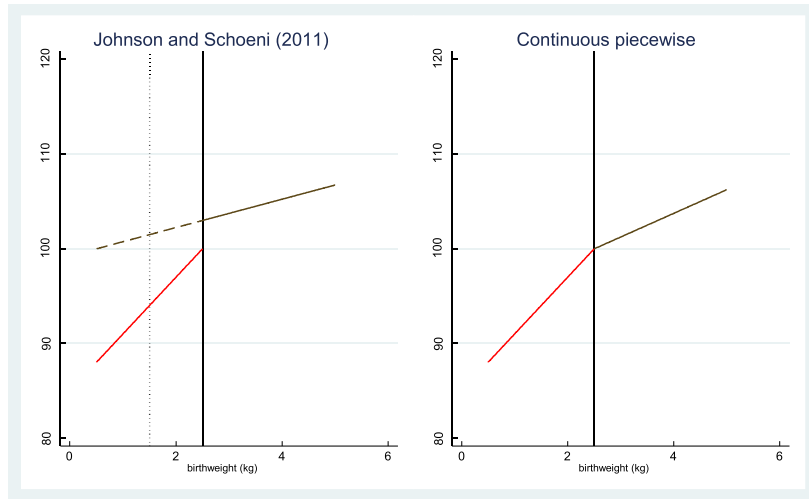
¹⁸See PSID-CDS User Guide for CDS-II (2010). http://psidonline.isr.umich.edu/CDS/cdsii_userGd.pdf.

¹⁹See Elardo and Bradley (1981) for a survey on the literature.

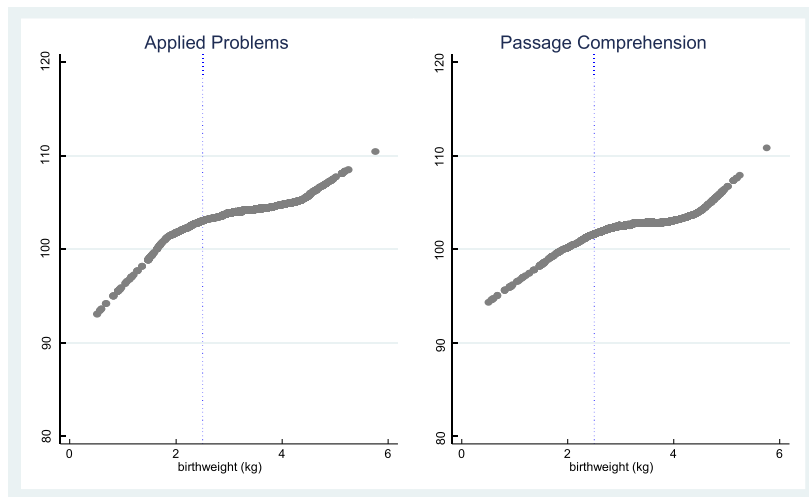
²⁰This implicit modeling assumption can be tested by adding a binary indicator of low birth weight in the usual continuous piecewise regression. We replicated Table III in Johnson and Schoeni (2011) and tested the assumption. The estimated jump is statistically insignificant with its magnitude unreasonably large. Moreover, the estimated slopes of birth weight spline were negative for some outcome measures. This result is counterintuitive as higher birth weight imposes a cognitive penalty rather than providing an advantage.

²¹The estimates are obtained after controlling for all the covariates included our baseline model.

A. Schematic representation of birthweight spline



B. Semiparametric estimates of birthweight effect



Source: 1997, 2002-3, 2007 waves of PSID-CDS.

Figure 2. Nonlinearity in birth weight effect.

4.2. Two-way nested error-component model

A two-way NEC model, which contains both child and mother heterogeneity, has been often used in the education production literature (Todd and Wolpin, 2003; Kim and Frees, 2006) and in other contexts (Baltagi *et al.*, 2001). In particular, Boardman *et al.* (2002) estimated this model by maximum likelihood assuming uncorrelated effects to find a significant test score gap among low birth weight children. The model can be written as follows:

$$y_{ijt} = x_{ijt}\beta + w_{ij}\gamma + z_i\delta + u_{ijt} \tag{2}$$

where test score y_{ijt} is a function of birth weight and a set of child and family characteristics. x_{ijt} denote the vector of time-varying child and family characteristics; w_{ij} the vector of time-invariant, child-specific characteristics including log of birth weight; and z_i the vector of time-invariant, mother-specific characteristics. We write the disturbance term u_{ijt} as follows:

$$u_{ijt} = m_i + c_{ij} + e_{ijt} \quad (3)$$

where m_i denote the heterogeneity of mother, i , c_{ij} denote the heterogeneity of child j nested in mother i , and e_{ijt} denote the global error term. Equation (3) corresponds naturally to the multilevel nested grouping in our data set. Equations (2) and (3) can be written compactly as follows:

$$y_{ijt} = x_{ijt}\beta + w_{ij}\gamma + z_i\delta + m_i + c_{ij} + e_{ijt} \quad (4)$$

Notice that the model (4) is more general than the MFE model (1) in that the child heterogeneity is included in addition to the maternal heterogeneity. Hence, the MFE estimation of the model (4) will be inconsistent if the child heterogeneity is correlated with any of x_{ijt} or w_{ij} (Kim and Frees, 2006).

5. ESTIMATION

We begin with decomposing Equation (4) into three constituent regressions, which will be useful in the discussion that follows. The within-child regression can be written as follows:

$$(y_{ijt} - \bar{y}_{ij.}) = (x_{ijt} - \bar{x}_{ij.})\beta + (e_{ijt} - \bar{e}_{ij.}) \quad (5)$$

where $\bar{y}_{ij.} \equiv \frac{1}{T_i} \sum_{t=1}^{T_i} y_{ijt}$ and $\bar{x}_{ij.}$ and $\bar{e}_{ij.}$ are defined similarly. The ordinary least squares (OLS) regression of Equation (5), which is equivalent to the child fixed effects (CFE) estimation, identifies β in the presence of correlated child or mother effects. The within-mother or within-sibling regression can be written as follows:

$$(y_{ijt} - \bar{y}_{i..}) = (x_{ijt} - \bar{x}_{i..})\beta + (w_{ij} - \bar{w}_{i.})\gamma + (c_{ij} - \bar{c}_{i.}) + (e_{ijt} - \bar{e}_{i..}) \quad (6)$$

where $\bar{y}_{i..} \equiv \frac{1}{T_i N_i} \sum_{t=1}^{T_i} \sum_{j=1}^{N_i} y_{ijt}$, $\bar{w}_{i.} \equiv \frac{1}{N_i} \sum_{j=1}^{N_i} w_{ij}$, and $\bar{x}_{i..}$, $\bar{c}_{i.}$, and $\bar{e}_{i..}$ are similarly defined. The OLS regression of Equation (6), which is equivalent to MFE estimation, identifies (β, γ) provided that the child heterogeneity c_{ij} is uncorrelated with (x_{ijt}, w_{ij}) . The between-mother regression can be written as follows:

$$\bar{y}_{i..} = \bar{x}_{i..}\beta + \bar{w}_{i.}\gamma + z_i\delta + m_i + \bar{c}_{i.} + \bar{e}_{i..} \quad (7)$$

where (β, γ, δ) are identified by OLS regression provided that the child and mother heterogeneity are uncorrelated with all the regressors.²²

5.1. Nested error-component generalized least squares

We discuss the NECGLS estimation that will serve as a building block for the NEC2SLS estimator. Under the assumptions that the error components m_i , c_{ij} and e_{ijt} are identically and independently distributed with mean zero and variance σ_m^2 , σ_c^2 , and σ_e^2 , and each error component is uncorrelated with the regressors, the NECGLS is consistent and efficient. For the NECGLS estimation, following Fuller and Battese (1973), we first transform the Equation (2) as follows:

$$\tilde{y}_{ijt} = \tilde{x}_{ijt}\beta + \tilde{w}_{ij}\gamma + \tilde{z}_i\delta + \tilde{u}_{ijt} \quad (8)$$

²²Between-child regression can be defined but is redundant in our context given Equations (5)–(7).

where for $i = 1, \dots, M, j = 1, \dots, N_i, t = 1, \dots, T_i,$

$$\tilde{x}_{ijt} \equiv x_{ijt} - \alpha_{1i}\bar{x}_{ij.} - \alpha_{2i}\bar{x}_{i..}, \tag{9}$$

$$\alpha_{1i} \equiv 1 - \left[\frac{\sigma_e^2}{\sigma_e^2 + T_i\sigma_c^2} \right]^{\frac{1}{2}}, \tag{10}$$

$$\alpha_{2i} \equiv \left[\frac{\sigma_e^2}{\sigma_e^2 + T_i\sigma_c^2} \right]^{\frac{1}{2}} - \left[\frac{\sigma_e^2}{\sigma_e^2 + T_i\sigma_c^2 + N_iT_i\sigma_m^2} \right]^{\frac{1}{2}}. \tag{11}$$

Equations (9)-(11) represent the Fuller and Battese transformation for two-way NEC model. The transformation makes the error term have a scalar covariance matrix and allows us to obtain the NECGLS estimates by the OLS regression of the transformed Equation (8). Given the assumptions required for a consistent NECGLS estimation, a number of ways to estimate variance components $\sigma_m^2, \sigma_c^2,$ and σ_e^2 are suggested in the literature (Baltagi *et al.*, 2001). We estimate the variance components using a method suggested in Fuller and Battese (1973). Intuitively, σ_e^2 and σ_c^2 can be estimated consistently from Equations (5) and (6) and the overall variance σ_u^2 from Equation (7). Then, consistent estimate for σ_m^2 can be obtained by subtracting $\hat{\sigma}_e^2$ and $\hat{\sigma}_c^2$ from the estimated overall variance $\hat{\sigma}_u^2$.

Note that the NECGLS estimates will be inconsistent if any explanatory variable is correlated with any error component in matching dimension of the multilevel data. The CFE estimation is robust to either the correlated maternal heterogeneity m_i or the correlated child heterogeneity c_{ij} , but it is not an option for our purpose because only the estimates for β will be obtained while the coefficients of interest lie in γ and δ . The MFE estimation has been widely used in the literature under the implicit assumption that only the maternal heterogeneity m_i is correlated with the covariates. However, in our model, even the MFE estimation can be inconsistent in the presence of correlated child heterogeneity c_{ij} .

5.2. Nested error-component two-stage least squares

Unlike MFE, our estimation strategy is to allow only a subset of the regressors to be endogenous, which will be tested using the Hausman (1978) test. In a two-way NEC model, there can be three different variants of the Hausman test depending on the two alternative hypotheses (Kim and Frees, 2006). On the basis of the results from three Hausman tests that will be presented in the next section, we will maintain that some of the covariates are correlated with the maternal heterogeneity m_i but not with the child heterogeneity c_{ij} , and the others are uncorrelated with either the child or mother heterogeneity. However, our estimation strategy does allow some of the covariates to be correlated with the child heterogeneity c_{ij} in general. To perform 2SLS estimation on the Fuller-Battese transformed equation (8), we follow the approach suggested in Breusch, Mizon, and Schmidt (1989) and Kinal and Lahiri (1993). Note that time-varying variable x_{ijt} can be decomposed into three components because we can always write as $x_{ijt} = (x_{ijt} - \bar{x}_{ij.}) + (\bar{x}_{ij.} - \bar{x}_{i..}) + \bar{x}_{i..}$. Likewise, for regressors w_{ij} that do not vary over time, but vary across siblings, the decomposition can be written as $w_{ij} = (w_{ij} - \bar{w}_{i.}) + \bar{w}_{i.}$. In matrix form, we can write the decomposition of x_{ijt} as $X = Q_1X + Q_2X + PX$ where $Q_1, Q_2,$ and P are defined in such a way (Baltagi *et al.*, 2001) that Q_1X denotes the deviation from the child mean, Q_2X denotes the child deviation from the mother mean, and PX denotes the mother mean. Similarly, we can write the decomposition of w_{ij} in matrix form as $W = Q_2W + PW$. Then, the NECGLS estimates can be obtained by performing the 2SLS estimation of Equation (8) where the list of instruments is as follows:

$$A = (Q_1X, Q_2X, PX, Q_2W, PW, Z)$$

where Z is a matrix representation of z_i in Equation (2). The aforementioned set of instruments give the GLS estimates because the set A includes all the regressors that are decomposed into their Q_1 , Q_2 , and P components.

Now, we partition $X = (X_1, X_2)$ where X_1 are uncorrelated with c_{ij} and m_i , but X_2 are allowed to be correlated with c_{ij} or m_i . Similarly, we partition $W = (W_1, W_2)$ where W_1 are uncorrelated with c_{ij} and m_i , but W_2 may be correlated with c_{ij} or m_i . Under the identification assumption that X_1 and W_1 are uncorrelated with c_{ij} , the consistent NEC2SLS estimator is the 2SLS estimator where the list of instruments is as follows:

$$B = (Q_1X, Q_2X_1, PX_1, Q_2W_1, PW_1, Z).$$

Note that, compared with the full instrument set A , the instruments Q_2X_2 , PX_2 , Q_2W_2 , and PW_2 are excluded in the set B because of their potential correlation with c_{ij} . Alternatively, under the assumption that X_2 and W_2 are correlated with only with m_i , which we will adopt eventually, the consistent NEC2SLS estimator is the 2SLS estimator where the list of instruments is as follows:

$$C = (Q_1X, Q_2X, PX_1, Q_2W, PW_1, Z).$$

Note that, compared with the set B , in the set C , we bring back the instruments Q_2X_2 and Q_2W_2 because these are now assumed to be uncorrelated with m_i . In the next section, we will discuss how we partition X and W based on the Hausman tests and determine which variables are correlated with c_{ij} or m_i .

Note that the NEC2SLS estimator that uses either the instrument set B or C is consistent under the identification assumptions weaker than those required for the GLS estimation because some regressors (X_2 , W_2) are allowed to be correlated with either c_{ij} or m_i . In particular, by excluding the components in the set C that are correlated with the child heterogeneity, the NEC2SLS estimation by using the instrument set B can potentially address the concern with correlated child heterogeneity, which will not be feasible in the MFE estimation.

On the other hand, the NEC2SLS estimator that uses the set of instruments C requires stricter identification assumptions than the MFE estimation where all regressors are allowed to be correlated with the maternal heterogeneity m_i . However, by relaxing some of the MFE assumptions, we can exploit information on the single-child families and between-mother variation in the data. The latter allows us to recover the coefficient estimates for the time-invariant, mother-specific covariates such as maternal education and race, which is not feasible in the MFE estimation. Often, these variables are of special interest to policy makers.

Intuitively, we can see how the model parameters are identified without any external instruments from the three regressions defined in Equations (5)–(7). When child heterogeneity is uncorrelated but maternal heterogeneity is correlated with X_1 and W_1 , consistent but inefficient estimates for β can be obtained from the OLS regression of Equation (5), whereas consistent estimates for β and γ can be obtained from the OLS regression of Equation (6). Hence, β are overidentified in this case. For consistent estimates for δ , we can consider the OLS regression of the following model:

$$\bar{y}_{i..} - \bar{x}_{i..}\hat{\beta} - \bar{w}_{i.}\hat{\gamma} = z_i\delta + m_i + \bar{c}_i + \bar{e}_{i..} \quad (12)$$

where consistent $\hat{\beta}$ and $\hat{\gamma}$ are obtained from within-mother regression (6) and z_i are assumed exogenous. Thus, given the panel data model with NECs, all parameters are naturally identified using internal instruments provided z_i are exogenous. Note that we obtain consistent estimates for σ_u^2 from the OLS regression of Equation (12) because the original variance components estimates σ_u^2 obtained from between-mother regression (7) will be inconsistent once we allow for (X_2, W_2) to be correlated with m_i .

Table II. Decomposition of total sum of squares of the key variables

Level of variation	Within-child	Within-sibling	Between-mother	Total
Applied problems	0.18	0.13	0.69	1.00
Passage comprehension	0.21	0.13	0.66	1.00
Log of birth weight	0.00	0.16	0.84	1.00
Gestational age	0.00	0.19	0.81	1.00
Fetal growth rate	0.00	0.15	0.85	1.00
Family income	0.12	0.00	0.87	1.00
Home environment	0.00	0.16	0.84	1.00

From the decomposition $x_{ijt} = (x_{ijt} - \bar{x}_{ij.}) + (\bar{x}_{ij.} - \bar{x}_{i..}) + \bar{x}_{i..}$, the total sum of squares of a variable x_{ijt} can be decomposed as follows:

$$\sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{t=1}^{T_i} (x_{ijt} - \bar{x})^2 = \sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{t=1}^{T_i} (x_{ijt} - \bar{x}_{ij.})^2 + \sum_{i=1}^M \sum_{j=1}^{N_i} T_i (\bar{x}_{ij.} - \bar{x}_{i..})^2 + \sum_{i=1}^M N_i T_i (\bar{x}_{i..} - \bar{x})^2$$

where $\bar{x} \equiv \frac{1}{MN_i T_i} \sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{t=1}^{T_i} x_{ijt}$ and the three sums of squares on the right-hand side correspond to the sum of squares from the three hierarchical levels: within-child, within-mother, and between-mothers. Table II shows the decomposed total sum of squares for some of the key variables as a proportion of the total sum of squares. It shows that over 65% of the total variation in the test scores is explained by the between-mothers variation. This is the case for the other key variables too, which highlights that a substantial amount of information in the data set will be lost if we use the MFE estimation where the entire between-mother variation is inadvertently discarded.

6. MAIN RESULTS

We begin by presenting the NECGLS estimates of Equation (2) where the independent variable of interest is log of birth weight. The estimates in Table III suggest that birth weight has strong and positive effects on test scores. However, the chi-squared statistics from a pair of Hausman tests, which are shown at the bottom of Table III, suggest inconsistency of the NECGLS estimates. For applied problems, the chi-squared statistics are large enough to reject the null hypothesis of uncorrelated child and mother endowment at 1% level of significance while inconsistency appears to be less severe for passage comprehension. In particular, the individual t statistics from the Hausman tests suggest that family income and home environment are the major sources of endogeneity regardless of the alternative hypotheses (i.e., either against CFE or MFE) and the academic achievement tests.

However, notice that we cannot pin down the unobserved heterogeneity that is correlated with family income and home environment based on the two Hausman tests in Table III because the rejection of the Hausman test between the NECGLS and the CFE estimation only indicates the correlation in child and mother heterogeneity jointly. Also, the Hausman test between NECGLS and MFE may not be valid if the child heterogeneity is correlated with the regressors. Therefore, we perform the Hausman test of the correlated child heterogeneity based on the CFE and MFE estimates. The test results are presented in Table IV. The null hypothesis of uncorrelated child heterogeneity cannot be rejected at the 10% significance level for both Woodcock–Johnson achievement tests. Overall, the results from three Hausman tests indicate that the major source of endogeneity in the NECGLS estimation is the correlated maternal heterogeneity, which is consistent with the extensive use of the MFE estimation in the literature. Therefore, we maintain the identifying assumption that the unobserved child heterogeneity is uncorrelated and all regressors, except for the maternal component of family income and home environment, are exogenous.

Table III. Estimated effects of infant health on test scores, nested error-component generalized least squares (NECGLS) estimates

WJ achievement tests	Applied problems				Passage comprehension			
	NECGLS		Hausman test		NECGLS		Hausman test	
	bhat	<i>t</i> stat.	vs. CFE	vs. MFE	bhat	<i>t</i> stat.	vs. CFE	vs. MFE
Family income	1.841	4.60	-4.20	-3.40	2.213	5.17	-2.58	-2.38
Mother working	-0.258	-0.51	-0.56	0.00	-0.115	-0.21	0.17	0.29
Child age	-0.030	-7.42	2.70	2.26	-0.084	-16.55	1.69	1.41
Log of birth weight	4.963	3.94		-0.17	3.447	2.70		0.73
Female	-0.523	-1.08		0.45	2.496	4.99		-0.67
Maternal age at child birth <20	-3.184	-3.29		-0.79	-2.357	-2.35		-0.40
Maternal age at child birth >35	2.042	1.99		0.35	2.195	2.11		-1.59
Mother single at child birth	-1.398	-1.82		-0.50	-1.927	-2.44		-0.20
Second born	-1.166	-2.23		-1.47	-1.274	-2.33		-1.97
Third born	-1.947	-2.65		0.15	-3.650	-4.85		-1.03
Fourth born	-0.084	-0.07		1.11	-3.692	-3.23		0.13
Fifth or more born	-5.385	-3.04		-0.93	-4.065	-2.26		0.67
Home environment	0.330	4.02		-1.95	0.356	4.24		-2.73
African American	-8.066	-10.39			-4.038	-5.11		
Latino	-4.626	-3.20			-5.320	-3.71		
Other	-1.331	-0.77			-1.904	-1.10		
Mother's education	1.108	7.34			0.808	5.21		
<i>N</i>	4609				4106			
σ_o^2 : Observation level	113.61				124.74			
σ_c^2 : Child level	16.62				12.76			
σ_m^2 : Mother level	88.43				82.67			
Chi-squared statistic			18.18	30.76			6.65	18.90
(<i>p</i> -value)			(0.000)	(0.004)			(0.084)	(0.126)

The regressions are based on the 1997, 2002–2003, and 2007 PSID-CDS data sets.
CFE, child fixed effects; MFE, mother fixed effects.

Table IV. The Hausman test of the correlated child endowment

WJ achievement tests	Applied problems				Passage comprehension			
	CFE	MFE	Diff.	<i>t</i> stat.	CFE	MFE	Diff.	<i>t</i> stat.
Family income	-0.244	-0.088	-0.156	0.94	0.593	0.589	0.004	0.02
Mother working	-0.460	-0.206	-0.254	1.62	-0.037	0.064	-0.101	0.45
Child age	-0.024	-0.023	0.000	0.40	-0.078	-0.079	0.000	0.13
Chi-squared	6.15				0.21			
(<i>p</i> value)	(0.105)				(0.975)			

All regressions include the following set of regressors that is suppressed in the table: log of birth weight, binary indicator for preterm birth, home environment, a set of indicators for mother being single at birth, mother's age at child's birth, and birth order of children. Both covariance matrices are based on the common estimated disturbance variance from the MFE.

CFE, child fixed effects; MFE, mother fixed effects.

Under these identifying assumptions, we estimate Equation (8) using 2SLS estimation where the list of instruments is equivalent to the instrument set *C* in the previous section. Given that mother heterogeneity is the only source of endogeneity, our three-level panel data have two extra dimensions that can be utilized to obtain internal instruments: within-child (or over-time) and within-sibling. Hence, the identifying information for each endogenous regressor comes naturally from its own variations: within-child variation for family income and within-sibling variation for home environment, as can be seen in Table II. Notice that this would not have been possible in a conventional two-dimensional panel model. Moreover, because of the absence of time-invariant

Table V. Estimated effects of infant health on test scores, nested error-component two-stage least squares (NEC2SLS) estimates

WJ achievement tests	Applied problems			Passage comprehension		
	NEC2SLS		Hausman test	NEC2SLS		Hausman test
	bhat	<i>t</i> stat.	<i>t</i> stat.	bhat	<i>t</i> stat.	<i>t</i> stat.
Family income	-0.110	-0.18	0.08	0.692	0.93	-0.37
Mother working	0.008	-0.02	-0.46	0.130	0.23	-0.12
Child age	-0.023	-5.37	0.02	-0.077	-14.10	-0.36
Log of birth weight	5.394	4.23	-0.11	3.808	2.94	0.53
Female	-0.562	-1.15	0.71	2.499	4.94	-0.60
Maternal age at child birth <20	-3.432	-3.51	-0.97	-2.542	-2.51	-0.41
Maternal age at child birth >35	2.380	2.27	0.28	2.395	2.25	-1.79
Mother single at child birth	-2.207	-3.06	0.70	-2.715	-3.16	-0.67
Second born	-1.408	-2.63	-1.47	-1.459	-2.61	-1.72
Third born	-2.394	-3.18	0.17	-3.984	-5.18	-0.94
Fourth born	-0.386	-0.34	1.88	-3.911	-3.38	0.35
Fifth or more born	-5.781	-3.23	-1.04	-4.329	-2.38	0.79
Home environment	0.115	0.91	-0.85	0.140	1.05	-1.83
African American	-9.267	-11.06		-5.075	-5.81	
Latino	-5.300	-3.57		-5.897	-4.00	
Other	-1.672	-0.96		-2.280	-1.29	
Mother's education	1.375	8.27		1.027	5.82	
N	4609			4106		
σ_e^2 : Observation level	113.88			125.20		
σ_c^2 : Child level	16.27			12.63		
σ_m^2 : Mother level	93.63			88.18		
Chi-squared statistic				16.84		
(<i>p</i> -value)				(0.207)		
				10.90		
				(0.619)		

The regressions are based on the 1997, 2002–2003, and 2007 PSID-CDS data sets.

endogenous variables in our model, we do not need the order condition derived by Hausman and Taylor (1981) that the number of time-varying exogenous variables is no less than the number of endogenous time-invariant regressors.

Table V presents the estimates on our baseline model from the NEC2SLS estimation described in the previous texts. After addressing the endogeneity arising from the maternal component of family income and home environment, we find a positive effect of birth weight on test scores of similar magnitude as the NECGLS estimates. The magnitude of estimated birth weight effect is somewhat larger for Applied Problems. Notice that the chi-squared statistics for the Hausman tests become small and insignificant as shown at the bottom of Table V. The estimated coefficients for the other covariates are as expected in general. We find that children born to young or single mothers perform worse on the achievement tests. Higher birth order leads to worse outcomes on the tests, which is consistent with the literature on the birth-order effects. After controlling for the endogeneity in family income and home environment, these variables are no longer statistically significant, but the effects of all mother-specific variables including mother's education have become substantially bigger compared with the NECGLS estimates for both test scores. Furthermore, we estimate a large Black-White gap in test scores and a highly significant effect of maternal education. The parameters for all these time-invariant, mother-specific regressors would not be feasible to estimate within the MFE estimation framework.

Table VI summarizes the coefficient estimates on birth weight and the fetal growth rate when alternative estimation methods are used. Panel A shows significant effects of birth weight on test scores across different estimation methods. The estimated coefficient of 5.394 implies that a 10% increase in birth weight is associated with a 0.036 standard deviation increase in math scores. The magnitude is comparable with but somewhat smaller than the size reported in Figlio *et al.* (2013). We find in panel B that these effects are concentrated over the low birth weight range. The size of the estimated effects of birth weight appears to be modest as

Table VI. Comparison between nested error-component generalized least squares (NECGLS), mother fixed effects (MFE), and nested error-component two-stage least squares (NEC2SLS) estimates

WJ achievement tests	Applied problems			Passage comprehension		
	NECGLS	NEC2SLS	MFE	NECGLS	NEC2SLS	MFE
<i>Panel A: birth weight</i>						
Log of birth weight	4.963*** (1.258)	5.394*** (1.276)	5.205** (2.440)	3.447*** (1.279)	3.808*** (1.297)	4.816** (2.309)
<i>Panel B: birth weight spline</i>						
Birth weight (<2.5 kg)	5.275*** (1.513)	5.404*** (1.518)	5.820** (2.369)	4.411*** (1.540)	4.534*** (1.545)	5.369** (2.558)
Birth weight (≥2.5 kg)	0.683 (0.555)	0.777* (0.557)	0.261 (0.933)	0.263 (0.567)	0.341 (0.569)	0.374 (1.021)
<i>Panel C: nutritional intake in utero</i>						
Fetal growth rate	0.056*** (0.021)	0.059*** (0.021)	0.031 (0.036)	0.034 (0.022)	0.036* (0.022)	0.028 (0.039)
Preterm birth	-1.170 (1.035)	-1.279 (1.044)	-2.229 (1.550)	-1.502 (1.072)	-1.664 (1.080)	-2.665 (1.735)
<i>N</i>	4609	4609	3004	4106	4609	2664

Standard errors are in parentheses. All regressions include the same set of covariates as in Table III. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VII. Interaction effects with mother's education

WJ achievement tests	Applied problems		Passage comprehension	
	(1)	(2)	(3)	(4)
<i>Panel A: birth weight</i>				
Log of birth weight	5.394*** (1.276)	11.632* (6.940)	3.808*** (1.297)	10.198 (7.098)
Log of birth weight * Mother's education		-0.501 (0.541)		-0.558 (0.553)
<i>Panel B: nutritional intake in utero</i>				
Fetal growth rate	0.059*** (0.021)	0.289*** (0.104)	0.036* (0.022)	0.222** (0.108)
Fetal growth rate * Mother's education		-0.018** (0.008)		-0.015** (0.008)
Preterm birth	-1.286 (1.040)	-1.200 (1.040)	-1.660 (1.075)	-1.588 (1.076)
<i>N</i>	4609	4609	4106	4106

Standard errors are in parentheses. All regressions include the same set of covariates in Table III. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

increasing birth weight by 1 kg within the low birth weight range translates into a 0.33 standard deviation increase in test scores. In panel C, we consider the fetal growth rate as a proxy for nutritional intake *in utero* to test the fetal nutrition hypothesis. We find positive effect of the fetal growth rate on test scores, although the statistical significance of the estimated effect of the fetal growth rate is marginal for passage comprehension. The NEC2SLS estimates imply that a one standard deviation increase in the fetal growth rate is associated with a 0.056 standard deviation increase in math scores and a 0.034 standard deviation increase in reading scores.²³

²³For example, the effect of one standard deviation increase in the fetal growth rate on math scores is calculated as $0.059 \times 14.32 / 15 = 0.056$ where 14.32 is a standard deviation of the fetal growth rate.

Very similar estimates are obtained when the standardized fetal growth rate together with preterm dummy were used in the regression (see footnote 14).

The size of these estimates is larger than that in Torche and Echevarría (2011) and is comparable with the estimated effect of mother's exposure to start of Ramadan in the first month of pregnancy (Almond *et al.*, 2011). We also estimate a negative effect of preterm birth on test scores, but it is not statistically significant, and its size is small. Finally, we note that these estimates are similar across different estimation methods. Currie (2009) has noted that the effects of early health on later outcomes tend to be very similar across different estimation methods. We reach a similar conclusion with newer data and a more efficient estimation method.

To see whether a better-educated mother can buffer the negative consequences of low birth weight, we estimate a model that contains an interaction term between the infant health measures and mother's education. The estimated negative interaction terms in Table VII suggest that mother's education buffers the negative consequences from compromised fetal growth. This buffer effect of mother's education is statistically significant for the fetal growth rate but not for birth weight.

6.1. Implications for racial/ethnic disparity in test scores

In our analytical sample, we find substantial racial/ethnic disparity in test scores. For applied problems, the average Black-White differential is 13.7 (0.91 standard deviation in test score), and the Latino-White gap is 12.2 (0.81 standard deviation). For passage comprehension, the Black-White test score gap is smaller (0.67 standard deviation) whereas the Latino-White gap is substantially more (0.81 standard deviation).

In Table VIII, we report the estimated racial/ethnic gap in test scores before and after controlling for a set of covariates including birth weight. Column (1) in Table VIII reports the estimated coefficients for race/ethnicity dummies for applied problems without controlling for maternal education and birth weight but with all other controls such as child characteristics and socioeconomic status (cf. Table III). The estimated coefficients for the African American and the Latino are -10.2 and -9.2, respectively, which implies that socioeconomic characteristics explain about 25% of the raw disparity in the test scores for applied problems. In column (2), when we introduce maternal education in the specification, we find that the estimated coefficient for the African American dummy changes little (-9.7) but that for the Latino is reduced by almost half (-5.37). This is because of a relatively lower average level of education for the Latino mothers (10.2 years) compared with the White (13.7 years) and the African American mothers (12.5 years) in our sample. In column (3), we add birth weight in the specification and find, interestingly, that the estimated coefficients for race/ethnicity dummies stay almost the same. These results imply that infant health as measured by birth weight does not constitute a pathway for the observed racial/ethnic disparity in test scores. In columns (5) through (7), we report similar experiments for passage comprehension, and the pattern of the results is seen to be very similar.

In columns (4) and (8), we introduce two additional regressors by interacting mother's education with two racial dummies for the African American and the Latino. That way, we allow gradients for mother's education to differ across race/ethnicity. The estimated coefficient for the interaction term for the African American is found to be negative and highly significant for both test scores. For the Latino, the estimated coefficients are qualitatively similar but smaller in size. These results imply that, compared with racial/ethnic minorities, the efficiency with which maternal education translates into gains in children's test score is higher among the non-Latino White families, but it is worse for the African American compared with the Latino families. Also, the results imply that the racial/ethnic disparity in test scores increases with mother's education. Although not included in the table, a similar regularity is observed with respect to family income. Thus, apart from the White-Latino gap in mother's education, a major reason for the observed racial/ethnic disparity in test scores can be attributed to the fact that returns to mother's education in producing child test scores is smaller among minority families compared with White families with comparable socioeconomic status.

Table VIII. Role of fetal growth rate in determining racial disparity in test scores, nested error-component two-stage least squares estimates

WJ achievement tests	Applied problems			Passage comprehension				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
African American	-10.169*** (0.866)	-9.706*** (0.838)	-9.267*** (0.838)	4.476 (4.082)	-5.565*** (0.891)	-5.368*** (0.866)	-5.075*** (0.874)	2.795 (4.194)
Latino	-9.169*** (1.503)	-5.369*** (1.492)	-5.300*** (1.485)	3.031 (4.761)	-8.706*** (1.475)	-5.936*** (1.461)	-5.897*** (1.475)	-0.537 (4.749)
Other	-1.645 (1.718)	-1.852 (1.754)	-1.672 (1.746)	-1.328 (1.743)	-2.228 (1.705)	-2.415 (1.766)	-2.280 (1.763)	-2.080 (1.764)
Maternal education		1.385*** (0.167)	1.375*** (0.166)	1.798*** (0.212)		1.031*** (0.177)	1.027*** (0.177)	1.276*** (0.225)
Log of birth weight			5.394*** (1.276)	5.476*** (1.483)			3.808*** (1.297)	3.270*** (1.514)
African American * Mother's education				-1.053*** (0.308)				-0.603*** (0.316)
Latino * Mother's education				-0.670* (0.420)				-0.440 (0.418)
N	4609	4609	4609	4609	4106	4106	4106	4106

Standard errors are in parentheses. All regressions include the same set of covariates as in Table III. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7. IS MATERNAL EDUCATION ENDOGENOUS?

In this section, we address a potential concern about endogeneity arising from mother-specific, time-invariant variables. Our strategy to identify that the endogenous variables based on the individual t statistics from the Hausman tests have a limitation that the t statistics are not available for the mother-specific variables such as mother's education and race. Indeed, Carneiro *et al.* (2013) provided some evidence that the mother's education is endogenous in academic achievement production, noting that both mother's education and child's academic achievement are determined by mother's ability. In principle, if mother's education (any of the mother-specific variables) is endogenous, all the coefficient estimates of the NEC2SLS will be inconsistent. In this case, the endogeneity arising from the mother-specific regressors should have been detected indirectly by the Hausman tests against the MFE estimation even though t statistics are not available for those regressors. In order to be doubly sure, in this section, we allow for mother's education to be endogenous to check the robustness of our previous NEC2SLS estimates.

As before, we partition $Z = (Z_1, Z_2)$ where Z_2 is endogenous mother's education and Z_1 are all other mother-specific, time-invariant regressors. Notice that, for the endogenous mother-specific, time-invariant regressor Z_2 , a valid instrument cannot be found within Z_2 , which varies over a single dimension. Now the situation is analogous to the Hausman and Taylor (1981) framework requiring the necessary condition for identification that the number of time-varying exogenous variables be no less than the number of endogenous time-invariant endogenous variables. Provided that this order condition is met, the consistent NEC2SLS estimates are obtained by using the smaller set of instruments:

$$D = (Q_1X, Q_2X, PX_1, Q_2W, PW_1, Z_1)$$

than the instrument set C now that the endogenous Z_2 should be excluded from the instrument set. In our case, the order condition is clearly satisfied because the number of exogenous regressors having extra dimensions beyond between-mother is greater than one.²⁴ That is, Q_1X_1 , Q_2X_1 , and Q_2W_1 provide a sufficient number of instruments for Z_2 . One potential concern in this estimation strategy is that these internal instruments may be weak. However, we find that many instruments have sufficient explanatory power in the first-stage regression of endogenous mother's education. The F statistic is 15.42, which is greater than 10, the cutoff for weak instruments suggested by Stock and Yogo (2005). To obtain additional identifying information, we also turn to the following covariance restrictions that have been proposed by Lewbel (2012) and used in many applications. More specifically, we will assume the following:

$$Cov(S, u_{ijt}) = 0, \quad (13)$$

$$Cov(S, v) = 0, \quad (14)$$

$$Cov(T, u_{ijt}v) = 0 \quad (15)$$

where $S = (X, W, Z_1)$, $T = (X_1, W_1, Z_1)$ and v denote the error term for reduced-form equation for mother's education. Under these assumptions, $(T - \bar{T})\hat{v}$ give a set of valid instruments for mother's education (Lewbel, 2012). \hat{v} can be obtained from the 2SLS regression of mother's education Z_2 on S where internal instruments are used for (X_2, W_2) .

²⁴Note that the estimates from the ordinary least squares regression of Equation (12) are inconsistent now that Z_2 is correlated with mother heterogeneity. However, we can consistently estimate δ by two-stage least squares estimation of Equation (12) if (X_1, W_1) provide a sufficient number of instruments for Z_2 . Note also, with mother's education endogenous, we reestimate the variance component σ_m^2 using this IV procedure.

Table IX. Robustness check: endogenous mother's education

WJ achievement tests	Applied problems			Passage Comprehension		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: birth weight</i>						
Log of birth weight	5.394*** (1.276)	5.394*** (1.279)	5.374*** (1.276)	3.808*** (1.297)	3.743*** (1.311)	3.803*** (1.305)
<i>Panel B: nutritional intake in utero</i>						
Fetal growth rate	0.059*** (0.021)	0.059*** (0.021)	0.058*** (0.021)	0.036* (0.022)	0.036 (0.022)	0.035 (0.022)
Preterm birth	-1.286 (1.040)	-1.280 (1.049)	-1.326 (1.045)	-1.660 (1.075)	-1.579 (1.095)	-1.709 (1.086)
<i>N</i>	4609	4609	4609	4106	4106	4106
Mother's education endogenous	No	Yes	Yes	No	Yes	Yes
Own instruments	Yes	Yes	Yes	Yes	Yes	Yes
Covariance restriction	No	No	Yes	No	No	Yes

Standard errors are in parentheses. All regressions include the same set of covariates in Table III. Breusch–Pagan tests strongly support the relevance of instruments from covariance restrictions in all cases. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IX provides comparison of the estimates under different identification assumptions. The original NEC2SLS estimates in Table VI are replicated in columns (1) and (4) for comparison, and all estimates in the other columns in Table IX are obtained when mother's education is allowed to be endogenous. Columns (2) and (5) show that the estimates obtained using only the internal instruments, without the covariance restrictions (13)–(15) imposed, are robust to endogenous mother's education. Columns (3) and (6) report the estimates when those covariance restrictions are imposed to increase the power of instruments. The Breusch–Pagan test indicates the instruments created by the covariance restrictions are highly relevant. Again, the NEC2SLS estimates remain robust under different sets of instruments, although the estimated effect of the fetal growth rate on passage comprehension is no longer statistically significant. The null hypotheses of uncorrelated maternal heterogeneity cannot be rejected in the Hausman tests in all cases. Overall, Table IX shows that the 2SLS estimates are not sensitive to the identification assumption on mother's education, and it can safely be treated as exogenous in our context.

8. CONCLUSION

In this paper, we investigate the association between birth weight and cognitive outcomes in childhood by using the fetal growth rate as a measure of nutritional intake *in utero*. In particular, we develop the NEC2SLS estimation method that can overcome many of the limitations associated with the use of MFE estimation. We also suggest a simple instrumental variable algorithm that does not depend on instruments external to our multilevel model. Using NEC2SLS, we find a positive and statistically significant effect of the fetal growth rate on academic achievement of children. Our finding is consistent with those found in some recent studies using relatively large samples (for example, Torche and Echevarría, 2011; Figlio *et al.*, 2013). The estimated effect of birth weight, however, is concentrated over the low birth weight range, and its magnitude is somewhat smaller than that in Figlio *et al.* (2013). Overall, our results imply that cognitive gains in childhood from better nutritional intake *in utero* may constitute a pathway through which birth weight determines adult outcomes, such as education and earnings.

The NEC2SLS approach allows us to investigate the effects of birth weight and time-invariant variables such as mother's education and race/ethnicity on academic achievement in childhood, avoiding a critical limitation of the fixed effect estimator. We find that mother's education has an important mediating effect of low birth weight; allowing it to be endogenous does not reduce its effect even when our optimal instrument set is supplemented with additional instruments coming from heteroskedastic covariance restrictions (Lewbel, 2012). We also find that low birth weight does not contribute much to explain the racial disparities in test scores.

Our study is not without limitations. First, the benefits over the MFE estimation are obtained at the possible cost of imposing an assumption that only a subset of the regressors is endogenous. However, using a battery of Hausman-type exogeneity tests, we found no credible evidence against these identifying assumptions. Second, our model does not account for potentially different parental investment across siblings (Rosenzweig and Wolpin, 1988; Behrman *et al.*, 1994) even though the evidence has not been clear whether the differential parental investment compensates or reinforces the lack of initial endowment of a particular child (Datar *et al.*, 2010; Hsin, 2012; Lynch and Brooks, 2013). If the parental investment is compensating, then our estimates should be interpreted as lower bounds. Finally, we did not examine the growth in child academic achievement at a particular age, whereas many studies focus on the age profile of the effects of birth weight, gender, or race on educational outcomes (Fryer Jr and Levitt, 2004; Bond and Lang, 2013). Addressing these limitations will be left for future research.

Research was supported by the National Institute on Minority Health and Health Disparities, National Institutes of Health (grant number 1 P20 MD003373). The content is solely the responsibility of the authors and does not represent the official views of the National Institute on Minority Health and Health Disparities or the National Institutes of Health. This is a revised version of our paper presented at the 22nd European Workshop on Econometrics & Health Economics in Rotterdam. We thank our discussant Stefan Holst Milton Bache, Maarten Lindeboom, Nigel Rice and other participants of the workshop, Badi Baltagi and an anonymous referee for many useful comments. However, the responsibility for any remaining errors is ours.

REFERENCES

- Akresh R, Akresh IR. 2011. Using achievement tests to measure language assimilation and language bias among the children of immigrants. *Journal of Human Resources* **46**(3): 647–667.
- Almond D, Chay KY, Lee DS. 2005. The costs of low birth weight. *The Quarterly Journal of Economics* **120**(3): 1031–1083.
- Almond D, Currie J. 2011. Killing me softly: the fetal origins hypothesis. *The Journal of Economic Perspectives* **25**: 153–172.
- Almond D, Mazumder B. 2011. Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics* **3**(4): 56–85.
- Almond D, Mazumder B, van Ewijk R (2011), 'Fasting during pregnancy and children's academic performance', Working Paper 17713, National Bureau of Economic Research, Cambridge, MA.
- Asher C. 1946. The prognosis of immaturity. *British Medical Journal* **1**(4455): 793–796.
- Baltagi BH, Heun Song S, Cheol Jung B. 2001. The unbalanced nested error component regression model. *Journal of Econometrics* **101**(2): 357–381.
- Barker DJP. 1966. Low intelligence. Its relation to length of gestation and rate of foetal growth. *British Journal of Preventive & Social Medicine* **20**(2): 58–66.
- Barker DJP. 1995. Fetal origins of coronary heart disease. *British Medical Journal* **311**(6998): 171–174.
- Barker DJP. 1998. *Mothers, Babies and Health in Later Life*. Churchill Livingstone: Edinburgh.
- Behrman JR, Rosenzweig MR. 2004. Returns to birthweight. *Review of Economics and Statistics* **86**(2): 586–601.
- Behrman JR, Rosenzweig MR, Taubman P. 1994. Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of Political Economy* **102**(6): 1131–1174.
- Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand K. 2002. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *The Journal of the American Medical Association* **288**(6): 728–737.
- Black SE, Devereux PJ, Salvanes KG. 2005. The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics* **120**(2): 669–700.
- Black SE, Devereux PJ, Salvanes KG. 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics* **122**(1): 409–439.
- Boardman JD, Powers DA, Padilla YC, Hummer RA. 2002. Low birth weight, social factors, and developmental outcomes among children in the united states. *Demography* **39**(2): 353–368.
- Bond TN, Lang K (2013), 'The Black-White education-scaled test-score gap in grades k-7', Working Paper 19243, National Bureau of Economic Research, Cambridge, MA.
- Boomsma D, van Beijsterveldt C, Rietveld M, Bartels M, Van Baal G. 2001. Genetics mediate relation of birth weight to childhood IQ. *British Medical Journal* **323**(7326): 1426–1427.

- Breslau N, Johnson EO, Lucia VC. 2001. Academic achievement of low birthweight children at age 11: the role of cognitive abilities at school entry. *Journal of Abnormal Child Psychology* **29**(4): 273–279.
- Breusch TS, Mizon GE, Schmidt P. 1989. Efficient estimation using panel data. *Econometrica* **57**(3): 695–700.
- Brown S, Ortiz-Núñez A, Taylor K. 2012. Parental risk attitudes and children's academic test scores: evidence from the US Panel Study Of Income Dynamics. *Scottish Journal of Political Economy* **59**(1): 47–70.
- Caldwell BM, Bradley RH. 1984. *Administration Manual: Home Observation for Measurement of the Environment*. University of Arkansas: Little Rock, AR.
- Campbell FA, Ramey CT, Pungello E, Sparling J, Miller-Johnson S. 2002. Early childhood education: young adult outcomes from the abecedarian project. *Applied Developmental Science* **6**(1): 42–57.
- Carneiro P, Meghir C, Patey M. 2013. Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association* **11**(s1): 123–160.
- Case A, Fertig A, Paxson C. 2005. The lasting impact of childhood health and circumstance. *Journal of health economics* **24**(2): 365–389.
- Christensen K, Petersen I, Skytthe A, Herskind AM, McGue M, Bingley P. 2006. Comparison of academic performance of twins and singletons in adolescence: follow-up study. *British Medical Journal* **333**(7578): 1095–1097.
- Currie J. 2009. Healthy, wealthy, and wise: is there a causal relationship between child health and human capital development? *Journal of Economic Literature* **47**(1): 87–122.
- Currie J. 2011. Inequality at birth: some causes and consequences. *American Economic Review* **101**(3): 1–22.
- Currie J, Hyson R. 1999. Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. *American Economic Review* **89**(2): 245–250.
- Currie J, Moretti E. 2007. Biology as destiny? short-and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics* **25**(2): 231–263.
- Currie J, Vogl T. 2013. Early-life health and adult circumstance in developing countries. *Annual Review Economics* **5**(1): 1–36.
- Datar A, Kilburn MR, Loughran DS. 2010. Endowments and parental investments in infancy and early childhood. *Demography* **47**(1): 145–162.
- Elardo R, Bradley RH. 1981. The home observation for measurement of the environment (home) scale: a review of research. *Developmental Review* **1**(2): 113–145.
- Figlio DN, Guryan J, Karbownik K, Roth J (2013), 'The effects of poor neonatal health on children's cognitive development', Working Paper 18846, National Bureau of Economic Research.
- Fletcher JM. 2011. The medium term schooling and health effects of low birth weight: evidence from siblings. *Economics of Education Review* **30**(3): 517–527.
- Fryer Jr RG, Levitt SD. 2004. Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics* **86**(2): 447–464.
- Fuller WA, Battese GE. 1973. Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association* **68**(343): 626–632.
- Hack M, Flannery DJ, Schluchter M, Cartar L, Borawski E, Klein N. 2002. Outcomes in young adulthood for very-low-birth-weight infants. *New England Journal of Medicine* **346**(3): 149–157.
- Hausman JA. 1978. Specification tests in econometrics. *Econometrica* **46**: 1251–1271.
- Hausman JA, Taylor WE. 1981. Panel data and unobservable individual effects. *Econometrica* **49**: 1377–1398.
- Hsin A. 2012. Is biology destiny? Birth weight and differential parental treatment. *Demography* **49**(4): 1385–1405.
- Johnson RC, Schoeni RF. 2011. The influence of early-life events on human capital, health status, and labor market outcomes over the life course. *The BE journal of economic analysis & policy* **11**(3): Article 3.
- Kim JS, Frees EW. 2006. Omitted variables in multilevel models. *Psychometrika* **71**(4): 659–690.
- Kinal T, Lahiri K. 1993. On the estimation of simultaneous-equations error-components models with an application to a model of developing country foreign trade. *Journal of Applied Econometrics* **8**(1): 81–92.
- Lewbel A. 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics* **30**(1): 67–80.
- Lindeboom M, Portrait F, Van den Berg GJ. 2010. Long-run effects on longevity of a nutritional shock early in life: the dutch potato famine of 1846–1847. *Journal of Health Economics* **29**(5): 617–629.
- Lynch JL, Brooks R. 2013. Low birth weight and parental investment: do parents favor the fittest child? *Journal of Marriage and Family* **75**(3): 533–543.
- Mather N. 1991. *An Instructional Guide to the Woodcock–Johnson Psycho-Educational Batter–Revised*. Wiley: New York.
- Mohanty LL, Raut LK. 2009. Home ownership and school outcomes of children: evidence from the PSID Child Development Supplement. *American Journal of Economics and Sociology* **68**(2): 465–489.
- Morgane PJ, Austin-LaFrance R, Bronzino J, Tonkiss J, Diaz-Cintra S, Cintra L, Kemper T, Galler JR. 1993. Prenatal malnutrition and development of the brain. *Neuroscience & Biobehavioral Reviews* **17**(1): 91–128.
- Murnane RJ, Willett JB, Levy F. 1995. The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics* **77**(2): 251–266.

- Neelsen S, Stratmann T. 2011. Effects of prenatal and early life malnutrition: evidence from the greek famine. *Journal of Health Economics* **30**(3): 479–488.
- Oreopoulos P, Stabile M, Walld R, Roos LL. 2008. Short-, medium-, and long-term consequences of poor infant health an analysis using siblings and twins. *Journal of Human Resources* **43**(1): 88–138.
- Petersen I, Jensen VM, McGue M, Bingley P, Christensen K. 2009. No evidence of genetic mediation in the association between birthweight and academic performance in 2413 danish adolescent twin pairs. *Twin Research and Human Genetics* **12**(06): 564–572.
- Phillips M, Brooks-Gunn J, Duncan GJ, Klebanov P, Crane J. 1998. Family background, parenting practices, and the black-white test score gap. Jencks C, Phillips M (eds). *The Black-white test score gap*. Brookings Institution Press: Washington, D.C; 103–145.
- Record R, McKeown T, Edwards J. 1969. The relation of measured intelligence to birth weight and duration of gestation. *Annals of Human Genetics* **33**(1): 71–79.
- Rosenzweig MR, Wolpin KI. 1988. Heterogeneity, intrafamily distribution, and child health. *Journal of Human Resources* **23**: 437–461.
- Royer H. 2009. Separated at birth: US twin estimates of the effects of birth weight. *American Economic Journal: Applied Economics* **1**(1): 49–85.
- Sørensen HT, Sabroe S, Olsen J, Rothman KJ, Gillman MW, Fischer P. 1997. Birth weight and cognitive function in young adult life: historical cohort study. *British Medical Journal* **315**(7105): 401–403.
- Stock JH, Yogo M. 2005. Testing for weak instruments in linear iv regression. Andrews DWK, Stock JH (eds). *Identification and inference for econometric models*, Cambridge University Press: Cambridge, MA; 80–108.
- Suloway FJ. 2007. Psychology. Birth order and intelligence. *Science (New York, NY)* **316**(5832): 1711–1712.
- Todd PE, Wolpin KI. 2003. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* **113**(485): F3–F33.
- Torche F, Echevarría G. 2011. The effect of birthweight on childhood cognitive development in a middle-income country. *International Journal of Epidemiology* **40**(4): 1008–1018.
- Van Ewijk R. 2011. Long-term health effects on the next generation of ramadan fasting during pregnancy. *Journal of Health Economics* **30**(6): 1246–1260.
- Woodcock RW, Johnson MB. 1989. *Woodcock–Johnson Tests of Achievement*. DLM Teaching Resources: Itasca, IL.