The Impact of Family Income on Child Achievement:

Evidence from the Earned Income Tax Credit: Comment

Samuel M. Lundstrom

University of California, Irvine

July, 2015

Abstract:

In a recent article in *The American Economic Review*, Dahl and Lochner (2012) use an instrumental variables (IV) strategy to estimate the causal effect of family income on children's math and reading achievement. Their identification derives from the large, nonlinear changes in the Earned Income Tax Credit (EITC) in the late 1980s and early 1990s. Their baseline IV estimates imply that a $1,000 increase in income raises combined math and reading test scores by 6 percent of a standard deviation. I document a critical error in Dahl and Lochner's code. In their construction of family income they fail to include EITC income. This is a critical omission considering that the correlation between EITC income and the instrument is, presumably, the principal source of income variation in their strategy. Correcting this error reduces the IV estimates by 32 percent. More importantly, identifying this error makes apparent that the instrument does not operate in the way that Dahl and Lochner suppose. I find evidence suggesting that identification does not rely solely on government changes in the EITC schedule. In particular, when I estimate models for time periods where the largest EITC expansions are excluded, I obtain statistically significant first stage estimates, and large and statistically significant reduced form estimates of either sign. Furthermore, even when instrument variation arising from government changes in the EITC schedule is preserved, I find evidence suggesting that the estimates are sensitive to which time periods that variation comes from. The evidence presented here raises questions regarding the validity of Dahl and Lochner's instrument.
The endogeneity of income has long prevented researchers from forming a consensus regarding the causal effect of income on child achievement (Haveman and Wolfe 1995; Duncan and Brooks-Gunn 1997; Mayer 1997). In particular, children in poor families are likely to have adverse home environments which might also affect their achievement levels. Moreover, changes in family circumstances over time such as moving, illness, or parental job loss or promotion may affect both family income and parenting behavior. Researchers have employed fixed effects strategies to eliminate biases caused by permanent family or child characteristics (Duncan et. al 1998; Blau 1999; Levy and Duncan 1999). While these fixed effects estimates represent a significant step forward from basic ordinary least squares (OLS) estimates, they do not control for endogenous transitory shocks (e.g., parental job loss or promotion, family illness, residential moves) and they likely suffer from severe attenuation bias since growth in income is usually measured noisily.

Dahl and Lochner (2012; hereafter, DL) outline an instrumental variables (IV) strategy to estimate the causal effect of family income on child achievement that intends to eliminate bias due to the omission of permanent and transitory achievement determinants that are correlated with income, as well as bias due to measurement error. Their strategy uses the expansions of the Earned Income Tax Credit (EITC) in the late 1980s and 1990s as an exogenous source of income variation for American families. As illustrated in Figure (1), from 1987 to 1999 both the maximum benefit amount increased, as did the range of qualifying incomes—particularly for families with two+ children. To the extent that income affects child achievement, DL expect to observe relative improvements in the test scores of children from families benefitting the most from the EITC expansions. Their analysis uses panel data on nearly 4,500 children matched to their mothers in the Children of the National Longitudinal Survey of Youth (NLSY). These data contain both a rich set of income and demographic measures, as well as up to five repeated measures of cognitive test scores per child, taken every other year. DL's instrumental variables (IV) estimates suggest that current
income has a significant effect on a child's math and reading achievement—a $1,000 increase in family income raises math and reading test scores by about 6 percent of a standard deviation.

While replicating DL’s study, I discovered several coding errors which, when corrected, reduce the IV estimates by 32 percent. The most important of these errors relates to DL's construction of family income. In their construction of family income DL fail to add federal tax credits, and they include state tax credits twice. Since the EITC is a federal credit, this means that DL's measure of family income does not include the EITC. This is a critical omission considering that the correlation between EITC income and changes in the EITC schedule is, presumably, the primary source of income variation in their strategy. When these mistakes are corrected the IV estimates fall by 32 percent. More importantly, the discovery of this error begs the question of how DL's instrument operates, given its strong correlation with the non-EITC components of family income.

I find evidence suggesting that identification does not rely solely on government changes in the EITC schedule, as supposed by Dahl and Lochner. When I estimate models for time periods where the largest EITC expansions are excluded, I obtain statistically significant first stage estimates, and large and statistically significant reduced form estimates of either sign. More specifically, when the sample is restricted to the years 1987-1993 (a period prior to the largest EITC expansions), I find that a $1,000 increase in simulated EITC income (DL's instrument) decreases combined math and reading test scores for young kids (age<12) by 36.4 percent of a standard deviation. When the sample is restricted to the years 1995-1999 (a period following the largest EITC expansions), I find that a $1,000 increase in simulated EITC income increases the combined math and reading test scores of black or Hispanic and male children by 14.0 and 22.1 percent of a standard deviation respectively. Moreover, even when instrument variation arising from government changes in the

---

1 For a detailed explanation of the error refer to Appendix 1.
EITC schedule is preserved, I find evidence suggesting that the estimates are sensitive to which time periods that variation comes from. The implication is that DL's IV estimates are likely driven by a spurious correlation between the instrument and child achievement shocks (even if the source of this spurious relationship is unclear). These findings raise questions regarding the validity of DL's instrument.

This paper proceeds as follows. In Section 1 I describe DL's study in detail, and I replicate their results. I show how both the magnitude and interpretation of their results are changed once family income is correctly measured. In Section 2 I take a step back from the coding error, and perform several tests to probe the plausibility of instrument validity. Section 3 concludes.

1 Summary of Dahl and Lochner's Study

In this section I provide a summary of Dahl and Lochner's study. I begin by describing the data and the estimation strategy. Next I replicate their results, and document the coding errors. I conclude this section by discussing the implications of the coding errors, and the questions that they raise in relation to instrument validity.

1.1 Data

Data are drawn from the Children of the National Longitudinal Survey of Youth (NLSY) matched to their mothers in the main NLSY. The NLSY is a panel survey that originally included a nationally representative sample of 12,686 men and women who were all 14 to 21 years of age on December 31, 1978. Annual interviews were completed with most of these respondents from 1979 through 1994 at which point there was a shift to biennial interviews. Beginning in 1986, the children born to NLSY female respondents have been assessed every two years. The estimation sample consists of child observations merged to their mothers for the survey years 1988-2000.
The NLSY provides many components of family income which are aggregated into three categories of pretax income: earned income, unearned income, and nontaxable income. Cognitive achievement is measured using standardized math and reading scores from the Peabody Individual Achievement Tests (PIAT). The tests measure a child’s ability in math, reading recognition (word recognition), and reading comprehension (the ability to derive meaning from printed words). The PIAT was administered to children ages five to fourteen\(^2\) and consists of a series of 84 questions which increase in difficulty from preschool to high school levels. To make PIAT tests more easily interpretable, DL create normalized test scores with a mean of zero and a standard deviation of one based on the random sample of test takers (i.e. excluding the poor, military, and minority oversamples). They also create a combined math-reading score by taking the average of the normalized math and reading scores. This is then renormalized to have a mean of zero and standard deviation of one in the random sample.

The main sample is restricted to children who are observed in at least two consecutive sample years between 1988 and 2000 and who have valid PIAT scores, family background characteristics, and family income measures.\(^3\) The sample is limited to children whose mothers did not change marital status during two-year intervals when tests are measured, and observations are excluded if family income levels exceeds $100,000 or if there are very large changes in income. NLSY oversamples of poor white families and military families are excluded from the sample.\(^4\)

---

\(^2\) The PIAT tests were administered to children older than age 14 prior to 1994. Beginning in 1994 a separate survey for older children (15 and over) was created and the PIAT was not part of this survey. Because of this, about two percent of children took the PIAT tests after their 15\(^{th}\) birthday. These observations are retained in the analysis. Furthermore, many children ages 5-7 do not have valid scores for the reading recognition test, because their scores were out of range based on the national norming sample in 1968. See DL's online appendix or the NSLY User’s Guide for more details.

\(^3\) Family income is reported for the year prior to the survey. Therefore, while observations are drawn from the 1988 to 2000 surveys, income is reported for the years 1987 - 1999. In this analysis I always make reference to the income year, not the survey year.

\(^4\) The reason for dropping these observations is, apparently, that the interviews with the military oversamples were discontinued after 1984 and interviews with the poor white oversamples were discontinued after 1990. See the NLSY79 User Guide for details.
1.2 Methods

In this section I reproduce DL’s child achievement model. The model intends to capture how changes in family income (through such policies as the EITC) affect child achievement. Let \( x_i \) reflect observable permanent characteristics, \( \mu_i \) reflect unobserved permanent ability (i.e. a child fixed effect), \( w_{ia} \) reflect time-varying characteristics, and \( I_{ia} \) reflect total family income for child \( i \) at age \( a \). Using this notation, DL present a simple model for child outcome \( y_{ia} \) as follows:

\[
(1) \quad y_{ia} = x'_i \alpha + w'_i \beta + I_{ia} \delta + \mu_i + \epsilon_{ia}
\]

Least squares estimation of this model faces three major challenges: omitted variable bias due to unobserved permanent child characteristics, omitted variable bias due to unobserved transitory shocks (e.g. illness, moves, job loss or promotion), and bias due to measurement error in income. DL propose an estimation strategy that intends to eliminate bias from all three sources. They eliminate the threat of bias due to unobserved child fixed effects by taking first differences, yielding the following equation:

\[
(2) \quad \Delta y_{ia} = x'_i \alpha + \Delta w'_i \beta + \Delta I_{ia} \delta + \Delta \epsilon_{ia}
\]

where permanent characteristics are allowed to affect the growth of child achievement (i.e. \( \alpha \equiv \alpha_a - \alpha_{a-1} \)). For \( I_{ia} \) DL intend to use total net family income (inclusive of EITC payments and net of other federal and state taxes and transfers). EITC income, \( X^{s_{ia}}(P_{ia}) \), is a function of pretax income, \( (P_{ia}) \), where the superscript \( s_{ia} \) denotes which schedule a child’s family is on (the EITC schedules differ based on the number of children in the household). They also net out other taxes, \( \tau^{s_{ia}}(P_{ia}) \).

Therefore, total net family income is given by:

---

5 In the most general model, Dahl and Lochner allow income to have long-run effects by incorporating lags into the model. However, the model that forms the basis for their main results excludes income lags. Since the effects of lagged income are not important for their main results, and in order to simplify the discussion, I exclude the lags in this model.
The coefficient of interest in Equation 2 is \( \delta_o \), which is intended to capture the causal effect of a contemporaneous change in family income on cognitive achievement. The main impediments to least squares estimation of Equation 2 are bias due to measurement error in income, and omitted variable bias due to unobserved transitory shocks. To address these issues DL employ an IV strategy that builds on Gruber and Saez (2002) and is loosely based on Feldstein (1995) and Currie and Gruber (1996). This IV strategy takes advantage of major changes in the EITC. As an instrument for \( \Delta I_{ia} \) they use:

\[
\Delta \chi^IV_a \equiv \chi^s_{ia-1} \left( \mathbb{E}[P_{i,a} | P_{i,a-1}] \right) - \chi^s_{a-1} (P_{i,a-1})
\]

where \( \mathbb{E}[P_{i,a} | P_{i,a-1}] \) is an estimate of pretax income given lagged pretax income. In practice, they regress pretax income on an indicator for positive lagged pretax income and a 5th-order polynomial in lagged pretax income in the calculation of \( \mathbb{E}[P_{i,a} | P_{i,a-1}] \). This yields predicted changes in EITC income as a function of lagged pretax income, taking into account the fact that income evolves over time in a predictable way and the EITC schedule changes in some years. By holding fixed the type of EITC schedule (one versus two+ children) \( s_{i,a-1} \) in generating the instrument, DL only exploit variation in predicted EITC income due to government changes in EITC schedules over time, and not due to changes in family structure.

Finally, they note that simply estimating equation (2) using \( \Delta \chi^IV_a \) as an instrument for \( \Delta I_{ia} \) will likely yield biased estimates for \( \delta_o \) since the changes in families’ simulated EITC payments are a function of lagged pretax income, which may also affect child achievement. Therefore, they augment equation (2) with a flexible function of lagged pretax income when instrumenting. Letting \( \Phi(P_{i,a-1}) \) reflect a flexible function of lagged pretax income, they estimate
8

\[ \Delta y_{ia} = \Delta x_{i}\alpha + \Delta w_{i}\beta + \Delta l_{i}\delta + \Phi \left( P_{i,a-1} \right) + \eta_{ia} \]

using \( \Delta x_{i}' \) as an instrument for \( \Delta l_{i} \). Empirically, DL use the same functional form for \( \Phi \left( P_{i,a-1} \right) \) as is used in calculating \( E \left[ P_{i,a} | P_{i,a-1} \right] \) (i.e. an indicator for positive lagged pretax income and a 5th-order polynomial in lagged pretax income). By using a fully flexible, but time-invariant control function, DL assume implicitly that the relationship between lagged income and shocks to child achievement does not vary over time. Moreover, with the inclusion of a fully flexible, but time-invariant control function, all identifying variation comes from differential changes in the EITC schedule over time, and not from differences in lagged income levels.

1.3 Replication Results

In Table (1), column (1) I present DL’s published estimates of the contemporaneous effects model, where child achievement is measured using the combined math and reading test score.\(^6\) In panel (A) I present results for the model estimated in levels (equation 1); in panel (B) I present results for the model estimated in differences (equation 2); in panel (C) I present results for the model estimated using the IV (equation 3). Income is measured in $1,000s and test scores are standardized, so the estimates represent the percent of one standard deviation that the test score changes given a $1,000 increase in family income. In column (2) I present results from my replication. This replication was performed as part of a larger project in which I estimate the impact of the EITC on child achievement for kids with single moms using a difference-in-differences strategy (Lundstrom 2015).\(^7\) Dahl and Lochner graciously provided me with copies of their programs and data which

\(^6\) In order to simplify the discussion I focus exclusively on the combined math and reading test score. My replication results for the individual tests also matched DL’s published results nearly perfectly. These results are available upon request.

\(^7\) The estimation strategy I employ is similar to one used by Hotz, Mullin, and Scholz (2006). It is a straightforward difference-in-differences model that exploits the fact that in the early 1990s, as part of the Omnibus Reconciliation Act of 1993 (OBRA93), the maximum EITC available to families with two or more children increased relative to the EITC available to families with only one child. If the cognitive abilities of children in EITC recipient families improve due to the increased generosity of the EITC, we should observe an improvement in the outcomes of
allowed me to replicate their results almost perfectly. Note that the IV estimate is more than fifty times larger than the differences estimate. DL propose two possible explanations for the IV estimates being so much larger than the levels or differences estimates. The first possibility is that measurement error in income produces severe attenuation bias in the levels and differences estimates. DL note that measurement error alone is unlikely to explain most of the gap, however. A second possible explanation is that the IV captures the effect of income for disadvantaged children—the marginal effect of income on child achievement might be particularly high for this group.

In the course of replicating DL's results I discovered two coding errors. First, contrary to DL's description, they do not hold the type of EITC schedule fixed (i.e. one versus two+ children) in generating the instrument. Failing to do so implies that instrument variation arises due to changes in family structure in addition to government changes in the EITC schedule. I correct this mistake and re-estimate the model. The new estimates are presented in Table (1), column (3). This correction reduces the IV estimate, though the change is very slight. The second coding error is less benign. In their construction of family income DL fail to add federal tax credits, and state tax credits are added twice. Since the EITC is a federal credit, this means that DL's measure of family income does not include the EITC. This omission is striking since DL assume that the correlation between the instrument and changes in EITC income provides the income variation in their identification strategy. Correcting this error does not affect the instrument, but it does strengthen the first stage since EITC income is highly correlated with the instrument. Results using the corrected measure of family income are presented in column (4) of Table (1). In total, the IV estimate falls by about one-third of a standard deviation after correcting the two errors.

---

children in two-child families relative to children in one-child families. The impact of the EITC on kids with single moms is of particular interest since the evidence suggests a large EITC-induced increase in labor force participation among single moms. The combined effect of this policy on these kids is unclear. On the one hand it increases family income. On the other hand it incentivizes the mother to enter the workforce and, quite possibly, spend less time with her kids.

8 See Appendix 1 for a detailed explanation of this error.
While correcting this bias is important, the corrected estimates are still statistically significant and are not significantly different from DL’s estimates. However, the discovery of these errors makes clear that DL’s interpretation of their first stage is not correct. DL assume that their first stage is primarily driven by the covariance between changes in the EITC schedule over time and changes in EITC income. They assume that the small deviation from one in the first stage is likely due to a labor supply response to the EITC expansions.\(^9\) However, since EITC income was not included in their construction of family income, this interpretation cannot be right. In the following section I determine what is driving the first stage.

1.4 What is Driving the First Stage?

In order to determine what is driving the first stage, I decompose the first stage into first stage estimates for various components of family income. I present this decomposition in Table (2). In column (1) I show the first stage for correctly measured family income. In column (2) I decompose the first stage for correctly measured income into two broad components: (A) the first stage for income as measured by DL and, (B) the first stage for the difference between correctly measured income and DL’s measure of income. In column (3) I decompose these two broad components into their subcomponents. What we discover in this final decomposition is that the first stage for income as measured by DL is entirely driven by maternal earnings. We also discover that the first stage for the portion of family income that is excluded from DL’s measure is—unsurprisingly—almost entirely driven by EITC income.

Since the EITC is tied to work it might not be surprising to find a strong first stage relationship with maternal earnings. Indeed, numerous studies have examined the impact of the EITC on labor supply and found consistent evidence that the EITC encourages work, primarily among single mothers (Eissa and Liebman 1996; Hoynes, Miller, and Simon 2015). Labor supply

\(^9\) See Dahl and Lochner (2012) page 1942, paragraph 1 for their interpretation of the first stage.
changes may affect parenting behavior and family dynamics, in addition to family income. Since changes in family dynamics and parenting behavior may also affect child achievement (as argued by DL in the motivation for their study), these variables will be lurking in the error term of the child achievement model and may cause the IV estimates to be biased.

In their study, DL recognize this threat but are unable to address it since labor supply changes are also endogenous and they do not have a second instrument. However, they consider the threat of bias to be minimal since they assume that the first stage is primarily driven by EITC income, with only a small earnings component. Correcting DL's coding error makes clear that, in reality, the first stage is primarily driven by maternal earnings and, thus, the threat of bias is potentially non-trivial. More generally, the fact that the instrument does not operate in the way that DL suppose begs the question of whether the instrument is valid. In the following section I explore this question.

2 Plausibility of Instrument Validity

In this section, I take a step back from any specific omitted variable bias concerns that are raised by the correlation between the instrument and maternal earnings and look more generally at the question of instrument validity. With the inclusion of the control function, \(\Phi(P_{t,a-1})\), DL assume that all identifying variation comes from government changes in the EITC schedule. Furthermore, they assume that the instrument is uncorrelated with the error term in the child achievement model, and that it only affects child achievement through its effect on family income (i.e. the exclusion restriction is valid). In this section, I present two sets of results that raise questions in regards to instrument validity. First, I show that I am able to obtain statistically significant first stage estimates, as well as large and statistically significant reduced form estimates—of either sign—when instrument variation arising from government changes in the EITC schedule is eliminated (or, at least, mostly eliminated). Second, I show that even when instrument variation arising from government changes

---

10 See Dahl and Lochner (2012), page 1948-1949, including footnote 20, for a discussion of this issue.
in the EITC schedule is preserved, the estimates are quite sensitive to which time periods this
variation comes from.

Before proceeding with the sensitivity analysis it is helpful to have a clear understanding of
how variation in the instrument is generated. In section 2.1 I fully explore the sources of instrument
variation. In section 2.2 I present results from my sensitivity analysis.

2.1 How is Variation in the Instrument Generated?

In this section I explain more fully how variation in the instrument is generated. To aid in the
discussion, I again present the instrument:

$$\Delta \chi_a^{W} (P_{i,a-1}) \equiv \chi_a^{s, a-1} (\hat{E}[P_{i,a} | P_{i,a-1}]) - \chi_a^{s, a-1} (P_{i,a-1})$$

where $i$ and $a$ denote child $i$ at age $a$, $\chi$ denotes EITC income, $P$ denotes pretax income, $s$ indicates
which EITC schedule the child’s family is on (i.e. one vs two+ child schedule), and $\hat{E}[P_{i,a} | P_{i,a-1}]$ is
an estimate of pretax income given lagged pretax income, where $\hat{E}[P_{i,a} | P_{i,a-1}]$ is estimated by
regressing pretax income on an indicator for positive lagged pretax income and a fifth order
polynomial in lagged pretax income.

First, conditional on family size and time, instrument variation arises due to differences in
lagged income levels between families. To illustrate this, suppose that the sample is restricted to the
1997-1999 period for families who have two+ children in 1997. All families with the same pretax
income ($P_{i,a-1}$) in 1997 will have the same predicted income ($\hat{E}[P_{i,a} | P_{i,a-1}]$) in 1999. Moreover, all
families with two+ children in 1997 face the same federal EITC schedule in 1997 and 1999 since the
type of EITC schedule (i.e. one vs. two+ children) is held fixed when creating the instrument. In this
case, instrument variation arises strictly from differences in lagged income levels between families.
This variation is illustrated in Figure (2a) where I plot the instrument against lagged income for the 1997-1999 time period for families with two+ children.

Second, conditional on time and lagged income, instrument variation arises due to differences in the EITC schedule by family size. To see this, suppose that we augment the 1997-1999 sample of two+ child families to also include one child families. In 1997 and 1999 the EITC schedule for one child families differs from the schedule for two+ child families. Because of this, simulated EITC changes for families with only one child differ from simulated EITC changes for families with two+ children. In Figure (2b) I illustrate these two dimensions of instrument variation (i.e. variation due to differences in lagged income levels, and variation due to differences in the EITC schedule by family size) by plotting the instrument against lagged income for both one and two+ child families for the 1997-1999 time period.

Third, conditional on lagged income and family size, instrument variation arises due to differences in the EITC schedule over time. To see this, suppose that we augment the 1997-1999 sample to include the 1987-1989 time period. Since predicted income is a function of lagged income only, all families with the same lagged pretax income will have the same predicted income (whether the lagged year is 1987 or 1997). However, the EITC schedules in 1987-1989 differ from the EITC schedules in 1997-1999, leading to different simulated EITC changes in 1987-1989 compared to the simulated EITC changes for families with identical lagged income in 1997-1999. In Figure (2c) I illustrate these three dimensions of instrument variation by plotting the instrument against lagged income for one- and two+ child families in 1997-1999 and all families in 1987-1989 (note: the EITC schedule does not vary by family size in 1987-1989).

Fourth, a very small amount of instrument variation is generated due to differences in state EITCs. I do not illustrate this source of variation as it is very minor. Indeed, as noted by DL, because few states had EITCs during the sample period considered in this study (five states by 1996
and ten states by 1999), the results are almost identical if state EITCs are excluded when constructing the instrument.

With the inclusion of the control function, $\Phi(P_{i,\alpha-1})$, DL assume that all instrument variation due to differences in lagged income levels is removed (recall that the control function consists of an indicator for positive lagged pretax income and a fifth order polynomial in lagged pretax income). More specifically, once the control function is added to the model, they assume that instrument variation comes exclusively from government changes in the EITC schedule. In the analysis that follows it becomes apparent that this is probably not true, and that this residual variation is likely causing some problems. I now proceed with the sensitivity analysis.

### 2.2 Sensitivity Analysis

In this sensitivity analysis I perform two tests. First, I test whether identification relies on government changes in the EITC schedule (as assumed by DL). Second, I test whether the estimates are sensitive to which time periods the EITC schedule variation comes from.

**Sensitivity test #1: Does identification rely on government changes in the EITC schedule?**

Looking at Figure (1), we see that for both types of families (i.e. one and two+ child families) most of the expansion of the EITC schedule occurs between 1993 and 1995. There are several small expansions that occur between 1987 and 1993, no expansions from 1995 to 1999 for one child families (in real terms), and only a small expansion between 1995 and 1997 for two+ child families. In this test I estimate a model for two+ child families during the 1987-1993 time period, and I estimate a separate model for two+ child families during the 1995-1999 time period. I also remove the state EITCs when constructing the instrument. By restricting the sample to two+ child families, I
completely eliminate instrument variation due to differences in the EITC schedule by family size. By excluding the state EITCs, I eliminate instrument variation due to differences across states. And by estimating separate models for the 1987-1993 and 1995-1999 time periods, I remove most instrument variation arising from differences in the EITC schedule over time. If identification relies on government changes in the EITC schedule, I should not obtain precise reduced form or first stage estimates given these sample restrictions.

I present reduced form (for combined math & reading test scores) and first stage estimates for kids from two+ child families in each time period in Table (3). Results are presented separately for a variety of demographic subgroups. The subgroups I use match those used by DL in Table 6 of their published paper. For the 1987-1993 time period, as shown in panel (A) of Table (3), the reduced form estimates are all negative, and for younger kids (age<12) the estimate is huge (-0.364) and significant at the 1% level. Moreover, I obtain large and statistically significant first stage estimates for kids whose moms have low education, for black or Hispanic kids, and for kids whose moms have low AFQT scores. For the 1995-1999 time period, as shown in panel (B), I obtain large positive, and statistically significant, reduced form estimates for black or Hispanic kids (significant at 10% level), and male children (significant at 5% level). I also obtain statistically significant (at the 10% level) first stage estimates for kids whose moms have low education and for black or Hispanic kids.

The evidence presented here suggests two things. First, that there is variation in the instrument even when EITC schedule changes are removed (apparently due to differences in lagged income levels). Second, for this portion of the variation at least, the exclusion restriction does not appear to hold. This second point can be seen in two ways. First, there is no apparent relationship

11 I could, of course, perform this test by restricting the sample to one child families instead. However, there are very few kids from one child families in the estimation sample. Using kids from two+ child families is an equally valid test and has the advantage of a far larger sample.
12 The results presented in Table (3) represent a partial list of the subgroups for which I obtained estimates. The full set of estimates is included in the appendix, Tables (A2) and (A3).
between the strength of the first stage and the strength of the reduced form estimates. In the
instances where the first stage estimates are strongest, the reduced form estimates are generally weak.
And where the reduced form estimates are strongest, the first stage estimates are quite weak. Second,
the relationship between the instrument and child achievement is almost certainly spurious. Consider
that when the sample is restricted to the 1987-1993 time period, these results suggest that a $1,000
increase in simulated EITC income reduces the test scores of young children by 36.4 percent of a
standard deviation. Conversely, when the sample is restricted to the 1995-1999 time period, the
reduced form effect for younger kids is small and imprecise while the effects for males and blacks or
Hispanics are large and positive (14.0 percent, and 22.1 percent of a standard deviation respectively).
Unless the instrument is unbelievably harmful to some kids in the earlier period, and unbelievably
helpful to some kids in the later period, the relationships must be spurious.

Sensitivity test #2: Are the estimates sensitive to which time period EITC schedule variation comes
from?

In the first test I produced evidence suggesting that the instrument is spuriously correlated
with child achievement when instrument variation arising from differences in EITC schedules is
excluded. In the second test, I present evidence suggesting that the relationship between the
instrument and child achievement is problematic even when variation arising from differences in
EITC schedules is included. I begin by constructing a sample that includes all observations from
1987 to 1993. As already noted, the EITC schedule is fairly stable (in real terms) over this time
period. In order to introduce variation arising from differences in EITC schedules I have a few
options. One option is to simply augment the sample with all observations from the later years.
Another option is to augment the 1987-1993 sample with observations from one two-year period at a
time. That is, I can estimate one model for a sample consisting of the 1987-1993 + 1993-1995 first-
differenced observations, another model for a sample consisting of the 1987-1993 + 1995-1997 first-
differenced observations, and a third model for a sample consisting of the 1987-1993 + 1997-1999 first-differenced observations. In each case I am augmenting the 1987-1993 sample with observations from a time period where the EITC schedules are quite different from the 1987-1993 schedules, thereby ensuring that instrument variation is influenced by EITC schedules changes. The advantage of this one-period-at-a-time approach is that it allows me to observe whether the estimates are sensitive to which time periods the EITC schedule variation comes from.

Estimates from these three samples are presented in Table (4). First stage estimates are statistically significant for all three samples, and are similar to each other in both magnitude and precision. However, the reduced form estimates are only positive and statistically significant for the 1987-1993 + 1993-1995 and 1987-1993 + 1997-1999 samples. For the 1987-1993 + 1995-1997 sample, the reduced form estimate is actually slightly negative. Since the first stage estimates are all very strong, we expect each estimate to be accompanied by a strong reduced form estimate. The fact that the reduced form estimate completely disappears for the 1987-1993 + 1995-1997 suggests that the relationship between the instrument and family income is not driving the relationship between the instrument and child achievement. The implication is that the reduced form relationships are likely driven by a spurious correlation between the instrument and shocks to child achievement (even if the source of this spurious relationship is unclear).

3 Conclusion

The causal effect of family income on child achievement is an important question. However, due to problems posed by the income endogeneity, it is a question that social scientists have struggled to answer convincingly for many years. The IV strategy that Dahl and Lochner (2012) employ to address this question is both creative and appealing. Unfortunately, the discovery of a coding error makes apparent that their IV estimates are overstated, and that their instrument does not covary with family income in the way that they suppose. Since the instrument does not operate in the
way that Dahl and Lochner describe in their paper, it is important to understand whether the instrument is valid. In this study I examine this question.

The evidence I present in this study raises questions in regards to instrument validity. In particular, when I estimate models using time periods where the largest EITC expansions are excluded, I obtain large and statistically significant reduced form estimates of either sign. These results are clearly spurious and indicate that there is at least some endogenous component of variation in the instrument. Furthermore, even when instrument variation arising from government changes in the EITC schedule is preserved, I find evidence suggesting that the reduced form (but not the first stage) estimates are sensitive to which time periods that variation comes from. These findings raise questions regarding the validity of the exclusion restriction.

The findings presented here do not say anything about whether or not income affects child achievement causally. There may very well be a causal effect of income on child achievement, and many studies point to that possibility. What this comment does do is raise some important questions in regards to the credibility of the specific identification strategy employed in this study. Moreover, if the instrument employed in this study is not valid, then the question of whether or not the EITC aids child achievement is still open. Considering that the EITC was originally designed as a program to assist working families with children (as opposed to childless families), the impact of this policy on child outcomes is of particular interest to policymakers. This important question is one which may be profitably explored by future researchers.
References


Figure 1. Federal EITC Schedules from 1987 to 1999 (Year 2000 Dollars)

Notes: This figure illustrates the changes in the EITC schedule from 1987 to 1999 for families with one child and for families with two+ children. The EITC became much more generous over this time period, particularly for larger families.
Figure 2. Illustration of instrument variation

(a) Variation due to differences in lagged income levels

(b) Variation due to EITC schedule differences by family size

(c) Variation due to EITC schedule differences over time

Notes: These figures illustrate the three sources of variation in the instrument. For a complete description, refer to section 2.1 of the paper.
Table 1. Estimates of the Contemporaneous Effects Model

<table>
<thead>
<tr>
<th>Panel A. Estimated in levels</th>
<th>Panel B. Estimated in differences</th>
<th>Panel C. Instrumental variables estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Math &amp; Reading:</td>
<td>Combined Math &amp; Reading:</td>
<td>Combined Math &amp; Reading:</td>
</tr>
<tr>
<td>0.0047** (0.001)</td>
<td>0.0011 (0.001)</td>
<td>0.0610** (0.023)</td>
</tr>
<tr>
<td>0.0047*** (0.001)</td>
<td>0.0011 (0.001)</td>
<td>0.0607*** (0.023)</td>
</tr>
<tr>
<td>0.0047*** (0.001)</td>
<td>0.0011 (0.001)</td>
<td>0.0575*** (0.022)</td>
</tr>
<tr>
<td>0.0048*** (0.001)</td>
<td>0.0012* (0.001)</td>
<td>0.0413*** (0.013)</td>
</tr>
<tr>
<td>Observations:</td>
<td>Observations:</td>
<td>Observations:</td>
</tr>
<tr>
<td>8,609</td>
<td>8,609</td>
<td>8,609</td>
</tr>
</tbody>
</table>

Notes: Child achievement is a normalized average of math and reading scores. Income is measured in $1,000 of year 2000 dollars. Panel A "levels" regressions (equation 1) control for all variables listed in Appendix Table A1. Panel B "difference" regressions (equation 2) use two-period differences and control for baseline variables in panel A of Table A1. Panel C "Instrumental variables" regressions (equation 3) use two-period differences and control for baseline variables in panel A of Table A1, an indicator for positive lagged pretax income, and a fifth-order polynomial in lagged pretax income. Samples include children taking a math or reading PIAT test in the 1988 survey year or later. Standard errors are reported in parentheses and are clustered at the family level. Significant at the ***1%, **5%, and *10% levels.
### Table 2. First Stage Decomposition

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First stage for correctly measured family income:</td>
<td>1.7548***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.362)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Decomposition of first stage for correctly measured family income:**

(A) First stage for income as measured by DL: 1.2584***

(B) First stage for the difference between correctly measured income and income as measured by DL: 0.4964***

**Notes:**
- The first stage estimates represent the dollar change (in year 2000 dollars) in the specified income component per $1 increase in a simulated EITC payment.
- In column 1 I present the first stage estimate for correctly measured family income. In column 2 I present a broad decomposition of correctly measured family income, consisting of family income as measured by DL, and the difference between correctly measured family income and income as measured by DL. In column 3 I present a more detailed decomposition of the first stage for income as measured by DL into maternal earnings, paternal earnings, unearned + nontaxable income, and "other income", which is a catch-all for any other income included in this measure. In column 3 I also present a decomposition of the first stage for the difference between correctly measured family income and income as measured by DL which consists of EITC income and other income. Significant at the ***1%, **5%, and *10% levels.
Table 3. Reduced form and first stage estimates for kids from families with two+ kids during time periods that exclude major EITC changes

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Column 5</th>
<th>Column 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother's education is high school or less</td>
<td>-0.0654 (0.100)</td>
<td>-0.1043 (0.112)</td>
<td>-0.1873 (0.135)</td>
<td>-0.0509 (0.109)</td>
<td>-0.3638*** (0.133)</td>
<td>-0.0192 (0.130)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A. 1987 - 1993</td>
<td>Reduced form: combined math &amp; reading</td>
<td>0.0822 (0.063)</td>
<td>0.1397* (0.082)</td>
<td>0.0417 (0.079)</td>
<td>0.0603 (0.078)</td>
<td>-0.0019 (0.096)</td>
</tr>
<tr>
<td></td>
<td>First stage: corrected family income</td>
<td>2.6115* (1.374)</td>
<td>3.3475* (2.007)</td>
<td>0.4491 (2.007)</td>
<td>1.7490 (1.812)</td>
<td>1.0390 (1.812)</td>
</tr>
<tr>
<td></td>
<td>Sample size:</td>
<td>3,096</td>
<td>2,270</td>
<td>1,372</td>
<td>2,123</td>
<td>2,205</td>
</tr>
<tr>
<td>Panel B. 1995 - 1999</td>
<td>Reduced form: combined math &amp; reading</td>
<td>0.0822 (0.063)</td>
<td>0.1397* (0.082)</td>
<td>0.0417 (0.079)</td>
<td>0.0603 (0.078)</td>
<td>-0.0019 (0.096)</td>
</tr>
<tr>
<td></td>
<td>First stage: corrected family income</td>
<td>2.6115* (1.374)</td>
<td>3.3475* (2.007)</td>
<td>0.4491 (2.007)</td>
<td>1.7490 (1.812)</td>
<td>1.0390 (1.812)</td>
</tr>
<tr>
<td></td>
<td>Sample size:</td>
<td>1,409</td>
<td>1,012</td>
<td>614</td>
<td>975</td>
<td>1,041</td>
</tr>
</tbody>
</table>

Notes: By estimating models for kids from two+ child families during these two time periods separately, I completely eliminate instrument variation due to differences in the EITC schedule by family size, and I mostly eliminate instrument variation due to differences in the EITC schedule over time. The reduced form estimates indicate the percent of one standard deviation that test scores change due to a $1,000 change in simulated EITC income. The first stage estimates indicate the dollar change (in 2000 dollars) in family income (using correctly measured income) per $1,000 change in simulated EITC income. Significant at the ***1%, **5%, and *10% levels.
Table 4. Reduced form and first stage estimates showing sensitivity to which time periods are included

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced form: combined math &amp; reading</td>
<td>0.0738***</td>
<td>-0.0108</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>First stage: corrected family income</td>
<td>2.0843***</td>
<td>1.6975***</td>
</tr>
<tr>
<td></td>
<td>(0.498)</td>
<td>(0.557)</td>
</tr>
<tr>
<td>Sample size:</td>
<td>6,172</td>
<td>5,839</td>
</tr>
</tbody>
</table>

Notes: the samples in each column are restricted to all first-differenced observations from the original estimation sample that occur during the indicated years. Adding observations from the 1993-1995, 1995-1997, or 1997-1999 time period to a sample composed of observations from 1987-1993 introduces instrument variation due to government changes in the EITC schedule. Significant at the ***1%, **5%, and *10% levels.
Appendix

1 Coding Error in Construction of Family Income

In this appendix I document the coding error that resulted in the construction of a measure of family income that did not include EITC income. The TAXSIM outputs that are included in DL’s construction of family income are: v10, v28, siitax, and v40 (see line 198 of the stata file taxsim-eitc.do for DL’s use of taxsim output in the construction of family income). These TAXSIM outputs are described below:

- v10: federal adjusted gross income
- v28: federal income tax before credits
- siitax: state income tax liability
- v40: state total credits

The variable v28 is federal income tax before credits. Since the EITC is a federal credit, it is not included with v28. The variable that DL should have used is fiitax, which is total federal income tax liability and does account for federal credits. The variable siitax is total state income tax liability and already nets out state credits. It is redundant to add state credits using v40. Detailed descriptions of the taxsim outputs can be obtained from the Taxsim website: http://nber.org/~taxsim/taxsim-calc9/index.html. For more information, the creator of Taxsim, Daniel Feenberg, can be contacted directly at feenberg@nber.org.

The mistake in the exclusion of federal credits was likely due to an incorrect description of the variable v28 in a previous version of TAXSIM. In a previous version of TAXSIM, the variable v28 is incorrectly defined as "Federal Income Tax After Credits" (see http://users.nber.org/~taxsim/taxsim-calc8/). In the current version of TAXSIM, the variable
v28 is correctly referenced as "Income Tax before Credits". The content of variable v28 did not change from the earlier version of TAXSIM to the current version, but the documentation did.
### Table A1. Control variables used in estimation of contemporaneous effects model

**Panel A. Baseline variables**
- Male
- Age
- No siblings
- One sibling
- Two+ siblings
- Black
- Hispanic

**Panel B. Additional variables**
- Mother's age
- Mother a high school dropout
- Mother a high school graduate
- Mother attended some college
- Mother graduated college
- Mother's AFQT score (normalized and age adjusted)
- Mother lived with both natural parents at age 14
- Mother's father present in household
- Mother's mother present in household
- Number of adults in household
- Highest grade completed by mother's father
- Highest grade completed by mother's mother
- Mother married last year
- Age of mother's spouse
- Mother's spouse a high school dropout
- Mother's spouse a high school graduate
- Mother's spouse attended some college
- Mother's spouse a college graduate
- Year

Missing observation indicators:
- Mother's AFQT score
- Mother lived with both natural parents at age 14
- Mother's father present in household
- Mother's mother present in household
- Number of adults in household
- Highest grade completed by mother's father
- Highest grade completed by mother's mother
- Age of mother's spouse
- Mother's spouse's education

**Notes:** Mother's education variables represent completed education when the mother is age 23. See Dahl and Lochner (2012) Table (A1) for summary statistics using these variables.
Table A2. Reduced form and first stage estimates for kids from families with two+ kids during time periods that exclude major EITC changes during, 1987 - 1993

<table>
<thead>
<tr>
<th></th>
<th>Mother's education</th>
<th>Race</th>
<th>Mother's marital status</th>
<th>Mother's AFQT</th>
<th>Child's age</th>
<th>Child's gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High school or less</td>
<td>Black or Hispanic</td>
<td>Not married</td>
<td>Low AFQT</td>
<td>Age&lt;12</td>
<td>Male</td>
</tr>
<tr>
<td>Reduced form: combined math &amp; reading</td>
<td>-0.0654 (0.100)</td>
<td>-0.1043 (0.112)</td>
<td>-0.1873 (0.135)</td>
<td>-0.0509 (0.109)</td>
<td>-0.3638*** (0.133)</td>
<td>-0.0192 (0.130)</td>
</tr>
<tr>
<td>First stage: corrected family income</td>
<td>3.5757** (1.601)</td>
<td>2.9426* (1.592)</td>
<td>0.0617 (1.705)</td>
<td>3.0299* (1.742)</td>
<td>1.5575 (1.905)</td>
<td>1.9810 (2.002)</td>
</tr>
<tr>
<td>Sample size:</td>
<td>3,096</td>
<td>2,270</td>
<td>1,372</td>
<td>2,123</td>
<td>2,205</td>
<td>1,924</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Some college or more</th>
<th>White (not hisp.)</th>
<th>Married</th>
<th>High AFQT</th>
<th>Age≥12</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced form: combined math &amp; reading</td>
<td>-0.2200 (0.246)</td>
<td>-0.0455 (0.162)</td>
<td>-0.0999 (0.128)</td>
<td>-0.1780 (0.179)</td>
<td>0.1960* (0.117)</td>
<td>-0.1443 (0.117)</td>
</tr>
<tr>
<td>First stage: corrected family income</td>
<td>-7.4347* (4.260)</td>
<td>-0.0455 (3.410)</td>
<td>3.0836 (2.991)</td>
<td>0.0636 (3.101)</td>
<td>3.0928* (1.857)</td>
<td>2.5872 (1.820)</td>
</tr>
<tr>
<td>Sample size:</td>
<td>793</td>
<td>1,619</td>
<td>2,517</td>
<td>1,632</td>
<td>1,684</td>
<td>1,965</td>
</tr>
</tbody>
</table>

Notes: reduced form estimates indicate the percent of one standard deviation that test scores change due to a $1,000 change in simulated EITC income. The first stage estimates indicate the dollar change (in 2000 dollars) in family income (using correctly measured income) per $1,000 change in simulated EITC income. Significant at the ***1%, **5%, and *10% levels.
Table A3. Reduced form and first stage estimates for kids from families with two+ kids during time periods that exclude major EITC changes during, 1995 - 1999

<table>
<thead>
<tr>
<th></th>
<th>Mother's education</th>
<th>Race</th>
<th>Mother's marital status</th>
<th>Mother's AFQT</th>
<th>Child's age</th>
<th>Child's gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High school or less</td>
<td>Black or Hispanic</td>
<td>Not married</td>
<td>Low AFQT</td>
<td>Age&lt;12</td>
<td>Male</td>
</tr>
<tr>
<td>Reduced form: combined math &amp; reading</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>0.0822</td>
<td>0.1397*</td>
<td>0.0417</td>
<td>0.0603</td>
<td>-0.0019</td>
<td>0.2214**</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.082)</td>
<td>(0.079)</td>
<td>(0.078)</td>
<td>(0.096)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>First stage: corrected family income</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>2.6115*</td>
<td>3.3475*</td>
<td>0.4491</td>
<td>1.7490</td>
<td>1.0390</td>
<td>2.0210</td>
</tr>
<tr>
<td></td>
<td>(1.374)</td>
<td>(2.007)</td>
<td>(1.661)</td>
<td>(1.802)</td>
<td>(1.876)</td>
<td>(1.668)</td>
</tr>
<tr>
<td>Sample size:</td>
<td>1,409</td>
<td>1,012</td>
<td>614</td>
<td>975</td>
<td>1,041</td>
<td>1,093</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Some college or more</th>
<th>White (not hisp.)</th>
<th>Married</th>
<th>High AFQT</th>
<th>Age≥12</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced form: combined math &amp; reading</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>-0.0269</td>
<td>-0.0496</td>
<td>0.0633</td>
<td>0.0149</td>
<td>0.0767</td>
<td>-0.0974</td>
</tr>
<tr>
<td></td>
<td>(0.126)</td>
<td>(0.083)</td>
<td>(0.089)</td>
<td>(0.089)</td>
<td>(0.073)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>First stage: corrected family income</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>-1.7122</td>
<td>-0.2121</td>
<td>2.8255</td>
<td>0.7609</td>
<td>2.5123*</td>
<td>1.6428</td>
</tr>
<tr>
<td></td>
<td>(3.168)</td>
<td>(1.751)</td>
<td>(2.488)</td>
<td>(2.168)</td>
<td>(1.490)</td>
<td>(1.770)</td>
</tr>
<tr>
<td>Sample size:</td>
<td>790</td>
<td>1,187</td>
<td>1,585</td>
<td>1,168</td>
<td>1,158</td>
<td>1,106</td>
</tr>
</tbody>
</table>

Notes: see notes for Table A2. Significant at the ***1%, **5%, and *10% levels.