Permanent Income and the Black-White Test Score Gap

Jesse Rothstein*
University of California, Berkeley and NBER

Nathan Wozny
Mathematica Policy Research

December 2010

Abstract

Analysts often examine the black-white test score gap conditional on family income. Typically only a current income measure is available. We argue that the gap conditional on permanent income is of greater interest, and we describe a method for identifying this gap using an auxiliary data set to estimate the relationship between current and permanent income. Current income explains only about half as much of the black-white test score gap as does permanent income, and the remaining gap in math achievement among families with the same permanent income is only 0.2 to 0.3 standard deviations in the CNLSY and ECLS samples. When we add permanent income to the controls used by Fryer and Levitt (2006), the unexplained gap in 3rd grade shrinks below 0.15 SDs, less than half of what is found with their controls.

* Corresponding author. E-mail rothstein@berkeley.edu. We thank Lars Lefgren, Bob Margo, Sean Reardon, Richard Rothstein, and seminar participants at UBC and Mathematica and in the NBER Labor Studies and Children's groups for helpful conversations and suggestions. We also thank Fanyin Zheng for excellent research assistance.

I. Introduction

The black-white test score gap has been extensively documented. On many different types of tests, administered to children of various ages over several decades, average scores for blacks are substantially lower than those for whites. Although the precise magnitude of the gap varies across samples, tests, and ages, gaps approaching one full standard deviation are not uncommon. Moreover, while the gap shrunk rapidly during the 1970s and 1980s, progress has largely stopped among cohorts born since the early 1970s (Neal 2006; Chay, Guryan, and Mazumder 2009). This has important economic implications: Neal and Johnson (1996) show that nearly all of the black-white wage gap can be attributed to differences in "premarket" factors such as test scores, so slow progress in closing the test score gap suggests that economic disparities will persist for many decades to come.

If there were no black-white gap among families with the same income, we might hope that eventual convergence of black and white family incomes (Neal 2006; Krueger, Rothstein, and Turner 2006) would lead to the disappearance of the test score gap. But while ethnographic evidence (Kozol 1992, Lareau 2003) suggests that material circumstances can account for much of the black-white gap, this view has not been supported by statistical analyses of representative samples. Jencks and Phillips (1998) summarize the state of knowledge: "Income inequality between blacks and whites appears to play some role in the test score gap, but it is quite small" (p. 9); and, "the gap shrinks only a little when black and white families have the same amount of schooling, the same income, and the same wealth" (p. 2). Fryer and Levitt (2004, 2006) are the most successful at explaining the gap via differences in observable characteristics, but even they find that the gap that remains after controlling for a vector of demographic and behavioral variables is nearly 0.4 standard deviations by the end of the 3rd grade (down from a raw gap of 0.88 SDs in their sample). Many have concluded from the robustness of the black-white gap to controls for income and other family characteristics that it is largely attributable to differences in genes (Herrnstein and Murray 1996), culture (Moynihan 1965), or parenting styles (Brooks-Gunn et al. 1996) between blacks and whites. These factors are unlikely to be amenable to simple policy interventions.

In this paper we argue that important shortcomings in the way that income is measured have led the existing literature to dramatically understate the role of family income differences in accounts of the black-white test score gap and, therefore, to dramatically overstate the gap among children with the same family incomes. Studies of the conditional black-white gap typically control for the family's measured income in the year that the child was tested. As has long been recognized (Modigliani and Brumberg 1954; Friedman 1957), annual income is a poor proxy for a family's consumption and investment possibilities, and in any case it may be measured with substantial error in population surveys. To see the implications of this, consider a regression that controls for annual income, where test scores are instead related to permanent income. Because measured annual income is a noisy proxy for the true variable, its coefficient is attenuated relative to what would be observed if permanent income were controlled.[1] As mean income is lower for blacks than for whites, this attenuation leads to overstatement of the black-white test score gap conditional on income.

In literatures where income is the explanatory variable of interest researchers often attempt to form better measures of permanent income. One common strategy is to average observed incomes over several years, perhaps five (e.g., Solon 1992, Mayer 1997). More recent studies have suggested that even short-run averages are quite noisy (Mazumder 2001, Haider and Solon 2006). Using a longer-term income average, Mazumder (2003) concludes that measurement error in 5-year-income averages leads to attenuation of the intergenerational elasticity of income by nearly 30%. But the insight that it is important to measure permanent income accurately has been slow to penetrate literatures where income is used only as a control variable, despite the well-known result that mismeasurement of one explanatory variable will bias the coefficients for all right-hand-side variables in OLS regressions. Researchers typically use annual income (Campbell et al. 2008) or short-run averages (Phillips et al. 1998; Blau and Grossberg 1992), or simply substitute other variables – like maternal education or socioeconomic

[1] We assume for the moment that annual income equals permanent income plus a i.i.d. error. Although this specific configuration is unlikely (Haider and Solon 2006), more general income processes produce similar results for the conditional black-white gap.

status indices – that are thought to be effective proxies for permanent income (Fryer and Levitt 2004, 2006).[2]

We begin by showing that the distinction between current and permanent income is an important one for understanding black-white differences. Using data from the Child Supplement to the National Longitudinal Study of Youth (CNLSY), we show that the black-white gap in permanent family income is fully half as large among families with the same observed annual income as it is unconditionally, suggesting that current income is at best a limited proxy for the permanent income that one would like to control.

One reason that researchers studying test score gaps do not control for permanent income is that the data requirements are onerous: Child achievement measures are rarely available in the same data sets as the longitudinal family income histories needed to measure permanent income. We describe how instrumental variables (IV) techniques can be used to identify the black-white test score gap conditional on permanent income even when only annual income is observed in the achievement sample, relying on an auxiliary data set with long income histories. When there are no covariates or all covariates are observed in both the primary and auxiliary samples, the estimator is conventional two-sample two-stage least squares (TS2SLS), using current income as an instrument for permanent income. We describe a modification of TS2SLS that can be used when some covariates are observed only in the test score sample.

In data from the Early Childhood Longitudinal Study (ECLS), we find that the black-white math score gap at the end of 5th grade, controlling for permanent income and for a very short list of family structure variables (primarily the mother's age at the child's birth), is only 0.18 standard deviations. By contrast, the gap without income controls is 0.62 standard deviations and the traditionally-estimated gap conditional on annual income is 0.38. We also reconsider Fryer and Levitt's (2006) analysis of 3rd grade math scores. When we add permanent income to the short vector of controls used by Fryer and Levitt – birth weight, number of children's books in the home, mother's age at first birth,

[2] An exception is Blau (1999), who finds that average incomes over 12 years have only small effects on child development. The literature on racial gaps in wealth accumulation has also explored the implications of transitory income variation (see, e.g., Altonji and Doraszelski 2005).

and a WIC recipient indicator – we find that the remaining black-white gap in 3rd grade falls to 0.15 standard deviations, less than half of the 0.34 that we obtain with the Fryer and Levitt controls.

Our approach is applicable beyond the analysis of test score gaps. An identical strategy would be useful whenever one is willing to impose our key identifying assumption, that transitory variations in annual income are unrelated with outcomes conditional on permanent income. This assumption frequently follows immediately from the permanent income hypothesis (PIH) and may be a tolerably close approximation in many contexts.[3] Thus, one might use our strategy to estimate the gap in homeownership or health outcomes between blacks and whites with the same permanent income, for example.

In Section II, we provide an overview of the black-white gap and of the role of income controls in studies of this gap. In Section III, we develop a simple econometric model of the role of income and race, and we describe our two approaches to identifying the black-white gap conditional on permanent income. In Section IV, we discuss the two data sets used in this paper, the CNLSY and the ECLS. In Section V, we present simple analyses of the dynamics of family income in the CNLSY data. Section VI presents results on black-white test score gaps. Section VII concludes.

II. Overview of the Problem

The gap in mean scores between black and white children is large and persistent, nearly always above 0.5 standard deviations and more commonly in the 0.75 to 1 range.[4] It is robust across many different samples and different tests, and in particular is apparent on both aptitude and achievement tests. It tends to be somewhat larger for older than for younger children (Phillips et al. 1998, Fryer and Levitt 2004, 2006). In longitudinally consistent data, it is smaller for children born since 1970 than for those born before 1960, though there was little additional progress for children born between the early 1970s and late 1980s (Chay, Guryan, and Mazumder 2009).

[3] We discuss the possibility that current income may have direct effects on student achievement below.

[4] See, e.g., the reviews by Neal (2006), Jencks and Phillips (1998), and Magnuson and Waldfogel (2008).

The test score gap has proven to be surprisingly resilient to controls for income and other family characteristics, for school segregation and integration, and for school spending (Jencks and Phillips 1998). Hedges and Novell (1998), for example, find that differences in parental education and family income explain only about 30% of the black-white gap. Phillips et al. (1998; see also Grissmer and Eiseman 2008) find that broader measures of family environment – including mother's perceived self-efficacy and parenting practices – can explain somewhat more.

In an important recent paper, Fryer and Levitt (2004; hereafter *FL*) control for a relatively short list of covariates, mixing family characteristics and choice variables: A socioeconomic status index (based on parental education and occupational status and on the family's income), gender, the number of children's books in the home, the child's age and birth weight; indicators for teen mothers and for older mothers, and an indicator for receipt of WIC benefits. They find that these variables can fully explain the black-white gap in reading and (to a lesser extent) math scores among entering kindergarteners in the ECLS sample. However, this result does not persist for long: By the time the same children finished third grade, the raw black-white gap had grown and the share explained by the FL covariates had fallen to two thirds (Fryer and Levitt 2006).

To understand why researchers often focus on the black-white gap in test scores conditional on income and other observable characteristics, it is useful to consider scores as an outcome of household investment decisions. We assume that a unitary household has utility that depends on consumption in each of T periods, $c_1, \ldots, c_T$, on the child's academic achievement, s, and on a preference parameter γ: $U=U(s, c_1, \ldots, c_T; \gamma)$.[5] The preference parameter might capture variation in parents' altruism, in parenting styles, or in direct tastes for education. Achievement is a function of innate aptitude, a, and educational investments, e: $s = f(a, e)$.[6]

[5] This can be seen as an indirect utility, where direct utility depends on consumption in the current generation and in the next generation and where the child's human capital affects her future earnings.

[6] Nothing that follows would be changed if the preference parameter γ entered directly into the educational production function f(), as might be the case if γ reflects variation in (for example) willingness to read to one's children.

The household receives income stream $Y_1, \ldots, Y_T$ and must allocate it between consumption and educational investments. For simplicity and without loss of generality, we assume that educational investments take place in period 0. Assuming for the moment that the household can borrow and save freely at rate r, the intertemporal budget constraint, expressed in period-0 dollars, is $e + \Sigma_t c_t(1+r)^{-t} = \Sigma_t Y_t(1+r)^{-t}$. With this assumption, educational investments and consumption decisions depend only on the discounted value of lifetime income, $\Sigma_t Y_t(1+r)^{-t}$, and not on the family's income in any particular year conditional on this.[7] In other words, the family's consumption and investment decisions are identical to those that would be seen if the family received constant annual income $P = (\Sigma_t Y_t(1+r)^{-t})/ (\Sigma_t (1+r)^{-t})$. We refer to P as the family's permanent income. The budget constraint can be written as $e = P\Sigma_t(1+r)^{-t} - \Sigma_t c_t(1+r)^{-t}$ or, in terms of achievement, as

$$s = f(a, P\Sigma_t(1+r)^{-t} - \Sigma_t c_t(1+r)^{-t}). \tag{1}$$

Using the implicit function theorem we can write the chosen achievement level s^* as a function of ability, preferences, and permanent income, $s^* = g(a, \gamma, P)$. Rescaling g() in terms of p=ln(P) and linearizing, we obtain

$$s^* \approx \beta_0 + a\,\beta_a + \gamma\,\beta_\gamma + p\,\beta_p. \tag{2}$$

Here, the effects of tastes and income operate solely through expenditure choices, while the ability coefficient β_a reflects both the direct effects of ability on achievement ($\partial f/\partial a$) and indirect effects operating through the choice of expenditures.

Equation (2) is not estimable, as a and γ are not readily observed. However, it is useful in understanding the sources of black-white gaps in s^*. Let b be an indicator for a black student, and let $\delta(X) = E[X \mid b=1] - E[X \mid b=0]$ be the black-white gap in some variable X. By (2), we can write the unconditional black-white test score gap as

$$\delta(s^*) = \delta(a)\,\beta_a + \delta(\gamma)\,\beta_\gamma + \delta(p)\,\beta_p. \tag{3}$$

[7] If families face credit constraints or are uncertain about future incomes, current and past income may have larger effects on investment decisions and test scores than does future income, producing direct effects of $\{Y_1, \ldots, Y_T\}$ on e (see, e.g., Carneiro and Heckman, 2002, who use this idea to construct a test for credit constraints). We maintain the permanent income assumption for expositional purposes in this Section, but we explore likely violations in our empirical analysis.

There are thus three sources of gaps in mean test scores: Differences in ability distributions ($\delta(a)<0$), differences in preferences ($\delta(\gamma)<0$), and differences in incomes ($\delta(p)<0$). The last is importantly different from the first two: One can easily imagine policy responses (e.g., changes in tax schedules) that would shrink the gap in disposable incomes between black and white families, but it would be difficult or impossible to design policies to reduce differences in ability or preferences between groups.

Next, consider the conditional expectation of test scores given race and income:

$$E\left[s^{*} \mid b,p\right] = \beta_0 + E\left[a \mid b,p\right]\beta_a + E\left[\gamma \mid b,p\right]\beta_\gamma + p\beta_p. \tag{4}$$

The black-white gap conditional on permanent income is simply the gap in this conditional expectation. Letting $\delta_p(X) = E[X \mid b=1, p] - E[X \mid b=0, p]$, we can write:

$$\begin{aligned} \delta_p\left(s^{*}\right) &\equiv E\left[s^{*} \mid b=1,p\right] - E\left[s^{*} \mid b=0,p\right] \\ &= \left(E\left[a \mid b=1,p\right] - E\left[a \mid b=0,p\right]\right)\beta_a + \left(E\left[\gamma \mid b=1,p\right] - E\left[\gamma \mid b=0,p\right]\right)\beta_\gamma \\ &= \delta_p(a)\beta_a + \delta_p(\gamma)\beta_\gamma \end{aligned} \tag{5}$$

Conditioning on p eliminates the β_p term from (3). Moreover, assuming that p is positively correlated with a and γ within race, $\delta_p(a) > \delta(a)$ and $\delta_p(\gamma) > \delta(\gamma)$. The conditional gap is thus smaller in magnitude than the unconditional gap, and is negative only if there are differences in ability or tastes between black and white families with the same incomes. Evidence that the black-white test score gap is largely robust to controls for permanent income would therefore suggest that the gap is primarily attributable to black-white differences in ability or preferences rather than to the direct effects of income. By contrast, evidence that the black-white gap was largely eliminated by the inclusion of income controls would be less conclusive, as a small $\delta_p\left(s^{*}\right)$ could be consistent with a raw gap that derives primarily from the causal effect of income on investments (i.e., from β_p) or with a gap due primarily to ability and attitudes that are well proxied by income (i.e., $\delta(a)\ \beta_a + \delta(\gamma)\ \beta_\gamma$ is large but $\left|\delta_p(a)\beta_a + \delta_p(\gamma)\beta_\gamma\right| << \left|\delta(a)\beta_a + \delta(\gamma)\beta_\gamma\right|$). We do not attempt to distinguish these two explanations.[8]

[8] Mayer (1997) and Dahl and Lochner (2010) review studies that attempt to identify the causal effect of income.

III. Estimation

A. Controlling for current income

We would like to estimate the following regression:

$$s_i = \theta_0 + b_i\,\theta_b + p_i\,\theta_p + \varepsilon_i. \qquad (6)$$

By the argument above, the θ_b coefficient would estimate $\delta_p(a)\beta_a + \delta_p(\gamma)\beta_\gamma$. Unfortunately, p_i is not commonly observed in educational data sets. Thus, existing studies typically condition on annual income in a particular year, typically the year of the test, rather than on p. The regression typically estimated is:

$$s_i = \theta_0' + b_i\,\theta_b' + y_{it}\,\theta_y' + \varepsilon_i', \qquad (7)$$

where y_{it} represents the log annual income of family i in year t. In general, $\theta_b' \neq \theta_b$. With the reasonable assumption that e_{it} is uncorrelated with ε_i we can apply the conventional errors-in-variables (EIV) formula to recover the relationship between (θ_b θ_p) and (θ_b' θ_y').[9] Following Haider and Solon (2006), we assume $y_{it} = \alpha_t p_i + e_{it}$, with $E[b_i e_{it}] = E[p_i e_{it}] = E[e_{it}] = 0$. We also adopt the simplifying but inessential assumption that var(p | b) and var(e | b) are constant across b=0 and b=1. We define

$$R_b \equiv \alpha_t^2 * \mathrm{var}(p_i \mid b_i) / \mathrm{var}(y_{it} \mid b_i). \qquad (8)$$

As $\mathrm{var}(y_{it} \mid b_i) = \alpha_t^2\,\mathrm{var}(p_i \mid b_i) + \mathrm{var}(e_{it})$, R_b can be interpreted as the within-race reliability of annual income as a proxy for permanent income. Thus, $R_b \leq 1$ with R_b=1 only if $\mathrm{var}(e_{it}) = 0$. R_b will be low if the transitory component of income is large or if current income is measured with a great deal of error.

Replacing the p in regression (6) with y in (7) has two effects. First, it rescales income by the multiplicative factor α_t. Thus, if $\mathrm{var}(e_{it}) = 0$, $\theta_y' = \theta_p / \alpha_t$. Second, if e_{it} is nondegenerate, it produces attenuation in the income effect and biases the black coefficient downward:

$$\theta_y' = (R_b / \alpha_t) * \theta_p \leq \theta_p / \alpha_t \text{ and} \qquad (9)$$

$$\theta_b' = \theta_b + (1-R_b)\,\delta(p)\theta_p. \qquad (10)$$

---

[9] The assumption would be incorrect if PIH fails or if the time path of family income carries information about the family's ability or tastes. We present specifications below that control for p_i and y_{it} simultaneously to investigate the importance of the assumption.

If black families have lower permanent incomes than white families (i.e., $\delta(p)<0$), $\theta_b' \leq \theta_b$. Indeed, equation (10) can be re-arranged to express θ_b' as a weighted average of the black-white gap conditional on permanent income, θ_b, and the unconditional gap, $\delta(s) = \theta_b + \delta(p)\theta_p$, with weights R_b and $(1-R_b)$, respectively:

$$\theta_b' = \theta_b R_b + (1-R_b)\, \delta(s). \tag{11}$$

Intuitively, black families on average have lower permanent incomes than white families with the same annual incomes, with the difference increasing in $(1-R_b)$. As a consequence, controlling only for annual income will produce a conditional gap shaded toward the unconditional gap, with more shading the lower is R_b.

If a permanent income measure is available in the test score sample, eliminating the bias is simple: One simply estimates specification (6) instead of (7). We nevertheless outline a more complex instrumental variables (IV) strategy that also recovers θ_b. This has two advantages over OLS estimation of (6). First, it can be extended to allow identification of θ_b when p is not observed in the same sample as s. Second, when p is measured with error, as will be the case in nearly any data set, the resulting bias may in empirically relevant situations be smaller in the IV estimator than in OLS. We discuss each point below.

B. An IV-based correction

It is well known that instrumenting a regressor that is measured with error with another measure with independent error yields a consistent estimate of the coefficient that would be obtained with an error-free measure. The same logic applies in our case, though we apply it in an unusual way: We consider instrumenting for the error-free measure, p, with the unreliable measure y. By our assumption above, y is uncorrelated with ε so is a valid instrument; it is of course correlated with p.

The intuition for this IV-based correction may be made clearer by considering the identification of the income coefficient θ_p. The first-stage regression is:

$$p = \lambda_0 + y\, \lambda_y + b\, \lambda_b + \eta. \tag{12}$$

Note that this is the reverse of the intuitive regression of current on permanent income. In that regression, we would expect a coefficient of α_t on permanent income and zero on race. In this regression, it is straightforward to show that plim $\lambda_y = R_b / \alpha_t$ and plim $\lambda_b =$

$(1-R_b)\delta(p)$. Equation (9) shows that the "reduced form" y coefficient is $\theta_y' = (R_b / \alpha_t) * \theta_p$, so the ratio θ_y' / λ_y – the two-stage least squares estimator for θ_p – identifies θ_p.

The primary appeal of the IV strategy is that many education data sets lack an adequate measure of p. The reduced form equation (7) can be estimated on any sample that contains measures of s, y, and b. If an auxiliary sample containing {p, y, b} is available for estimation of the first stage (12), the two can be combined to form a two-sample instrumental variables estimator (Angrist and Krueger, 1992).

Converting to matrix representations, let W represent non-income covariates, including b, a constant, and any other controls that are to be added to (6). Let Z = [y W] and X = [p W]. Suppose that Z and s are observed in sample 1 (the "test score sample") and that Z and X are observed in sample 2 (the "auxiliary sample"), and let subscripts denote the sample in which a variable is measured. The estimand is the coefficient vector from equation (6), β.

In a contrast from traditional IV applications, the OLS estimator $\hat{\beta}_{OLS} = (X'X)^{-1}X's$ would be consistent for β. However, it is infeasible if X and s are not observed in the same sample. A two-sample two-stage least squares (TS2SLS) estimator,

$$\hat{\beta}_{TS2SLS} = \left((Z_2'Z_2)^{-1}(Z_2'X_2)\right)^{-1}(Z_1'Z_1)^{-1}Z_1's_1. \tag{13}$$

exploits the fact that Z is a valid instrument for X. If the four moment matrices in (13), when scaled by the appropriate sample sizes n_1 and n_2, consistently estimate the corresponding population moments, the TS2SLS estimator identifies β.[10]

In some of the specifications below, we include covariates V that are available only in sample 1. This introduces complications. Redefining Z = [y W V] and X = [p W V], we require a consistent estimate of

---

[10] A numerically equivalent estimator starts by estimating the first-stage coefficients in sample 2, $\hat{\lambda}_2 = (Z_2'Z_2)^{-1}Z_2'X_2$, then forming predicted values $\hat{X}_1 = Z_1\hat{\lambda}_2$ and regressing s on these, $\hat{\beta}_{TS2SLS} = (\hat{X}_1'\hat{X}_1)^{-1}\hat{X}_1's_1$. The two-sample instrumental variables estimator, $\hat{\beta}_{TSIV} = \frac{n_2}{n_1}(Z_2'X_2)^{-1}Z_1's_1$ is also consistent but distinct. Solon and Inoue (2010) show that TS2SLS is more efficient than TSIV.

$$E[Z'X] = \begin{pmatrix} E[y'p] & E[y'W] & E[y'V] \\ E[W'p] & E[W'W] & E[W'V] \\ E[V'p] & E[V'W] & E[V'V] \end{pmatrix}. \tag{14}$$

This cannot be obtained solely from sample 1, which lacks p, or from sample 2, which lacks V. However, with one additional assumption the two samples can be combined to estimate each element of (14). Specifically, let $\pi = [\pi_p \ \pi_W']'$ be the coefficients of a linear projection of y on p and W. Note that π can be estimated from sample 2. We assume that $E[V'(y - p\pi_p - W\pi_W)]=0$; that is, that V is uncorrelated with the transitory component of income conditional on the control variables W. With this assumption, $E[V'y] = E[V'p]\pi_p + E[V'W]\pi_W$, so $E[V'p] = (E[V'y] - E[V'W]\pi_W)\pi_p^{-1}$. This permits a hybrid estimator for E[Z'X]:

$$n_2^{-1}\tilde{Z}'\tilde{X} = n_2^{-1}\begin{pmatrix} y_2'p_2 & y_2'W_2 & \frac{n_2}{n_1}y_1'V_1 \\ W_2'p_2 & W_2'W_2 & \frac{n_2}{n_1}W_1'V_1 \\ \frac{n_2}{n_1}\left(V_1'y_1 - V_1'W_1\hat{\pi}_{W,2}\right)\hat{\pi}_{p,2}^{-1} & \frac{n_2}{n_1}V_1'W_1 & \frac{n_2}{n_1}V_1'V_1 \end{pmatrix} \xrightarrow{p} E[Z'X]. \tag{15}$$

We form a corresponding Z'Z matrix:

$$n_2^{-1}\tilde{Z}'\tilde{Z} = n_2^{-1}\begin{pmatrix} y_2'y_2 & y_2'W_2 & \frac{n_2}{n_1}y_1'V_1 \\ W_2'y_2 & W_2'W_2 & \frac{n_2}{n_1}W_1'V_1 \\ \frac{n_2}{n_1}V_1'y_1 & \frac{n_2}{n_1}V_1'W_1 & \frac{n_2}{n_1}V_1'V_1 \end{pmatrix} \xrightarrow{p} E[Z'Z]. \tag{16}$$

We use these to form a hybrid two-sample 2SLS estimator:

$$\hat{\beta}_{HTS2SLS} = \left(\left(\tilde{Z}'\tilde{Z}\right)^{-1}\left(\tilde{Z}'\tilde{X}\right)\right)^{-1}\left(Z_1'Z_1\right)^{-1}Z_1's_1. \tag{17}$$

This is consistent for β.[11]

---

[11] We could replace $n_2^{-1}\tilde{Z}'\tilde{Z}$ with any consistent estimator for E[Z'Z], including $n_1^{-1}Z_1'Z_1$. The choice of $n_2^{-1}\tilde{Z}'\tilde{Z}$ follows from Solon and Inouye's (2010) intuition for the superiority of TS2SLS to TSIV (see note 10), as it adjusts for sampling differences between samples 1 and 2 that appear in $n_2^{-1}\tilde{Z}'\tilde{X}$. Consistent with this, Monte Carlo

C. Measurement error in permanent income

In practice, we are able to follow families only through the middle of their lives. Thus, the best available proxy for p is $\bar{y}_{15}$, the log of the family's average income over a 15-year period. Mazumder (2001) finds that the reliability of a 15-year income average, viewed as a proxy for lifetime income, is about 0.8. The same errors-in-variables results used above imply that both an OLS regression of s^* on $\bar{y}_{15}$ and b and an IV regression that instruments for $\bar{y}_{15}$ with y will overstate the magnitude of the black-white gap conditional on p.[12] Some simple algebra shows that with $\alpha_t = 1$ the ratio of the bias in OLS to the (asymptotic) bias in IV equals var($\bar{y}_{15}$ | p, b) / cov(y, $\bar{y}_{15}$ | p, b). If the e_{it} are i.i.d., the numerator and denominator of this expression are identical, so we should expect the two regressions to yield similar (and similarly biased) estimates. Under more general e_{it} processes, however, the asymptotic bias in the IV estimator may be larger or smaller than that in OLS.

In an empirically relevant case, the bias is smaller in IV than in OLS. Suppose that the e_{is} are independent but not identically distributed across s. Then the above ratio will be greater than one – the bias in the IV estimator will be smaller than that in OLS – if t is chosen so that $\mathrm{var}(e_{it}) < (1/15)\Sigma_s \mathrm{var}(e_{is})$.[13] The noise in annual family income diminishes as mothers reach middle age, so an IV estimator constructed using income from that period is less biased by measurement error in permanent income than is OLS. As we discuss in Section VI, the construction of our sample ensures that relatively low-noise ages are used for the current family income measure.

IV. Data

Our analyses draw on two nationally representative samples. The first is the 1979 National Longitudinal Survey of Youth, a sample of over 12,000 teens and young adults

---

simulations suggest that the estimator based on $n_2^{-1}\tilde{Z}'\tilde{Z}$ performs better than one based on $n_1^{-1}Z_1'Z_1$.

12 In the presence of uncertainty about future incomes, young families' educational investments will depend on *expected* permanent income. Even a very long-run average of realized income will measure this with error, producing further bias.

13 With non-independent e_{it}, the condition is $\mathrm{cov}\left(e_{it}, \frac{1}{15}\sum_s e_{is}\right) < \frac{1}{15}\sum_r \mathrm{cov}\left(e_{ir}, \frac{1}{15}\sum_s e_{is}\right)$.

in 1979 who have been surveyed frequently (annually until 1994 and biennially thereafter) ever since. We use data through 2006, when the youngest respondents were 41 years old. At each survey, respondents are asked detailed questions about their family incomes from various sources. Biological children of female members of the initial sample have been surveyed biennially since 1986, and have been administered standardized tests periodically as they have aged. This sample is known as the "Children of the NLSY," or CNLSY.

The CNLSY testing regime has changed over time, so that the tests administered to (for example) 6-year-olds depend on the year in which they were born. We focus on three scores are relatively consistently available: The Peabody Individual Achievement Test (PIAT) in math, the PIAT reading recognition and reading comprehension tests (which we average and refer to as a "reading" score), and the Peabody Picture Vocabulary Test-Revised (PPVT-R). We use scores on these three tests from the biennial survey corresponding to the year when the child was 10 or 11, as all CNLSY participants should have been administered these tests at that time, and we control for the age (in months) at which the exam was taken.[14] Scores on each test are normalized to mean zero and unit variance based on the CNLSY's 1968 norming sample.

The NLSY sample is representative of people who were age 14-21 at the end of 1978, so our CNLSY subsample is representative of children born before 1996 to women born between 1958 and 1965. It is not representative of all 10-11 year old children from any particular cohort. Most importantly, children born to older mothers are underrepresented in the CNLSY sample. Accordingly, in most of our analyses of the CNLSY data we control for a quadratic in the mother's age at the child's birth.

In each survey year, NLSY respondents are asked detailed questions about income from a variety of sources, such as wages and salary, income from self-employment, unemployment insurance, child support, and public benefits. We form family incomes for each year by summing across each of the various components, including income of the spouse if present. To preserve comparability over time, we

[14] In some cases, the testing protocol was not followed perfectly. We allow testing to have taken place anytime after age 9.5 or before age 12.5. We exclude children with no scores in this three-year window.

exclude any income from an unmarried partner – available only in later waves – from the family income calculation.

We use the family income in the year in which a CNLSY child was tested as his or her current income. To form permanent income, we average the real family income (in 2005 dollars) over the years in which the mother was aged 25 to 39. We also sometimes examine averages over several years prior to the CNLSY test. For these averages, we use only income from even numbered years. That is, we might examine the average of income in the year of the CNLSY test and two years prior, or the average of these two and the income four years prior to the test. This ensures that all of our income measures come from NLSY survey years.

In each year, roughly one-fifth of our sample has missing values in one or more of the income components. If any respondent with a missing component in any survey year were excluded from our permanent income calculation, we would have values for only 29% of CNLSY children with test score data. This would be an excessively restrictive rule, as it would exclude (for example) someone with missing food stamp benefit information in a single year even if all other information were complete. Moreover, even observations with complete data from each survey are missing income data for odd-numbered calendar years after 1994, when the NLSY survey became biennial.

To permit consistent measurement of permanent income for as many observations as possible, we developed an extensive imputation algorithm based loosely on that used by Dahl and Lochner (2010). Where possible, we used information about income of a particular type (e.g., food stamps or child support) from surrounding years to interpolate values for years in which this information is missing. Where there was too much information missing to permit this, we used coarser imputation procedures, though we use these only to construct permanent income; we exclude observations for which current income needed to be imputed this way. Our full imputation algorithm – described in the data appendix – allowed us to form a usable current income for 99% of CNLSY children (unweighted) and a permanent income measure for 94%, with missing values arising primarily when mothers permanently attrited from the NLSY sample before age 39.[15] Log current incomes average 10.72 (standard deviation 0.97), while the log permanent incomes average 10.77 (SD 0.70).

Our second sample is the Early Childhood Longitudinal Survey (ECLS) Kindergarten Cohort. This panel, the basis for several recent studies of the black-white test score gap (e.g., Fryer and Levitt 2004, 2006), follows a random sample of 21,000 students who were enrolled in kindergarten in the 1998-1999 school year. We rely on data updated to the spring of students' 5th grade years. Our analysis focuses on students' math scores from the spring of 5th grade (and, in some analyses, from the spring of 3rd grade), as this permits a rough comparison to the similarly-aged CNLSY sample. We use scaled item response theory (IRT) scores, standardized to have mean zero and unit variance.[16]

The income measures in the ECLS are of much lower quality than those in the NLSY. Each wave of the ECLS contains a single income variable, the parent's report of the total income of all persons in the household, assigned to one of 13 bins. We assign each bin to its midpoint, using $300,000 for the "$200,000 or more" bin, then convert these values to real 2005 dollars. We use the income reported in the spring of 5th grade as the current income for analyses of 5th grade test scores. We also construct a short-run average income from the responses in the springs of Kindergarten, 1st, 3rd, and 5th grades. We set this to missing unless there are at least three non-missing values, two non-imputed.

To zero in on the black-white test score gap, we exclude from our analyses of both the CNLSY and ECLS any respondent who is not either black or non-Hispanic white. Tables 1A and 1B show summary statistics for the two data sets. The first column of each table conditions only on the availability of a math score. The second columns

---

[15] Most respondents from the NLSY's military sample and economically disadvantaged white oversample were dropped from the panel relatively early. Thus, these subsamples represent only 0.1% of our main analysis sample.

[16] Strictly, the sample is students in 2004 who were in kindergarten in 1999 or first grade in 2000. Most but not all were in 5th grade in 2004. The IRT model is updated with each wave of the ECLS, producing changes in both past and current scores. Our analyses of 3rd grade scores uses scores from the 5th grade data release.

exclude families for which we are unable to construct the relevant income variables or our core demographic controls (age, gender, maternal age).[17] The third and fourth columns show statistics for the black and white subsamples.

V. Permanent and Current Income in the NLSY

Table 2 presents several simple analyses of the relationships between race, permanent income, and annual income in the CNLSY. Our sample for all analyses is the same as that in Column 2 of Table 1A: Black and non-Hispanic white children with non-missing demographics and family income (current, lagged, and permanent). Columns 1 and 2 report simple bivariate regressions of current and permanent income, respectively, on an indicator for being black. Column 1 shows that black students' families have current log incomes 0.87 below those of white students' families, on average. This gap falls to 0.64 when we control for gender, child's age, mother's age, year, and birth order (Panel B). The raw gap in permanent incomes (Column 2) is somewhat smaller, but most of the difference disappears when the maternal age control is added.

Column 3 presents a regression of current income on race and long-run average income. When we include our simple demographic controls, the average income coefficient is statistically indistinguishable from one and the R^2 just over 0.6.[18] The black coefficient is indistinguishable from zero, consistent with our maintained assumption that the transitory component of current income when children are age 10-11 is pure noise.

Column 4 reverses this regression, placing long-run average income on the left-hand side and current income on the right as in equation (12). Here, the current income coefficient (which corresponds to R_b / α_t) is just above 0.5. The black coefficient is negative, -0.27, and highly significant. This demonstrates the central fact that underlies our analysis: Even when current incomes are controlled, the black-white gap in

---

[17] In Table 1A, we exclude observations for which we were unable to measure or impute the family income two or four years before the test date, as we use these measures in our core specifications below.

[18] The income coefficient in this regression estimates α_t. Note, however, that our sample is based on the child's age rather than the mother's. Thus, if α is assumed to vary over the life cycle of the mother (Haider and Solon, 2006), the regression estimates a weighted average with weights corresponding to the distribution of the age of mothers of 10-11 year olds in the CNLSY. See the discussion in the appendix.

permanent income remains substantial. Indeed, the residual gap is just a bit less than half as large as the permanent income gap without current income controls from Column 2. Thus, test score regressions that include only a current income measure will dramatically understate – by nearly half – the explanatory power of family income for the black-white test score gap.

Columns 5 and 6 repeat the specifications from Columns 3 and 4, this time excluding the race control. Unsurprisingly, the income coefficients are largely unchanged. Similarly, a comparison between Panels A and B demonstrates that the income patterns are quite robust to the inclusion or exclusion of our demographic controls.

VI. Results

A. Evidence from the CNLSY

Table 3 presents regressions for student scores on the PIAT math exam, given to members of the CNLSY sample at age 10 or 11. Column 1 shows that the raw black-white gap is 0.77 standard deviations when estimated on the maximal possible sample. Column 2 (and the remainder of the table) restricts the sample to families for whom we observe enough information to compute a permanent income as well as annual incomes in the year of the test, two years prior, and four years prior. The gap in this subsample is nearly identical, 0.76.

Column 3 adds the vector of demographic controls used in Table 2: Child gender, the child's age at the time of the exam and its square, the mother's age at the child's birth and its square, the child's parity (entered as dummy variables), and calendar year indicators.[19] These controls bring the gap down to 0.56; a substantial portion of the raw black-white gap is attributable to between-race differences in the distribution of mother's age at the child's birth.

Column 4 adds a control for contemporaneous log family income. This has coefficient 0.21, indicating that a 10% increase in family income is associated with an increase in student test scores of about 0.02 standard deviations. The black coefficient

[19] We have estimated all specifications without the controls, with similar results.

shrinks to -0.43, about one quarter smaller than in Column 3. Columns 5-7 present specifications that use alternative income measures: The average of current income and that two years prior (Column 5); the average of current, two years prior, and four years prior (Column 6); and our long-run average (Column 7). As expected, when we use more information to construct our income measures, the income coefficient gets larger and the black coefficient shrinks toward zero. In Column 7, the black coefficient has fallen to -0.36, down 15% from that in Column 4.

One possible explanation for this result is simply that current income is noisily measured in survey data, and that the reduction of this noise rather than any life cycle considerations explains the rising income coefficient as we use longer and longer income averages. To evaluate this, we re-estimated the specification in Column 4, instrumenting for current income with annual income four years prior. To the extent that the attenuation of the income coefficient in Column 4 is attributable to measurement error, this IV specification should eliminate the bias (insofar as measurement errors in surveys conducted four years apart are independent). In fact, the income coefficient is 0.283, larger than in the OLS specification but significantly lower than in the specification that uses the 15-year income average. The gap is 0.376, similarly between these two.

This indicates that the results reflect true dependence of test scores on long-run average income. It does not, however, indicate that only long-run average income matters. If PIH does not hold exactly then the time path of income may be important as well. Columns 8 and 9 present two specifications aimed at probing the sensitivity of our results to possible violations. First, in Column 8, we include both long-run average and current income in the same specification. If families are unable to borrow and save freely, current income could have a direct effect on current educational investment. Indeed, the current income coefficient is significant in Column 8, albeit much smaller than the long-run average income coefficient. The black coefficient is essentially unchanged from Column 7.

Column 9 probes another possible PIH violation. If young families are uncertain about future income, early-career income may be a more important determinant of educational investments than is realized future income. For roughly two thirds of our child observations, our long-run average income includes annual income observations after the year in which the child took the test. In Column 9 we add a control for the log of average income in the years that the mother was aged 25-31 along with the long-run 25-39 average. The positive coefficient is consistent with the view that early-career income is indeed more important, but it is not significantly different from zero. As in Column 8, however, the the richer income control has essentially no effect on the black coefficient of interest.

Column 10 presents a two-stage-least-squares estimate of the specification in Column 7, using the log of current income as an instrument for the log of the long-run average income. The income coefficient is larger and the black coefficient smaller (in magnitude) than in the corresponding OLS specification in Column 7. As noted earlier, under simple models of the income process the two specifications should identify the same parameters, but they may differ if transitory income shocks are not i.i.d. Appendix Figure 1 presents the standard deviation of (y_{it} - $\bar{y}_{15}$) at different points in the mother's life cycle. This is high in the early 20s, as incomes are quite volatile at this point in the life cycle, then declines to a low point that is maintained from the late 20s through about 35 before rising again. Appendix Figure 2 shows what this implies for the transitory component of family incomes at different points in children's lives. We see that the shocks are smallest at ages 7-12 and higher on either side of this point. As discussed in Section III, bias in our estimates due to measurement error in the permanent income proxy will be smaller in the IV specification than in OLS if the current income instrument is drawn from a point in the life cycle where transitory shocks are below-average in magnitude; Appendix Figure 2 suggests that by using family income when the child is aged 10 or 11 we have done just that.[20] Indeed, when we repeat the IV specification using family income six years before the test date (when the child was aged 4 or 5) as an instrument, the income coefficient drops to 0.326, and the gap increases to 0.365.

Taking the estimates in Table 3 together, it is clear that simply including annual income in a regression severely under-controls for differences in permanent income between black and white families. The black coefficient in Column 10 is only three-

[20] It can also be shown that var($y_{it} - p_i$) < var($\bar{y}_{15}$ - p_i) will generate a non-zero current income coefficient in the specification in Column 8 even if y_{it} has no effect conditional on p_i. Thus, this specification does not provide a sharp test of PIH.

quarters as large as that in Column 4. Stated somewhat differently, the inclusion of an annual income control explains just over half as much of the raw black-white gap (as in Column 3) as is explained by permanent income.

Table 4 presents estimates for all three of the test scores available in the NLSY. The raw black-white gap is much larger on the PPVT than on the PIAT math, and is somewhat smaller on the PIAT reading.[21] However, the general pattern as we compare different income controls is, not surprisingly, very similar: Controlling for current income gets us only about half way to the black-white gap conditional on permanent income.

B. Evidence from the ECLS

Table 5 presents estimates for students' 5th grade math scores in the ECLS. Column 1 presents a regression that includes only a single independent variable, the student's race. The raw black-white gap in the ECLS data is 0.85 standard deviations. This is shrinks slightly to 0.78 when we restrict the sample to observations for which we have data on family income and the mother's age. Column 3 adds controls for the child's gender and age (entered as a quadratic). These have essentially no effect on the black-white gap. Column 4 adds additional quadratic controls for the mother's age at the child's birth. These are necessary for our two-sample analyses, as the CNLSY sample is only representative conditional on the mother's age. The maternal age control explains a notable portion of the gap, shrinking it to 0.62.

Column 5 adds a control for the family's income in the year that the test was taken. This reduces the black-white gap dramatically, to 0.38. Column 6 replaces the current income control with the average of family income across all four ECLS survey waves. The income coefficient is about one-third larger here, and the black-white gap shrinks to 0.34.

In Column 7, we present our TS2SLS specification that uses the CNLSY data to estimate the first-stage relationship between permanent income and the instrument,

---

[21] Recall from Table 1A that the sample standard deviation of PPVT scores is much larger than those for PIAT scores, perhaps indicating a problem with the NLSY score norms. Nevertheless, the black-white gap on the PPVT is notably larger than on either PIAT component even when measured in within-sample z-score units.

current income, and uses the ECLS data to identify the reduced-form relationship between current income and test scores (as in Column 5).[22] The identifying assumption here is that the transitory component of income is uncorrelated with achievement, conditional on race and our other controls. The income coefficient is over 50% larger than in Column 6 (and more than double that in Column 5). The conditional black-white gap is 0.18, less than a third of the raw gap and just over half of the gap controlling for the average income over the ECLS panel.

These estimates almost certainly undercorrect for the role of true permanent income. We assume that the current income measure in the ECLS is equivalent to that in the NLSY, when in practice the former is much inferior and likely less reliable.[23] If so, the income coefficient in Column 7 remains somewhat attenuated, and the black coefficient somewhat negatively biased.

C. Additional controls

It is common when analyzing the conditional black-white test score gap to control for other factors in addition to family income. For example, Phillips et al. (1998) explore controls like parental occupational status, parental wealth, neighborhood average income, and variables capturing the quality of the school and home environment. Some common controls (e.g., parental education and the presence of a father) may proxy for permanent income conditional on current income, thus partly correcting the biases that are the focus of this study. We can use our methods to investigate whether simple controls can adequately address the problem.

Table 6 presents estimates that control for two widely-available and commonly controlled variables that are plausibly good measures of permanent family income,

[22] Unfortunately, the age distributions of mothers in the CNLSY and the ECLS differ due to differences in the sampling schemes of the two surveys, and the first-stage coefficients may depend on the maternal age distribution due to variation in α_t. When we reweight the ECLS sample to match the maternal age distribution in the CNLSY sample, the raw black-white gap is somewhat smaller (0.62 vs. 0.78 standard deviations) but the pattern of coefficients across specifications is similar.

[23] In constructing the first-stage specification, we assign the CNLSY current income measure to the 13 bins used in the ECLS. However, we do not adjust for the added precision presumably gained by the use of a full income module in the CNLSY survey rather than a single question in the ECLS.

maternal education and for the presence of a father.[24] The first panel presents estimates from the CNLSY, while the second presents estimates from the ECLS. Both of the new variables are available in each sample. Column 1 presents estimates without income controls, Column 2 adds current income, Column 3 uses an average income over a longer period instead, and Column 4 presents estimates using our 2SLS (TS2SLS in Panel B) correction. Not surprisingly, the black-white gap is reduced by the inclusion of maternal education and father presence controls (compare column 1 of Table 6 to Column 3 of Table 3 and Column 4 of Table 5). Less expected is that the specification that includes current income yields larger black-white gaps than in the analogous specifications without the new controls. Evidently, conditional on income black students have somewhat *better* family situations than whites. Or, put somewhat differently, mother's education and father's presence do not fully explain the black-white gap in family incomes. The pattern of results across Columns 2, 3, and 4 of Table 6 is similar to that seen earlier: Even when maternal education and family structure are controlled, a model with current family income overstates the conditional black-white gap by 23 (CNLSY) to 62 (ECLS) percent relative to what is obtained when long-run income is controlled via our 2SLS estimator.

As a final exercise, we explore the implications of our analysis for Fryer and Levitt's (2006; hereafter FL) investigation of the black-white test score gap among 3rd graders in the ECLS. In an earlier paper (Fryer and Levitt 2004), these authors showed that a relatively small set of covariates could fully explain the black-white test score gap in kindergarten; in the 2006 paper they find that this is no longer true by the time the same students finish 3rd grade. Columns 1 and 2 of Table 7 report their estimates of the raw gap and the gap conditional on a list of nine covariates, ranging from the child's age and birth weight to measures of mother's age to the number of children's books in the

[24] We use only a contemporaneous family structure variable here. We have also explored specifications that control for the fraction of the child's life in which the father was present, with similar results.

home.[25] The raw gap is 0.88 standard deviations, and the inclusion of the FL controls reduces this to 0.38.

Columns 3 and 4 reproduce Fryer and Levitt's analysis, restricting the sample to just blacks and non-Hispanic whites to correspond with the other estimates presented in this paper.[26] Columns 5 and 6 repeat the estimates on the subsample of students for whom we have non-missing, non-imputed family income. The black-white gap, both unconditional and conditional on the FL covariates, is notably smaller in this subsample, but the conditional gap remains large and significant. Column 7 adds the log of current family income to the specification. The income coefficient is small but significant, while the black coefficient shrinks slightly but is generally similar to that seen in Column 6.

Because not all of the FL variables are available in the NLSY, we cannot estimate the FL specification via TS2SLS. Rather, we apply our proposed adaptation of TS2SLS that blends information from the two samples to estimate the first-stage regression. The key assumption of this estimator is that the transitory component of current income is uncorrelated with any of the ECLS-only control variables conditional on the covariates that are available in both samples. This is clearly false for the socioeconomic status index, as this is constructed from current family income. Columns 8 and 9 repeat the estimates from Columns 6 and 7 without this index. The specification without our family income control yields a slightly larger black-white gap, but that with a control for current income yields a notably smaller gap (and much larger income coefficient) than when the SES index is included.

Even with the mechanically related SES index excluded, the exclusion restriction may be incorrect. It would be violated, for example, if current income were correlated

---

[25] FL include Hispanics and Asians in their sample, with dummy variables for each group. The coefficients on these dummies are not reported in Table 7.

[26] Even when we include the other racial groups, we do not precisely reproduce Fryer and Levitt's sample or results. The most likely explanation is that we use the 5th grade wave of the ECLS data where they (presumably) used the 3rd grade wave. Students who attrited from the survey after 3rd grade are missing from our sample. Other differences between our analysis and the FL specification are that we take all control variables from the 3rd grade survey where possible, while FL appear to have used the kindergarten survey as the source of most covariates; we use the ECLS's 3rd grade cross-sectional weights in place of FL's longitudinal weights; and we present heteroskedasticity-robust standard errors where FL appear to report classical standard errors.

with the number of children's books in the home conditional on the family's permanent income. (Note, however, that there would be no correlation if the household behaved according to the permanent income hypothesis and faced no credit constraints.) Nevertheless, it seems likely to be a reasonably accurate approximation.

Applying our estimator, in Column 10, we see that the long-run average income coefficient is more than double the current income coefficient in Column 9, while the black coefficient is only -0.15. This is just over half of the estimate from a specification with a current income control and much less than half of what is estimated without income controls at all (with or without the SES control). Evidently, even FL's rich specification is unable to effectively control for income differences between black and white families.

VII. Discussion

Previous research has found that family income and other variables measuring a family's external circumstances do a relatively poor job of explaining the black-white test score gap. However, these studies typically control only for family income in the year that the student is tested, perhaps accompanied by weak proxies for permanent income like maternal education. There is little theoretical justification for believing that current income, rather than permanent income, is an important determinant of student achievement, and empirically both current income and human capital measures turn out to be very poor proxies for long-run measures of families' financial circumstances.

We describe a method for identifying the black-white test score gap conditional on permanent income that can be used even when the data set containing student test scores does not itself permit accurate measurement of a family's permanent income. Our method would also be useful for examinations of racial gaps in other outcomes such as educational attainment, asset accumulation (Hurst, Luoh, and Stafford 1998; Mayer 1997), and consumption patterns (Charles, Hurst, and Roussanov 2009).

We find that the association between family permanent income and student achievement is roughly twice as strong as that between current income and achievement. In our preferred 2SLS and TS2SLS estimates, a 10% increase in family permanent income is associated with an increase in child math scores of 0.04 (CNLSY) to 0.07

(ECLS) standard deviations. These coefficients cannot be interpreted causally, as they reflect both the true causal effects of family resources and the confounding effects of other factors that are correlated with both income and economic outcomes. The most obvious omitted variables – e.g., parental ability – would tend to bias the income coefficient upward relative to the causal effect of family income. Our estimated income coefficients are much *smaller*, however, than the plausibly causal effects of family income estimated by Dahl and Lochner (2010).[27]

Understatement of the income coefficient produces overstatement of the black-white test score gap conditional on income. In both the CNLSY and ECLS samples, we find that conventional methods understate the share of the black-white test score gap that is attributable to family income differences by about half. Where the prior literature has indicated that relatively little of the gap can be attributed to family income, we find that family financial circumstances can explain 40 to 75% of the raw gap at age 10 or 11. Moreover, we find that the addition of a control for permanent income to the already-rich covariates considered by Fryer and Levitt (2006) halves the already-small unexplained gap in their specification. Other variables – like maternal education, the presence of a father, or occupation-based socioeconomic status indices – do not do nearly as good of a job of capturing the family circumstances that are related with student achievement and that differ between races. This is not the pattern that one would expect if income is merely proxying for noneconomic family factors. Our analysis thus offers some hope that improvements in black families' economic circumstances could, absent any other changes, lead to substantial closing of the black-white test score gap.

---

[27] Dahl and Lochner (2010) exploit expansions of the Earned Income Tax Credit to identify the effects of shocks to family income that are expected to persist indefinitely. They estimate that a permanent $1,000 increase in family income causes test scores to rise by about 0.06 SDs. As the median family income in their sample is about $18,000, this implies that a 10% increase in family income would lift scores by 0.11 SDs. By contrast, their OLS specifications show much smaller effects of permanent income than those that we obtain. Dahl and Lochner speculate that the large effect in their IV specifications may reflect the low incomes of the disadvantaged families on which their estimates are based or a higher propensity to invest lump-sum EITC payments than ordinary income.

References

Altonji, Joseph G. and Ulrich Doraszelski (2005). "The Role of Permanent Income and Demographics in Black/White Differences in Wealth," *Journal of Human Resources* 40(1), Winter, 1-30.

Angrist, Joshua D. and Alan B. Krueger (1992). "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association* 87 (418), Jun., 328-336.

Blau, David M. (1999). "The Effect of Income on Child Development," *Review of Economics and Statistics* 81(2), May, 261-276.

Blau, Francine D. and Adam J. Grossberg (1992). "Maternal Labor Supply and Children's Cognitive Development," *Review of Economics and Statistics* 74 (3), August, 474-481.

Brooks-Gunn, Jeanne, Pamela Klebanov, and Greg J. Duncan (1996). "Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment, and Maternal Characteristics," *Child Development* 28(6), 1048-1055.

Campbell, Mary E., Robert Haveman, Tina Wildhagen, and Barbara L. Wolfe (2008). "Income Inequality and Racial Gaps in Test Scores." In Magnuson, Katherine and Jane Waldfogel, eds., Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap. Russell Sage Foundation: New York, 110-136.

Carneiro, Pedro and James J. Heckman (2002). "The Evidence on Credit Constraints in Post-Secondary Schooling," *Economic Journal* 112, 705-734.

Charles, Kerwin, Eric Hurst, and Nick Roussanov (2009). "Conspicuous Consumption and Race," *Quarterly Journal of Economics* 124(2), May, 425-467.

Chay, Kenneth Y., Jonathan Guryan, and Bhashkar Mazumder (2009). "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth." National Bureau of Economic Research Working Paper No. 15078.

Dahl, Gordon and Lance Lochner (2010). "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit." Unpublished manuscript, Nov. 19.

Friedman, Milton (1957). A Theory of the Consumption Function. National Bureau of Economic Research: New York.

Fryer, Roland G., Jr. and Steven D. Levitt (2004). "Understanding the Black-White Test Score Gap in the First Two Years of School," *Review of Economics and Statistics* 86(2), May, 447-464.

Fryer, Roland G., Jr. and Steven D. Levitt (2006). "The Black-White Test Score Gap Through Third Grade," *American Law and Economics Review* 8(2), Summer, 249-281.

Grissmer, David and Elizabeth Eiseman (2008). "Can Gaps in the Quality of Early Environments and Noncognitive Skills Help Explain Persisting Black-White Achievement Gaps?" In Magnuson, Katherine and Jane Waldfogel, eds., Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap. Russell Sage Foundation: New York, 139-180.

Haider, Steven and Gary Solon (2006). "Life-Cycle Variation in the Association Between Current and Lifetime Earnings," *American Economic Review* 96, September, 1308-1320.

Hedges, Larry V. and Amy Nowell (1998). "Black-White Test Score Convergence since 1965." In Jencks, Christopher and Meredith Phillips, ed., The Black-White Test Score Gap, Brookings Institution Press: Washington DC, 149-181.

Herrnstein, Richard J. and Charles Murray (1996). The Bell Curve: Intelligence and Class Structure in American Life. Free Press: New York.

Hurst, Eric, Ming Ching Luoh, and Frank Stafford (1998). "Wealth Dynamics of American Families: 1984-1994," *Brookings Papers on Economic Activity* 1.

Jencks, Christopher and Meredith Phillips (1998). "The Black-White Test Score Gap: An Introduction." In Jencks, Christopher and Meredith Phillips, ed., The Black-White Test Score Gap, Brookings Institution Press: Washington DC, 1-55.

Kozol, Jonathan (1991). Savage Inequalities: Children in America's Schools. Crown Publishers: New York.

Krueger, Alan, Jesse Rothstein, and Sarah Turner (2006). "Race, Income, andCollege in 25 Years: Evaluating Justice O'Connor's Conjecture," *American Law and Economics Review* 8(2), Summer, 282-311.

Lareau, Annette (2003). Unequal Childhoods: Class, Race, and Family Life. University of California Press: Berkeley CA.

Magnuson, Katherine and Jane Waldfogel (2008). "Introduction." In Magnuson, Katherine and Jane Waldfogel, eds., Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap. Russell Sage Foundation: New York, 1-30.

Mayer, Susan E. (1997). What Money Can't Buy: Family Income and Children's Life Chances. Harvard University Press: Cambridge, MA.

Mazumder, Bhashkar (2001). "The Mismeasurement of Permanent Earnings: New Evidence from Social Security Earnings Data." Federal Reserve Bank of Chicago Working Paper 2001-24.

Mazumder, Bhashkar (2003). "Revised Estimates of Intergenerational Income Mobility in the United States." Federal Reserve Bank of Chicago Working Paper 2003-16.

Modigliani, F. and R. Brumberg (1954). "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In Kurihara, Kenneth K., ed., Post-Keynesian Economics, Rutgers University Press: New Brunswick NJ.

Moynihan, Daniel Patrick (1965). The Negro Family: The Case for National Action. United States Department of Labor, Office of Policy Planning and Research: Washington DC.

Neal, Derek (2006). "Why Has Black-White Skill Convergence Stopped?" In Hanushek, Eric and Finis Welch, Handbook of Economics of Education. Elsevier: Amsterdam.

Neal, Derek and William Johnson (1996). "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy* 104 (5), October, 869-895.

Phillips, Meredith, Jeanne Brooks-Gunn, Greg J. Duncan, Pamela Klebanov, and Jonathan Crane (1998). "Family Background, Parenting Practices, and the Black-White Test Score Gap." In Jencks, Christopher and Meredith Phillips, ed., The Black-White Test Score Gap, Brookings Institution Press: Washington DC, 103-148.

Solon, Gary (1992). "Intergenerational Income Mobility in the United States," *American Economic Review* 82, June, 393-408.

Solon, Gary and Atsushi Inoue (2010). "Two-Sample Instrumental Variables Estimators," *Review of Economics and Statistics* 92, August, 557-561.

Appendix A: Data

In this appendix, we describe the imputation procedure that we use to fill in missing values in the NLSY income variables. Our procedure is based loosely on that used by Dahl and Lochner (2010), who generously provided us with their programs.

We divide the family's income into 19 components that are reasonably consistently measured in the NLSY. The most important are own wage and salary, spouse's wage and salary, military income for the respondent and for the spouse, self employment income for the respondent and for the spouse, and income "from all other sources," but there are also components reflecting various categories of government transfers (unemployment insurance, welfare, food stamps, SSI, etc.), as well as alimony, child support, and gifts.

We impute missing values for each of these separately. Wage and salary income, which accounts for 77% of total income in our sample, is quite variable across years for many individuals. Much of this variation appears to come from changes in employment status, so we treat employment status – measured as annual weeks worked – and annual full-year-equivalent earnings as distinct sources of variation, imputing the two separately and then multiplying them together. Similarly, we impute marital status and spouse's age separately, and impute values for the spouse's income only if the respondent appears to have been married in the relevant year.

We use the following strategy to impute full-year-equivalent wage and salary income, military income, self-employment income, "other" income, and the corresponding components for the spouse. If there are five or more non-missing values for a specified component for an individual, we estimate an individual-level regression using all non-missing values, with the respondent's (or her spouse's) age and its square as explanatory variables. We then impute missing values using the fitted values from this regression. If fewer than five non-missing values are available, or if the fitted value from the individual-level regression is negative, we instead impute with fitted values from a global regression that uses all individuals in the sample and includes individual fixed effects along with a single quadratic age control.

Information on employment status is available weekly for all years, even if a survey was not conducted. We linearly interpolate to fill in missing values of the fraction of the year the respondent (or spouse) was employed, using data from the year before and the year after the missing observation. We do not extrapolate employment status or interpolate across gaps greater than three years, so wage and salary income cannot be imputed in these years.

For the other income components, we use a simpler procedure: We simply impute the person-specific mean. We do not impute values if there are fewer than three non-missing values for the component.

If we are able to produce actual or imputed values for wages and salary, military income, and self-employment income, we form total family income as the sum of all available income components, using imputed values when actual values are unavailable and assigning zero to components that cannot be imputed. If we are unable to impute any of these three primary income categories, however, we revert to interpolating family income itself using fitted values from a person-specific regression of total family income on age and its square.

We convert family incomes to 2005 dollars and censor the annual values at \$3,373 (the 5th percentile in our sample). We form our permanent income by averaging these censored real incomes over the years when the mother is aged 25-39.

We exclude from our samples observations for which our current or permanent income measures require excessive imputation. First, we drop individuals who attrit from the survey before age 39, for whom we would have to extrapolate family income to years outside of the range for which we have actual values. Second, we exclude individuals for whom we have to interpolate the family income aggregate for any survey year or for more than two of the non-survey years used in the permanent income calculation. Finally, we drop individuals for whom employment status in the year that the child took the test must be imputed.

Our analysis of the NLSY uses custom sampling weights generated for the universe of CNLSY respondents who appear in any survey between 1986 and 2006. In the ECLS, we use weights appropriate for the grade-5 cross-section of children (C6CW0).

Appendix B: Income process

Haider and Solon (2006) assume that $y_{it} = \alpha_t p_i + e_{it}$. Letting t index maternal ages, we estimate α_t by regressing current income y_{it} on long-run average income $\bar{y}_i$ for different values of t. The regressions are estimated on our main NLSY sample, estimating $\bar{y}_i$ as the average of family income from age 25 to 39 and allowing t to vary over the same range. The α_t coefficients, along with 95% confidence intervals, are shown in the left panel of Figure A1, while the right panel shows the root mean squared errors from these regressions. α_t begins low but rises to about 1.1 by the early 30s and stays relatively constant through the end of the 30s. The transitory component, $y_{it} - \alpha_t \bar{y}_i$, is most variable among the oldest and youngest women, with relatively little variation for women in their early 30s.

Our main analysis focuses on 10-11 year old children, whose mothers vary in age. As Table 1A indicates, the average mother in our NLSY sample gives birth in her mid 20s but there is substantial variation around this average. Figure A2 shows the average of α_t and of the standard deviation of the transitory component as functions of the child's age. α rises monotonically, while the transitory component is less variable for 9-year-old children than for older or younger children. Vertical lines in the figure show the average age of CNLSY children at the date of testing. This is near the minimum of the transitory variation curve.

Figure A1: Income process parameters, by mother's age

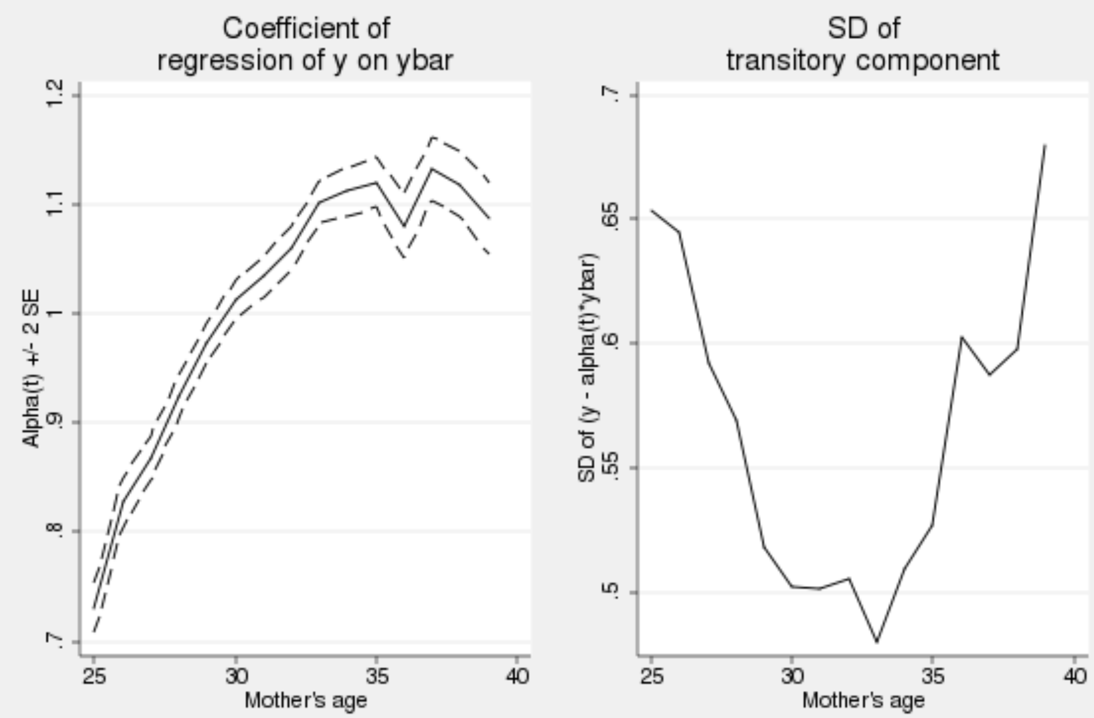

Figure A2: Average income process parameters, by child's age