The role of schools in the production of achievement*

Maria E. Canon†
Federal Reserve Bank of St Louis
October 2011

Abstract

What explains differences in pre-market factors? Three types of inputs are believed to determine the skills agents take to the labor market: ability, family inputs and school inputs. Therefore to answer the previous question it is crucial to understand first the importance of each of those inputs. The literature on the production of achievement has not been able to provide an estimation that can take the three factors into account simultaneously at the student level. This paper intends to fill this gap by providing an estimation of the production function of achievement where both types of investments (families and schools) are considered in a framework where the inputs are allowed to be correlated with the unobserved term, ability to learn. I do that by applying Olley and Pakes’ (1996) algorithm which accommodates for endogeneity problems in the choice of inputs for the production of achievement and by using parents’ saving for their child’s postsecondary education to control for the unobserved component (i.e. ability to learn) in the production of skills. The estimates for the role of family inputs are in line to previous findings. Additionally, the estimates of school inputs show that they are also important for the formation of students’ skills even after controlling for ability to learn.

Keywords: Cognitive Achievement, Literacy
JEL Classification Codes: I21, J15

* I would like to thank Ronni Pavan for his constant guidance and encouragement. For helpful discussions and insightful comments, I thank Mark Bils, Gregorio Caetano, Joshua Kinsler, Ronald Oaxaca, Juan Sanchez, Uta Schoenberg, Nese Yildiz and seminar participants in the Applied Lunch Seminar at the University of Rochester, in the WEAI Graduate Student Dissertation Workshop, LAMES/LACEA 2009, University of Richmond, College of William & Mary, VCU, University of Virginia, Johns Hopkins University, IAD, Yeshiva University, Universitat Autònoma de Barcelona, Universitat Pompeu Fabra, Federal Reserve Bank of St. Louis and University of Western Ontario. I would also like to thank the Institute of Education and Sciences at the U.S Department of Education for providing me the data. The research results and conclusions are mine and do not necessarily reflect the views of the U.S Department of Education. This paper has been screened to insure that no confidential data are revealed.

†E-mail: mcanon@mail.rochester.edu
1 Introduction

The literature on sources of inequality finds that “pre-market” factors; i.e., skills individuals acquire before entering the labor market, explain most of income inequality across individuals and between groups of individuals. In that line, Neal and Johnson (1996) conclude that the observed wage gap between black and white students mostly disappears once we control for “pre-market” factors, measured by the Armed Forces Qualifying Test (AFQT). Likewise, Keane and Wolpin (1997, 2001) and Cameron and Heckman (1998) suggest that labor market outcomes are largely determined by skills acquired during the school-age period. More recently, Carneiro et. al. (2005) find that factors determined outside of the market play a major role in accounting for minority-majority wage differentials in modern labor markets.

Several studies also document that the differences in test scores between blacks and whites widen with age in the school period (see for example Cunha and Heckman (2007a, 2007b), Currie and Duncan (1995), Blau and Currie (2006), Fryer and Levitt (2004)).

However the question of how skills are acquired and evolve over time; i.e., their dynamics, and the importance of the three main inputs (ability, family inputs and school inputs) still remains unsolved. The existing literature has not been able to provide an estimation of the production function of achievement that can take these three factors into account simultaneously at the student level.

Cunha (2007) and Cunha and Heckman (2007a, 2007b) had focused on the unobserved component of the production function; i.e., ability, while keeping school inputs implicit using the National Longitudinal Survey of Youth of 1979 Children and Young adults (NLSY79-CS) which provides reliable information about families’ characteristics and home investments in children. Cunha (2007) estimates a production function using factor analysis in nonlinear settings. He recovers the unobserved distribution of initial skills and that of ability (or heterogeneity at later ages). For the distribution of initial skills he uses characteristics at birth (weight and height), and for heterogeneity at later ages he considers some choices people make later in their lives (such as age at highest grade completed, the number of children they have by 2004, the frequency respondent consumes alcohol in 2004, probation by year 2004).

By merging NLSY79-CS with the Common Core Data (CCD) Lui, Mroz and van der Klaauw (2010) and Todd and Wolpin (2006) provide estimates of the production function of achievement. The CCD includes school variables at the county level as proxies of the true
school inputs students receive. Todd and Wolpin (2006) find that differences in mothers' ability (measured by AFQT) and home inputs explain large portions of test score gaps, while school characteristics have a very small impact. Lui, Mroz and van der Klaauw (2010), estimate a structural model of migration and maternal employment decision. Their finding is that once parental responses are taken into account, that is, after parents adjust their location and the mother’s labor supply decision, policies that exogenously change school characteristics have only minor impact on the child’s test scores.

This paper intends to overcome the aforementioned problems by providing an estimation of the production function of achievement where both types of investments (families and schools) are considered in a framework that accommodates for their relation with the unobserved term, ability. I achieve that by combining Olley and Pakes’ (1996) identification strategy for production functions with a very suitable dataset, NELS:88, which provides information not only on home and school inputs, but also on how much parents save for their children’s postsecondary education. This saving decision helps to identify the third input of the production function: “students’ ability to learn”. What makes this saving measure informative is the fact that parents decide it at the same time they choose the family and school inputs that will affect the observed test score (the current outcome). However those savings will not affect the current outcome, but instead will affect future labor market outcomes through the choice to go to college. To the best of my knowledge this is the first paper using savings for postsecondary education to recover a measure of parents’ knowledge about their children’s ability at the moment they make their input decision. More importantly, this approach allows me to disentangle the effect of ability from the effect of previous test score on current test score (see section 2.2). Identifying the effect of previous achievement on current achievement is important for analyzing the dynamics of educational policies.

The idea of the identification strategy is the following: Every period parents observe their child’s characteristics as well as other house-related characteristics (such as family income, parents’ education) and decide whether to send the child to school, how many inputs to invest in that period and whether or not to save for the child’s postsecondary education (postsecondary education being significantly more expensive than elementary or high school). This

1With “ability to learn” I am referring to the capacity to understand principles, truths, facts or meanings, acquire knowledge, and apply it to practise; the ability to comprehend.
parental saving decision is used to recover the ability, or unobserved, component.\footnote{Cooley (2007) uses a similar invertibility condition to identify students’ ability. She assumes that the portion of leisure time spent reading for fun is a function of students’ ability. She then inverts this function to recover students’ ability.} Therefore the identification strategy relies on the assumption that, conditional upon all the other observable variables affecting the saving decision, there exists a one-to-one mapping between savings and the unobserved term. Section 2.2 provides evidence supporting this assumption. Another potential problem of the identification strategy is whether we are capturing “ability to learn”; i.e., whether savings for postsecondary education are a good proxy for students’ unobserved “ability to learn”. To study that in Section 4.3 I test for the presence of measurement error and a preference shifter, a students’ fixed effect, in the saving function. I do that by incorporating the identification strategy for polynomials errors-in-variable model (Hausman, Newey, Ichimura and Powell (1991) and Hausman, Newey and Powell (1995)) to the standard Olley and Pakes’ algorithm.

I use the estimates of the production function of achievement to perform some counterfactual exercises. Following the literature, I focus on the black-white test score gap. In particular, I do an out of the sample exercise where I equalize the inputs of black students by the differential that white students in the sample are receiving. As opposed to what was found previously in the literature, the results suggest that schools are important in helping blacks to catch up to their white counterparts.\footnote{One exception is a recent work by Hanushek and Rivkin (2009). They find that teacher and peer characteristics explain a substantial share of the widening on the black-white achievement gap between third and eighth grade. Also, using data from Israel, Goud, Lavy and Paserman (2004) find evidence that early schooling environment has an important effect on high school dropout rates, repetition rates, and the passing rate on matriculation exams necessary to enter college.} Moreover, if inputs are altered only in 12th grade, home and school inputs have similar impact on students’ achievement. A policy that gives inputs in both 8th and 12th grade is found to be more effective than intervening only in 12th grade, a result that is consistent with the findings in Cunha (2007).

The order of the paper is the following: Section 2 starts by pointing out the potential estimation problems by using OLS and presents the proposed estimation strategy. Section 3 describes the data and the variables to be used in the estimation and Section 4 presents the empirical results. Finally, section 5 concludes.
2 A proposed framework to overcome estimation problems

The goal is to get consistent estimates of the following production function of achievement (ACH):

\[ h_t^* = f_t(x_t, e_t, H, h_{t-1}, \eta_t) \] (1)

In order to illustrate the estimation problems faced when trying to estimate the \( f_t(.) \) function, consider the value added specification. That is, assume that the production function of achievement is:

\[ h_t = \beta_0 + \beta_1 x_t + \beta_2 e_t + \beta_3 H_t + \beta_4 h_{t-1} + \eta_t + \varepsilon_t \] (2)

where

\[ h_t = h_t^* + \varepsilon_t \]

i.e., \( \varepsilon \) allows for classical measurement error.

If we want to estimate the production function of achievement a natural starting point is Ordinary Least Squares (OLS). In the first subsection I describe the problems we face if we choose that strategy. Then the econometric strategy used in this paper is explained.

2.1 Basic problems of OLS estimates

To illustrate potential biases suppose that cognitive skills are produced according to the following technology:

\[ h_t = \gamma_0 + \gamma_1 (I_t) + \gamma_2 h_{t-1} + \eta_t + \varepsilon_t \] (3)

where \( I \) accounts for all inputs. The main problem of estimating such a technology by OLS is that as econometricians we do not observe students' ability. Thus the error term becomes \( \mu_t = \eta_t + \varepsilon_t \). In this case the OLS estimates are:

\[ \hat{\gamma}_1 = \gamma_1 + \frac{\hat{\sigma}_{h,I} \hat{\sigma}_{I,\eta}}{\hat{\sigma}_{h,h} \hat{\sigma}_{I,I} - \hat{\sigma}_{h,I}^2} - \frac{\hat{\sigma}_{h,\eta} \hat{\sigma}_{I,\eta}}{\hat{\sigma}_{h,h} \hat{\sigma}_{I,I} - \hat{\sigma}_{I,\eta}^2} \]

\[ \text{endogeneity bias} \quad \text{selection bias} \]

\footnote{This specification was first suggested by Hanushek (1986) and has been widely used in the education literature since then. Todd and Wolpin (2006) find, based on a RMSE criterion, the value added specification to be preferred over different reduced form specifications (contemporaneous, cumulative, child fixed effect and sibling fixed effect) for the estimation of the production function.}
where $\hat{\sigma}_{a,b}$ denote the sample covariance between $a$ and $b$. This introduces two types of biases.

- **Endogeneity bias**: following the arguments in Todd and Wolpin (2003) any economic model of optimizing behavior predicts that the amount of resources allocated to a child will be responsive to the parent’s perception of a child’s ability. That is, parents choose inputs once they observe their child’s ability, so we should expect $\sigma_{I,\eta}$ to be different from zero. It could be either positive or negative. On the one hand, it could happen that parents observing that their child is of high ability expect a higher return to their investment and so invest more in him. In this case it would exist a positive correlation between children’s ability to learn and the amount of inputs. On the other it could be that parents who observe that their child is of low ability try to compensate this by investing more inputs. This type of behavior would induce a negative sample correlation between student’s ability to learn and the level of inputs.

- **Selection bias**: it will exist if only those children for whom it is profitable to attend school do not drop out. Therefore, the distribution of unobserved ability to learn in the sample is not the unconditional distribution, but the truncated distribution. If in this context children of higher ability are sent to school under lower realization of previous test scores; i.e., when they do bad in school, because they can catch up later, the truncation point of the ability distribution will be negatively correlated with the previous test score. Hence the sample average of ability will be decreasing in previous test score.

To overcome the two types of biases we need a framework which determines both the information available when inputs decisions are made and an exit rule from schools. The proposed econometric strategy deals with both problems.

### 2.2 Econometric Strategy

The baseline econometric strategy is analogous to Olley and Pakes (1996). I assume that every school year parents observe the stock of human capital of their child, the child’s ability to learn, and other household characteristics such as their family income. I assume as well that ability to
learn follows a first order Markov process. If the child is old enough so that school attendance is no longer compulsory parents can decide, given the information they have, whether the child will go to school or dropout. After high school the child might attend a postsecondary institution, and parents can start saving for their child’s postsecondary education in advance. The basic idea of the identification strategy is that parents make a saving decision to afford postsecondary education while their child is at middle and high school. They make this decision based on their family income, their stock of savings, the child’s achievement and the child’s ability\(^5\); i.e., :

\[ s_t = s_t(FI_t, S_{t-1}, h_{t-1}, \eta_t) \]

In turn, the child’s ability will determine the likelihood of attending college, whether he/she will get financial aid and which type of college he/she will be able to attend. Observing savings for postsecondary education allows me to recover the distribution of ability, as ability affects savings but savings do not affect the child’s level of achievement in the current period.

Conditional on the other set of variables, we can invert this function to back out \( \eta_t \):

\[ \eta_t = f_t(FI_t, S_t, h_{t-1}) \] \hspace{1cm} (4)

Note that to be able to invert this function I rely on the assumption that, conditional on the other set of variables, there exists a continuous and monotonic choice of savings based on the child’s ability.

Two cases might break this monotonicity assumption: if savings depend only on family income or if savings depend only on parents’ expectations about financial aid or some other way their children might find to face the cost of postsecondary education.

In the first case, if savings depend only on family income, then we should expect richer families to save more independently of their children’s level of achievement. If this is the case, conditional on family income there will be no relation between savings and achievement. Figure 1 shows the relation between students performance in math test score and savings for each quartile of family income. It can be observed that parents of children that do better tend to save more for their child’s postsecondary education within each income quartile. Thus, there is evidence that conditional on family income there exists a relation between the saving decision and students’ achievement. The data also shows that savings are not

\(^5\)This result can be obtained for instance from a simple overlapping generation model where parents care about the wellbeing of all their dynasty.
completely explained by the observables included in the saving functions Figure 2 presents the distribution of savings conditional on family income, race, sex, past achievement.

The second case is more complex and requires the consideration of more variables. If
parents of high ability students think their children will be able to get enough merit-based financial aid to face all his/her postsecondary expenses, it might not be optimal to save and reduce current consumption. The child could afford the expenses anyway and parents could increase lifetime utility by increasing current consumption. The same logic could apply for need-based financial aid; i.e., assigned based on family income. If this was the case we could observed a hump shaped relation between savings and ability breaking the monotonicity relation and so the identification strategy. But this argument is based on the assumption that the education decision depends only on the cost. However there could also be a quality dimension on that decision. Therefore, understanding how financial aid works and how it affects net price of attendance\(^6\) at different postsecondary institutions is of vital importance to study whether the identification assumption is correct.

There are three sources of financial aid for students attending postsecondary institutions. The first, and most important in terms of budget, are financial need-based governmental sources (such as the Pell Grant, Perkin Loans, Stafford subsidized loans and the Supplemental Educational Opportunity Grant (SEOG)). The second are governmental loans that are not need based, as the Stafford unsubsidized loans and the PLUS loans, but those are not commonly used in practice. Finally, colleges and universities offer their own grants which in most cases are merit based—in that they are assigned to students that attained a certain GPA in high school (usually above 3.0) or satisfy some other academic criteria.

An additional important aspect is that financial aid does not usually cover all postsecondary expenses. Parents should expect to pay at least half to two-thirds of their children’s college costs through a combination of savings, current income, and loans. Gift aid from the government, colleges and universities, and private scholarships account for only about a third of total college costs. According to data from the Education Department for the 1995-96 academic year Lee (2001) finds that 91.9% of students attending postsecondary schools with tuition and fees above $12,000 receive some direct financial contribution from their parents. Among those students attending institutions with tuitions and fees below $12,000 this percentage was 79.6% in public research universities and 70.8% for other institutions. In sum, for most students the net price of attendance is positive.

Therefore the empirical evidence suggests that: financial aid is based on observables that

\(^6\)Net price of attendance is usually defined as tuition plus room and board minus financial aid, i.e. what the student and/or family must cover after financial aid.
we take care of in our estimation, and for most of the students it does not cover all of their postsecondary education expenses. However, this argument does not rule out the existence of a hump shaped relation between ability and cost of postsecondary education. On one extreme of the distribution we could have students from economically disadvantage families getting financial aid and on the other extreme high ability students getting it. This could imply that only students from the center of the achievement distribution, who would not qualify for either need based or merit based financial aid, would face higher costs. However, this argument does not take into account other aspects that parents might take into account when choosing the postsecondary institutions for their children as for example the quality dimension of postsecondary education. If postsecondary costs are related to the quality of the institution and if the return to schooling depends on this quality, the previous argument does need not hold. To study the relation between a measure of students’ ability and the cost of postsecondary education, I do a nonparametric regression between students’ standardized composite SAT score and the actual net price of attendance they face for the students in the NELS88 sample. Figure 2 shows the result. The relation between SAT and net price of attendance is monotonically increasing, giving us evidence that there are some aspects other than price affecting the choice of the postsecondary institution. The next subsections explain in more detail each of the algorithm’s steps.
2.2.1 First Stage

Replacing \( \eta_t = f_t(H, s_t, h_{t-1}) \) in the production function gives:

\[
h_t = \beta_0 + \beta_1 x_t + \beta_2 e_t + \beta_3 Y_t + \beta_4 h_{t-1} + f(FI_t, S_t, h_{t-1}) + \varepsilon_t \tag{5}
\]

where \( Y_t \) accounts for all the control variables.

The first stage involves estimating (5) semiparametrically using a fourth order polynomial for \( \phi(.) \); i.e., treating \( f(.) \) flexibly. As is noted by Ackerberg et al. (2006) treating \( f(.) \) flexibly has important advantages. The saving function might be a complicated function that depends on the primitives of a model and might be the solution of a dynamic optimization problem. The OP algorithm allows us to avoid both the necessity of specifying the primitives of such a model and the burden of solving it numerically.

Note that because by assumption we are able to completely proxy for \( \eta_t \) the residual in (5) represents factors that are not observed by parents when making their inputs decisions. Therefore we can get consistent estimates of \( \beta_1, \beta_2 \) and \( \beta_3 \). However, the non-parametric treatment of \( f(.) \) does not allow one to separate the effect of \( h_{t-1} \) on the production of current human capital from its effect on the saving decision.

2.2.2 Second Stage

The second stage allows us obtain the intermediate estimates needed to back out the coefficient on lag test score. Note that from the estimates in the first stage we get:

\[
\tilde{\eta}_t = \tilde{\phi}(H, S_t, h_{t-1}) - \beta_0 - \beta_4 h_{t-1} \tag{6}
\]

that is, given a particular set of parameters \( (\beta_0, \beta_4) \) we can have an estimate of \( \eta_t \).

Consider the expectation of human capital net of variable inputs in \( t + 1 \) conditional on the information at the beginning of the period and not dropping from school \( (\chi_{t+1} = 1) \):

\[
E \left[ h_{t+1} - \tilde{\beta}_1 x_{t+1} - \tilde{\beta}_2 e_{t+1} - \tilde{\beta}_3 \tilde{Y}_{t+1}/\eta_t, \chi_{t+1} = 1 \right] \tag{7}
\]

\[
= \beta_0 + \beta_4 h_t + E \left[ \eta_{t+1}/\eta_t, \chi_{t+1} = 1 \right] \\
= \beta_0 + \beta_4 h_t + E \left[ \eta_{t+1}/\eta_t, \eta_{t+1} \geq \eta_{t+1}(H, S_t, h_t) \right] \\
= \beta_0 + \beta_4 h_t + g(\eta_t, \eta_{t+1})
\]
where $g(\eta_t, \eta_{t+1}) = \int_{\eta_{t+1}}^{\eta_t} \frac{F(d\eta_{t+1}/\eta_t)}{\int F(d\eta_{t+1}/\eta_t)}$. To go from the second to the third line we assume that the decision to not drop out from school depends on the child’s ability to learn. Students above a certain threshold will continue in school. I assume that this threshold depends on his stock of achievement, the household characteristics, and the level of savings. All these variables will affect the probability of attending a postsecondary institution and therefore the probability of remaining in school.

In order to control for the selection, we need a measure of both: $\eta_t, \eta_{t+1}$. This is an important difference between the OP algorithm and the standard propensity score literature as in the later only a measure of the threshold value is needed.

We can get a measure of $\eta_t$ from (6), but do not have a measure for $\eta_{t+1}$. What the OP algorithm suggests is to use the data on observed exit to control for $\eta_{t+1}$. Given the previous assumptions, we can write the probability of not dropping out from school in period $t+1$ conditional on the information available in period $t$ as:

$$
\Pr \left( \chi_{t+1} = 1/\eta_t, \eta_{t+1}(H, S_t, h_t) \right) = \Pr \left( \eta_{t+1} \geq \frac{\eta_t}{\eta_{t+1}}(H, S_t, h_t) \right) = p_t \left( \eta_t, \eta_{t+1}(H, S_t, h_t) \right) = P_t
$$

We can estimate (8) non-parametrically, using a fourth order polynomial in $(H, S_t, h_t)$ as the latent index. I assume that the i.i.d. shock received every period to the level of ability to learn follows a normal distribution. This implies that the probability of not dropping out from school follows a normal distribution as well.

Once we have an estimate for $P_t$, we can invert $\hat{P_t}$ to get a measure of $\eta_{t+1}$, i.e. $\eta_{t+1}(\hat{P_t}, \eta_t)$. The only condition we need for that inversion to be possible is that the density of $\eta_{t+1}$ given $\eta_t$ is positive in an area around $\eta_{t+1}(H, S_t, h_t)$. 


2.2.3 Third stage

Substituting $\tilde{P}_t$ and $\tilde{\phi}$ in the production function we can get consistent estimates of the effect of lag test score:

$$h_{t+1} - \tilde{\beta}_1 x_{t+1} - \tilde{\beta}_2 \varepsilon_{t+1} - \tilde{\beta}_3 \bar{Y}_{t+1}$$

$$= \beta_0 + \beta_4 h_t + \eta_{t+1} + \varepsilon_t$$

$$= \beta_0 + \beta_4 h_t + g(\eta_t, \eta_{t+1}) + \varsigma_t + \varepsilon_t$$

$$= \beta_0 + \beta_4 h_t + g \left( \frac{\tilde{\phi}(.) - \beta_0 - \beta_4 h_{t-1}, P_t}{\beta_0 - \beta_4 h_{t-1}, P_t} \right) + \varsigma_t + \varepsilon_t$$

To go from the second to the third line I use the assumption that the child’s ability to learn receives an i.i.d shock received every period ($\varsigma_t$). The estimation is similar to the first stage where we use a fourth order polynomial to approximate $g(\cdot)$, and estimate it by NNLS. Note that because of the approximation for $g(\cdot)$, $\beta_0$ cannot be identified. The identification of $\beta_4$ comes from comparing students with the same $\eta_t$ and $P_t$ but different $h_{t-1}$.

3 Data

NELS:88 is a nationally representative sample of eighth-graders who were first surveyed in the spring of 1988. The original sample employed a two-stage sampling design, selecting first a sample of schools and then a sample of students within these schools. In the first stage the sampling procedure set the probabilities of selection proportional to the estimated enrollment of eighth grade students. In the second stage 26 students were selected from each of those schools, 24 randomly and the other two among hispanic and Asian Islander students. Along with the student survey, NELS:88 included surveys of parents, teachers, and school administrators. A sample of these respondents were resurveyed through four follow-ups in 1990, 1992, 1994 an 2000. Consequently, NELS:88 represents an integrated system of data that tracked students from middle school through secondary and postsecondary education, labor market experiences, and marriage and family formation.(See Appendix C for more details about the survey’s sample and characteristics of each of the five collection years).

**Sample** From the 12,144 individuals in the NELS:88/2000 sample, I exclude those students that in 1988 belong to the “hearing impaired” sample, those students whose parent, teacher or school administrator did not return the questionnaire, and those students with missing test scores.
In 1990 these students can be divided into three groups. For those students who in 1990 dropped out from school, only their 8th grade observation was kept in the sample. Among those students attending school in 1990, if the student was attending a different grade (that is, not in 10th grade) his complete history was deleted. If the student was in 10th grade and his teacher and school administrator answer the questionnaire I keep him in the sample. If any of these questionnaires are not available, his complete history was deleted.

I repeat the same procedure for the 1992 data. If the student dropped out, their observations while in school are kept in the data. For those in school, if the student was attending a grade different from 12th grade, all his observations are deleted from the sample. If the student is attending 12th grade, but either his parents, school administrator or teacher did not completed the questionnaire, all the observations for that student are deleted. Finally, students in 12th grade that had all relevant questionnaires completed are in the sample.

The following figure summarizes how I construct the sample:
The sample resulted in 6,293 students with answers for 1988. In the following years the sample shrinks due to dropouts. Consequently, the final sample is an unbalanced panel of students from 1988 to 1992.

**Achievement Measures** I use the percentile in the math test score distribution as a measure of students’ achievement. All students in the sample took the same test provided by the Education Department Table 1 in Appendix A shows mean values of the math percentile test score in different school grades. Previous literature has focused in both math and read test scores finding qualitatively similar results with both types of tests. Because only the math teacher can be observed across different grades in school, all the analysis is based on the math test score. The descriptive statistics in Table 1 are in line with other datasets, the gap between whites and blacks is important and it is increasing over students life. In 8th grade, the average black student tend to perform in the third decile while the average white student is in percentile 52. At the end of the high-school period, the average white increase one percentile; i.e., it performs in percentile 53, while the average black was in the same percentile as in 8th grade.

**Home inputs** NELS:88 includes a large set of questions related to families activities that might foster or discourage a students’ test scores. Table 1 present some of the measures used in the empirical analysis by race. All the measures show that whites are more likely to receive home inputs that foster achievement than blacks. For example, black children tend to read outside school on average 0.3 hours less than whites in 8th grade, increasing the difference to half an hour by 12th grade. Blacks tend to watch on average more hours of TV than whites. While attending 8th grade a higher proportion of white children are sent to classes outside school (music, language, art) than whites, although this difference seems to narrow over time.

**School inputs** There are two types of school variables in NELS:88. One set of variables are observed at the school level, while the other correspond directly to the class level and are answered by the teacher. Table 1 shows the average value for some of these variables. As in the case of home inputs, there are important differences across race. Black students attend on average schools with worst characteristics to foster their human capital accumulation than white students. For example, blacks are on average in classes with almost twice as many students receiving remedial classes than their whites counterparts (12.21% versus 5.38%). On average blacks are also in classes in which the teacher spends a higher proportion of
the class time just maintaining the order. Both groups seems to have teachers with similar characteristics, in terms of wages, experience and certification.

**Parents characteristics** One advantage of NELS:88 is that it provides information on both parents, mothers and fathers. Table 1 shows the proportion of parents with different years of education. Only 30% of black fathers attain some college or more, while almost 70% of white fathers do. Among mothers differences are important though not as high as with fathers. Having both parents’ education is important to identify the home environment of the child. This seems important as at least in this sample there is evidence that mothers’ education is not always the same as fathers’ education (see Table 2, Appendix A). For example, within children whose mothers just completed high school, only 37% of the fathers had just completed high school, while 17% attended some college, 12% finished college and 34% are high school dropouts. Similar differences are observed for mothers with other schooling levels.

**Savings for postsecondary education** Savings parents make for postsecondary education are observed in every round of the survey where parents are surveyed. First parents are asked whether they have done anything to have some money for their child’s education after high school. For those that answer yes, they are asked how much money they had set aside. Table 1 shows that although there are big differences by race, the average saving to family income ratio is similar across races. In the empirical analysis, this differences jointly with other family and child’s characteristics are going to give the identification for heterogeneity across kids.

4 Empirical Results

This section presents the estimates for the production function of achievement using the OP algorithm described in section 2.2. Two additional controls will be added to the ability function. The first one is race. Equation (4) assumes that there are not systematic differences in savings for postsecondary education across races while the empirical evidence suggests that it might not be the case, see for example Oliver and Shapiro (2006). I will control for race in equation (4) to take care of this potential effect. Equation (4), assumes also that there are no gender differences. However, there is evidence that girls perform relatively below boys in math test scores (see for example Todd and Wolpin (2006)). Without taking into account gender differences and given that all the analysis is based in math test scores, I might
underestimate the “ability to learn” of girls and this difference would be captured by the sex coefficient. Therefore I will control for gender in equation (4) as well. The identification of these coefficients is similar to the coefficient on the lag test score explained in section 2.2.

I also need to search for the mean and standard deviation for the shock to ability to learn. For that, I added a loop on top of the algorithm described in section 2.2. The starting point is to assume that the shock follows a standard normal distribution (as in the original OP algorithm). Using the estimates I calculate the implied distribution for the shock, and use this as the assumed distribution. I stop once the difference on the parameters between the assumed and implied distributions is sufficiently small.

I present first the estimates for a value added production function and compared them with the standard OLS estimates for the same function. Then the estimates of a technology that allows for dynamic complementarity/substitutability are presented.

In Section 4.2 I use these estimates to perform some counterfactual exercises. Following the literature, most of the counterfactual exercises focus on the black-white test achievement gap. In particular, I study how the actual gap would change by exogenously altering the inputs a group of students receive. Finally Section 4.3 presents robustness analysis for the assumption that savings for postsecondary education is a good proxy for students’ ability to learn.

4.1 Estimates of the Production Function of Achievement

Table 3 presents the estimates for the production function under alternative specifications. The first column presents the estimates using OLS. Both home and school inputs are statistically significant for the production of achievement. The coefficient in the lag test score is statistically significant, positive and less than one. This result is in line previous literature. For example Currier and Thomas (2000) find that the effect of Head Start tend to fade out over time when not followed up by later investments.

The second column shows the estimates when the OP algorithm is used. The most important change with respect to the OLS estimates is in the effect of the lag test score, with its coefficient increasing in more than eight standard deviations. This result is in line with the standard predictions in the IO literature. Parents of children with larger stocks of human capital, $h_{t-1}$, should expect future higher returns for any level of their kid’s ability to learn.
learn, hence they will choose to send their children to school under lower realizations of their ability. Consequently, we should expect the truncation point of students’ ability to learn to be decreasing in $h_{t-1}$ and if the production function of achievement is increasing in $h_{t-1}$ this would imply a negative bias in the OLS estimate of its coefficient.\footnote{It could be argued that the change in the coefficient of the lag test score is due to measurement error. However, if the lag test score is instrumented with its lag the coefficient of the lag test score increases only by three standard deviations.}

For the home and school inputs their coefficients are jointly statistically different under the OP estimations and the OLS estimation. Under the null hypothesis that they are equal, we get a $\chi^2$ of 41.85 with a p-value of 0.0186. As was mentioned in section 2.1, a priori the bias from the OLS estimates in the home and school inputs could go either way depending on whether the substitution or wealth effect dominates. In this sample there is evidence that the substitution effect dominates as in most cases the impact of individual inputs goes down once we eliminate the endogeneity problem. If we increase all the inputs by one standard deviation the OLS estimates predict an increase of 0.95 standard deviations of the average math test score, while the OP estimates predict that the average math test score would increase by 0.68 standard deviation.

In line with previous results in the literature there is evidence of sensitive periods. In this sense, the return to reading an extra hour decreases by 96% from 8th grade to 12th grade and it is not significant in the last case. In terms of the school inputs, the effect of the proportion of time that teachers use just to maintain order in the class decreases by 56%.

Table 4 presents the estimates of a technology that allows for dynamic complementarity/substitutability in the production of skills. In particular, I estimate the following production function:

$$h_t = \beta_0 + \beta_1 x_t + \beta_2 e_t + \beta_3 Y_t + \beta_4 h_{t-1} + \gamma h_{t-1}(\beta_1 x_t + \beta_2 e_t) + \eta_t + \xi_t$$

Note that in order to make the estimation more tractable all the inputs are compacted in an index, where each input’s weight is their contribution to the production of achievement. This joint effect can be identified in the first stage of the algorithm, estimating this stage through NNLS, because current period inputs do not enter in the fourth order polynomial used to approximate the unobservable component.

The estimates show that the effect of skills in the previous period changes significantly.
Now $\beta_4$ is smaller and $\gamma$ is positive and significant. From equation (9):

$$\frac{\partial h_t}{\partial h_{t-1}} = \beta_4 + \gamma(\beta_1 x_t + \beta_2 e_t)$$

(10)

that is if investments are not followed by subsequent investments achievement tends to fade out more rapidly. There is also strong evidence for dynamic complementarity, the return to current investment is 1.55 times higher the higher the stock of skills acquired in previous periods.

The estimated mean of the shock to ability to learn is 0.041 while the standard deviation is 0.258.

4.2 Accounting Exercise

There exists a growing literature interested in understanding the production of skills because they are important determinants of labor market outcomes. To understand the effect of different inputs, in this section I use the estimates from table 4 to study the impact of exogenously altering different types of inputs on students’ achievement. Table 5 shows how the predicted black-gap would change under different scenarios. The first exercise shows the test score gap if black students would receive in addition the differential of what white students receive. That is, for each input I regress the quantity of the input on family income, parental education, sex, race and past achievement and a white dummy. I reassigned black students the actual amount they receive plus the differential that whites students receive. The estimated math test score imply that home inputs would reduce the achievement gap by 15.6% while equalizing school inputs would do it with 9.2%. The second exercise is a “late remediation policy” where blacks receive additional inputs only in 12th grade. The effect of both types of inputs in closing the gap decrease. School inputs would reduce the gap by 7.2% while home inputs would do it with 7.4%. That is, during the high school period the role of school inputs is closer to that of home inputs.

"The estimates of the production function show that the lag test score is an important input. One important limitation of NELS:88 is that all the inputs information begins in 8th grade, when already many things had occurred in children’s life. To see how much of the gap is due to these differences in skills at the beginning of 8th grade, I calculate predicted tests scores using the estimated parameters and giving every student the maximum initial test score observed in the sample. In this case 16.4% of the gap would be closed, implying
that initial conditions are important for future performance. This result is also in line with previous findings in the literature which suggest that early investments are more productive than investment in latter ages."

4.3 Robustness: Estimation problems when savings is a poor proxy of unobserved ability to learn

In order to estimate the contribution of home and school inputs to the production of achievement I suggest that: $s_{it} = f(\eta_{it}, controls)$, and assumed that this is the true function generating savings. In a more general set up, it could be that $s_{it} = f(\eta_{it}, controls, \zeta_{it})$, where $\zeta_{it}$ is an additional, unobserved for the econometrician, component. For example $\zeta_{it}$ could be measurement error. Alternatively, we could think that $\zeta_{it}$ is a variable that affects the true savings function, like parents’ generosity towards their children. In this section I study whether the inclusion of such a variable could bias the results and how I could solve this issue.8

To derive the estimation problems we would face, assume the production function of achievement is governed by the following technology:

$$h_{it} = \gamma_0 + \gamma_1 I_{it} + \nu_{it}$$

where $\nu_{it} = \eta_{it} + \varepsilon_{it}$, and where $\eta_{it}$ is ability to learn and $\varepsilon_{it}$ represents some idiosyncratic shock to the production of achievement or just classical measurement error in test scores. Without loss of generality, assume both components have mean zero. As mentioned in Section 2.1, any economic model of optimizing behavior would predict that the amount of resources allocated to a child will be responsive to the parent’s perception of a child’s ability, that is $\text{cov}(I_{it}, \eta_{it}) \neq 0$. To simplify the analysis assume that $\text{cov}(I_{it}, \varepsilon_{it}) = 0$.

In order to be illustrative lets keep aside the control variables considered in the saving function, that is suppose that in addition to $\zeta_{it}$ the only determinant of savings is ability to learn, and consider the simple case where the previous function is linear; i.e.,:

$$s_{it} = \beta \eta_{it} + \zeta_{it}$$

In this section I study what would be the estimates if the saving function depends on other unobservable different than ability to learn. It could be argued that parents could choose other type of investments in child’s achievement. Moving to a neighborhood with better schools or sending their child to private school could be an alternative. However, the estimates do not change significantly when the sample is restricted either to those students that do not change neighborhood or to students attending public schools. Tables 7 and 8 in the Appendix show how much of the gap would family and school inputs closed under these two restricted samples.
If this is the true relation between savings and ability to learn, then can we use savings to solve the endogeneity and selection problems of the OLS estimation? Yes, in fact it can be shown (see Appendix B) that the bias using the savings as a proxy variable for unobserved ability to learn can never be higher than the bias of the OLS estimates and equals:

$$E(\widehat{\gamma}_1 - \gamma_1) = \frac{-\sigma_{sIt} \sigma_{ist}}{\sigma_{ist}^2} + \frac{\sigma^2_{\zeta}}{\beta^2 \sigma^2_{\eta} \left( 1 - r^2_{\eta\mu} \right) + \sigma^2_{\zeta}^2} E(\gamma_{OLS}^1 - \gamma_1)$$

The extent of this bias will depend on the correlation between the regressors of interest (home and school inputs) and the proxy variable ($\sigma_{ist}$) and the correlation between the proxy variable and the error term ($\sigma_{sIt}$).

To study whether the estimators are consistent we need additional information. Given that the panel gives only two saving observations per student, I will take into account two cases for the stochastic component for which I can get identification in the sample.

### 4.3.1 Identification under alternative specifications

**Measurement error** Suppose that the unobserved term $\zeta_{it}$ is pure measurement error. That is, assume it is i.i.d across individuals and time, so that $cov(\zeta_{it}, \zeta_{it-1}) = 0$. In this case we can use savings from a different period, say $s_{i,t-1}$, to instrument for $s_{it}$, as in this case $cov(s_{i,t-1}, \zeta_{it}) = 0$. This procedure is called “multiple indicator solution”. To see why the bias vanishes, consider again

$$E(\widehat{\gamma}_1 - \gamma_1) = \frac{-\sigma_{sIt} \sigma_{ist}}{\sigma^2_{ist}}$$

if we use $s_{i,t-1}$ to predict $s_{it}$ ($s_{it} = \theta s_{i,t-1} + \zeta_{it}$), then:

$$\sigma_{sIt} = E \left[ \sum \hat{\theta} s_{i,t-1} \mu \right] = E \left[ \sum \hat{\theta} s_{i,t-1} \left( \varepsilon_{it} - \frac{1}{\beta} \zeta_{it} \right) \right]$$

where we use $\mu_{it} = \varepsilon_{it} - \frac{1}{\beta} \zeta_{it}$ as defined in (13) Appendix B. It follows that

$$\sigma_{sIt} = E \left[ \hat{\theta} \sum s_{i,t-1} \varepsilon_{it} - \frac{\hat{\theta}}{\beta} \sum s_{i,t-1} \zeta_{it} \right] = 0$$

given that $cov(s_{i,t-1}, \zeta_{it}) = 0$. Which implies $E(\widehat{\gamma}_1 - \gamma_1) = 0$.

To test the presence of measurement error in the saving function we can run the same set of regressions as in section 4.1 but instrumenting the current period saving with savings
in a different period. The problem in this case is that in the estimation of Section 4.1 a
different order polynomial expansion of parent’s savings for postsecondary education is used.
That is, we have a polynomial errors-in-variables model. Identification in this case is still
possible, but not by using a linear projection in the first stage. The problem is that powers
of the measurement error get interacted with the coefficients, and to identify them we need
to identify those moments of the error term as well. If a linear projection of the type of
\( s_{i,t} = \theta s_{i,t-1} + \xi_{it} \) is used, the estimated coefficients would be a linear combination of the true
coefficients and different moments of the measurement error term. Hausman, Newey, Ichimura
and Powell (1991) and Hausman, Newey and Powell (1995) proposed an identification strategy
for polynomial errors in variable models, and that is the one I follow in this case. To get
identification, I combine a linear projection in the first stage ith moment condition between
the dependent variable and the instrument, and the variable measured with error and the
instrument (see the cited literature for details).

Preference Shifter Suppose \( \xi_{it} \) is persistent over time for each student; i.e., \( \xi_{it} = \xi_{it-1} = \xi_i \). This scenario could arise if the unobserved component of savings is due to parents’ pre-
ference parameter, like parents’ generosity toward their children. In this case, instrumenting
current period savings with savings in a different period would not eliminate the bias as
\( \text{cov}(s_{i,t-1}, \xi_i) \neq 0 \). Instead the change in savings \( s_{i,t} - s_{i,t-1} = \beta(\eta_{it} - \eta_{it-1}) \) does. Change
in savings is correlated with savings, but not with \( \xi_i \) and so it allows us to get consistent
estimates of our parameters of interest.

4.3.2 Estimation Results

Both instruments require that we observe two savings observations for each students, therefore
all the estimations in this section include only the balance panel of students that do not drop
out from school between 8th grade and 12th grade. Consequently I cannot estimate the
survival probabilities needed to identify the coefficients in the third stage of the algorithm.
To be able to run the third stage, I use each student’s survival probabilities estimated from
the unbalance panel. I use these probabilities for the three measures of savings: current period
saving (the suggested measure), savings in a different period (measurement error case), and
change in savings (preference shifter case). I reestimate the OLS coefficients for this balance
panel as well.
Using the estimates of each alternative specification I compute how much of the black-white gap could be closed by giving black students the differential of white students receive on top of what they are receiving. Table 6 presents the results under the alternative specifications, the original OLS and OP value added specifications and the two OP estimations using each instrument. In comparison with its OLS alternative, the OP estimator using current period saving does much better. The OP estimator predicts that both types of inputs, home and school, would close the gap in a smaller fraction than what OLS predicts. In both cases the OP prediction is closer to what the instrument for each extreme case predicts.

5 Conclusion

The existing literature on sources of inequality find that “pre-market” factors, skills individuals acquire before entering the labor market, explain most of income inequality across individuals and between groups of individuals. But what explains differences in pre-market factors? A growing literature in economics tries to provide an answer to this question by studying children’s performance in test scores. This paper contributes to that literature by proposing an identification strategy that accommodates for usual endogeneity problems in the choice of inputs and to the choice on whether to attend schools or not and applying it to a very suitable data set for this problem: NELS:88. NELS:88 provides information of both home and school inputs at the student level as well as parents’ saving for their child postsecondary education that I use to control for the unobserved component (i.e., ability to learn) in the production of skills. This allows me to recover the parameters of interest in the production function of achievement: the effect of period by period investment as well as the impact of the achievement acquired in previous periods. The estimates show that in fact the most significant change from applying the proposed strategy rather than an OLS estimation occurs in the lag test score. Additionally, I find evidence that these savings are not a poor proxy for students’ unobserved ability to learn.

The estimates for the role of family inputs are in line to previous findings, they foster students achievement and there exists sensitive periods. However, the estimates of school inputs show that, contrary to what has been found in the literature, they are important for the formation of students’ skills and they seem to be as important as home inputs if late remediation policies are considered.
References


6 Appendix A: Tables

Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>8th Grade</th>
<th></th>
<th>12th Grade</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Black</td>
<td>White</td>
<td>Black</td>
<td>White</td>
</tr>
<tr>
<td>Standardized math test score</td>
<td>0.300</td>
<td>0.520</td>
<td>0.290</td>
<td>0.530</td>
</tr>
<tr>
<td></td>
<td>(0.245)</td>
<td>(0.284)</td>
<td>(0.249)</td>
<td>(0.284)</td>
</tr>
<tr>
<td><strong>Home Inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours reading outside school</td>
<td>1.748</td>
<td>2.032</td>
<td>2.147</td>
<td>2.585</td>
</tr>
<tr>
<td></td>
<td>(1.719)</td>
<td>(1.912)</td>
<td>(2.118)</td>
<td>(2.497)</td>
</tr>
<tr>
<td>Hours of TV per week</td>
<td>3.499</td>
<td>2.673</td>
<td>3.209</td>
<td>2.085</td>
</tr>
<tr>
<td></td>
<td>(1.651)</td>
<td>(1.514)</td>
<td>(1.694)</td>
<td>(1.447)</td>
</tr>
<tr>
<td>Special lessons</td>
<td>0.296</td>
<td>0.419</td>
<td>0.140</td>
<td>0.138</td>
</tr>
<tr>
<td><strong>School Inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private school</td>
<td>0.142</td>
<td>0.225</td>
<td>0.069</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.157)</td>
<td>(0.144)</td>
<td>(0.141)</td>
</tr>
<tr>
<td></td>
<td>(7.644)</td>
<td>(7.540)</td>
<td>(10.306)</td>
<td>(9.168)</td>
</tr>
<tr>
<td>Teacher certified</td>
<td>0.478</td>
<td>0.503</td>
<td>0.630</td>
<td>0.645</td>
</tr>
<tr>
<td>Class enrollment</td>
<td>26.596</td>
<td>24.476</td>
<td>27.083</td>
<td>27.058</td>
</tr>
<tr>
<td></td>
<td>(13.096)</td>
<td>(11.005)</td>
<td>(14.935)</td>
<td>(17.026)</td>
</tr>
<tr>
<td>Prop. of stud receiving remedial classes</td>
<td>12.227</td>
<td>5.377</td>
<td>12.175</td>
<td>5.797</td>
</tr>
<tr>
<td></td>
<td>(14.091)</td>
<td>(7.316)</td>
<td>(13.576)</td>
<td>(7.286)</td>
</tr>
<tr>
<td>Proportion of class maintaining order</td>
<td>2.100</td>
<td>1.990</td>
<td>1.815</td>
<td>1.591</td>
</tr>
<tr>
<td></td>
<td>(0.853)</td>
<td>(0.785)</td>
<td>(1.019)</td>
<td>(0.761)</td>
</tr>
<tr>
<td><strong>Family characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother High School</td>
<td>0.667</td>
<td>0.777</td>
<td>0.713</td>
<td>0.811</td>
</tr>
<tr>
<td>Mother Some College</td>
<td>0.346</td>
<td>0.429</td>
<td>0.370</td>
<td>0.441</td>
</tr>
<tr>
<td>Mother College</td>
<td>0.128</td>
<td>0.210</td>
<td>0.131</td>
<td>0.205</td>
</tr>
<tr>
<td>Father High School</td>
<td>0.357</td>
<td>0.642</td>
<td>0.367</td>
<td>0.672</td>
</tr>
<tr>
<td>Father Some College</td>
<td>0.205</td>
<td>0.422</td>
<td>0.216</td>
<td>0.442</td>
</tr>
<tr>
<td>Father College</td>
<td>0.112</td>
<td>0.263</td>
<td>0.104</td>
<td>0.261</td>
</tr>
<tr>
<td>Savings post secondary education</td>
<td>4.022</td>
<td>5.938</td>
<td>5.443</td>
<td>9.099</td>
</tr>
<tr>
<td></td>
<td>(4.138)</td>
<td>(5.087)</td>
<td>(6.830)</td>
<td>(9.357)</td>
</tr>
<tr>
<td>Family Income</td>
<td>29,729</td>
<td>49,464</td>
<td>36,658</td>
<td>57,026</td>
</tr>
<tr>
<td></td>
<td>(26,781)</td>
<td>(39,993)</td>
<td>(32,500)</td>
<td>(42,512)</td>
</tr>
</tbody>
</table>
Table 2: Distribution of fathers’ education by mothers’ education

<table>
<thead>
<tr>
<th>Father’s Education</th>
<th>Mother’s Education</th>
<th>High School</th>
<th>Some College</th>
<th>College</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School</td>
<td>37.1%</td>
<td>16.4%</td>
<td>6.6%</td>
<td></td>
</tr>
<tr>
<td>Some College</td>
<td>16.6%</td>
<td>26.2%</td>
<td>10.9%</td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>12.0%</td>
<td>28.3%</td>
<td>66.2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OLS</td>
<td>O-P estimator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------</td>
<td>---------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag test score</td>
<td>0.4951</td>
<td>0.6911</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0142)</td>
<td>(0.0209)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read in 8thG</td>
<td>0.0366</td>
<td>0.0375</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0073)</td>
<td>(0.0074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read in 8thG squared</td>
<td>-0.0030</td>
<td>-0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Go to class with pencil 8thG</td>
<td>0.0244</td>
<td>0.0280</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0110)</td>
<td>(0.0111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special lessons 8thG</td>
<td>0.0290</td>
<td>0.0275</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0088)</td>
<td>(0.0089)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family have books at home 8thG</td>
<td>0.0513</td>
<td>0.0465</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0182)</td>
<td>(0.0183)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours of TV per week 8thG</td>
<td>-0.0169</td>
<td>-0.0154</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0028)</td>
<td>(0.0028)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time with parents 8thG</td>
<td>-0.019</td>
<td>-0.0209</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0523)</td>
<td>(0.0536)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(teacher wage 8thG)</td>
<td>0.1038</td>
<td>0.0820</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0282)</td>
<td>(0.0285)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher experience 8thG</td>
<td>0.0011</td>
<td>0.0012</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/(% students with single parents 8thG)</td>
<td>0.4297</td>
<td>0.4524</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1596)</td>
<td>(0.1606)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours of classes 8thG</td>
<td>0.0668</td>
<td>0.0552</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0356)</td>
<td>(0.0356)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/(% students attending remedial classes 8thG)</td>
<td>0.0272</td>
<td>0.0209</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td>(0.0104)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/(prop class time teacher spends maintain. Order 8thG)</td>
<td>0.0970</td>
<td>0.0973</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0188)</td>
<td>(0.0188)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 (contd.): Regression results grade-dependent technology

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>O-P estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read in 12th G</td>
<td>0.0032</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0024)</td>
</tr>
<tr>
<td>Go to class with pencil 12th G</td>
<td>0.0545</td>
<td>0.0563</td>
</tr>
<tr>
<td></td>
<td>(0.0210)</td>
<td>(0.0212)</td>
</tr>
<tr>
<td>Family rule about hw 12th G</td>
<td>0.0353</td>
<td>0.0274</td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
<td>(0.0136)</td>
</tr>
<tr>
<td>Participate in community activities 12th G</td>
<td>0.0358</td>
<td>0.0304</td>
</tr>
<tr>
<td></td>
<td>(0.0119)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>Go to the Theater 12th G</td>
<td>-0.0028</td>
<td>-0.0017</td>
</tr>
<tr>
<td></td>
<td>(0.0178)</td>
<td>(0.0179)</td>
</tr>
<tr>
<td>Time with parents 12th G</td>
<td>0.0099</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>(0.0184)</td>
<td>(0.0185)</td>
</tr>
<tr>
<td>Log(teacher wage 12th G)</td>
<td>0.0776</td>
<td>0.0660</td>
</tr>
<tr>
<td></td>
<td>(0.0396)</td>
<td>(0.0404)</td>
</tr>
<tr>
<td>Teacher has a master degree 12th G</td>
<td>0.0224</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>(0.0115)</td>
<td>(0.0112)</td>
</tr>
<tr>
<td>Teacher certified in math 12th G</td>
<td>0.0306</td>
<td>0.0221</td>
</tr>
<tr>
<td></td>
<td>(0.0128)</td>
<td>(0.0129)</td>
</tr>
<tr>
<td>1/(% students attending remedial classes 12th G)</td>
<td>0.0079</td>
<td>0.0081</td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
<td>(0.0138)</td>
</tr>
<tr>
<td>1/(% students with single parents 12th G)</td>
<td>0.2334</td>
<td>0.2150</td>
</tr>
<tr>
<td></td>
<td>(0.1009)</td>
<td>(0.1023)</td>
</tr>
<tr>
<td>1/(prop class time teacher spends maintain. Order 12th G)</td>
<td>0.0885</td>
<td>0.0663</td>
</tr>
<tr>
<td></td>
<td>(0.0208)</td>
<td>(0.0213)</td>
</tr>
</tbody>
</table>

all specifications control for parents’ education, race and sex
standard errors in parenthesis
Table 4: Regression results grade-dependent technology with cross effects

<table>
<thead>
<tr>
<th>Term</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag test score</td>
<td>0.5515</td>
<td>(0.0191)</td>
</tr>
<tr>
<td>Interaction Lag test score current inputs</td>
<td>1.5491</td>
<td>(0.6628)</td>
</tr>
<tr>
<td>Read in 8thG</td>
<td>0.0233</td>
<td>(0.0059)</td>
</tr>
<tr>
<td>Read in 8thG squared</td>
<td>-0.0020</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>Go to class with pencil 8thG</td>
<td>0.0182</td>
<td>(0.0076)</td>
</tr>
<tr>
<td>Special lessons 8thG</td>
<td>0.0176</td>
<td>(0.0063)</td>
</tr>
<tr>
<td>Family have books at home 8thG</td>
<td>0.0297</td>
<td>(0.0125)</td>
</tr>
<tr>
<td>Hours of TV per week 8thG</td>
<td>-0.0097</td>
<td>(0.0024)</td>
</tr>
<tr>
<td>Time with parents 8thG</td>
<td>-0.0869</td>
<td>(0.0343)</td>
</tr>
<tr>
<td>Log(teacher wage 8thG)</td>
<td>0.0397</td>
<td>(0.0186)</td>
</tr>
<tr>
<td>Teacher experience 8thG</td>
<td>0.0008</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>1/(% students with single parents 8thG)</td>
<td>0.2004</td>
<td>(0.1055)</td>
</tr>
<tr>
<td>Hours of classes 8thG</td>
<td>0.0286</td>
<td>(0.0223)</td>
</tr>
<tr>
<td>1/(% students attending remedial classes 8thG)</td>
<td>0.0127</td>
<td>(0.0067)</td>
</tr>
<tr>
<td>1/(prop class time teacher spends maintain. Order 8thG)</td>
<td>0.0600</td>
<td>(0.0153)</td>
</tr>
</tbody>
</table>
Table 4 (Contd.): Regression results grade-dependent technology with cross effects

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read in 12thG</td>
<td>0.0009</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>Go to class with pencil 12thG</td>
<td>0.0254</td>
<td>(0.0123)</td>
</tr>
<tr>
<td>Family rule about hw 12thG</td>
<td>0.0103</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>Participate in community activities 12thG</td>
<td>0.0138</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>Go to the Theater 12thG</td>
<td>-0.0002</td>
<td>(0.0091)</td>
</tr>
<tr>
<td>Time with parents 12thG</td>
<td>0.0018</td>
<td>(0.0096)</td>
</tr>
<tr>
<td>Log(teacher wage 12thG)</td>
<td>0.0281</td>
<td>(0.0211)</td>
</tr>
<tr>
<td>Private School 12thG</td>
<td>0.0071</td>
<td>(0.0061)</td>
</tr>
<tr>
<td>Teacher certified in math 12thG</td>
<td>0.0128</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>1/(% students attending remedial classes 12 G)</td>
<td>0.0053</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>1/(% students with single parents 12thG)</td>
<td>0.0933</td>
<td>(0.0549)</td>
</tr>
<tr>
<td>1/(prop class time teacher spends maintain. Order 12thG)</td>
<td>0.0265</td>
<td>(0.0123)</td>
</tr>
<tr>
<td>Description</td>
<td>Value</td>
<td>Percentage</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>Predicted gap</td>
<td>0.214</td>
<td></td>
</tr>
<tr>
<td>gap closed by home inputs</td>
<td>0.034</td>
<td>15.6%</td>
</tr>
<tr>
<td>gap closed by school inputs</td>
<td>0.020</td>
<td>9.2%</td>
</tr>
<tr>
<td>gap closed by giving inputs only in 12th grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>home inputs</td>
<td>0.016</td>
<td>7.4%</td>
</tr>
<tr>
<td>school inputs</td>
<td>0.015</td>
<td>7.2%</td>
</tr>
<tr>
<td>gap closed by giving the same initial test score</td>
<td>0.035</td>
<td>16.4%</td>
</tr>
<tr>
<td>Model</td>
<td>Predicted Gap</td>
<td>Gap Closed by Home Inputs</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>---------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>OLS predicted gap</td>
<td>0.235</td>
<td>0.023</td>
</tr>
<tr>
<td>OP current saving predicted gap</td>
<td>0.235</td>
<td>0.021</td>
</tr>
<tr>
<td>OP other period saving predicted gap</td>
<td>0.235</td>
<td>0.020</td>
</tr>
<tr>
<td>OP change in savings predicted gap</td>
<td>0.235</td>
<td>0.020</td>
</tr>
</tbody>
</table>
Table 7: Robustness Analysis, students that do not move. Accounting exercise

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Restricted sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted gap</td>
<td>0.214</td>
<td>0.205</td>
</tr>
<tr>
<td>gap closed by family inputs</td>
<td>0.034</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>15.6%</td>
<td>16.8%</td>
</tr>
<tr>
<td>gap closed by school inputs</td>
<td>0.020</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>9.2%</td>
<td>10.6%</td>
</tr>
</tbody>
</table>

Table 8: Robustness Analysis, public school students. Accounting exercise

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Restricted sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>predicted gap</td>
<td>0.214</td>
<td>0.200</td>
</tr>
<tr>
<td>gap closed by family inputs</td>
<td>0.034</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>15.6%</td>
<td>16.2%</td>
</tr>
<tr>
<td>gap closed by school inputs</td>
<td>0.020</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>9.2%</td>
<td>7.6%</td>
</tr>
</tbody>
</table>
Appendix B: Derivation of the bias when savings are a poor proxy of ability to learn

From (12): \( \eta_{it} = (1/\beta)(s_{it} - \zeta_{it}) \), and plugging back in (11) we get:

\[
\begin{align*}
    h_{it} &= \gamma_0 + \gamma_1 I_{it} + \varepsilon_t + (1/\beta)(s_{it} - \zeta_{it}) \\
          &= \gamma_0 + \gamma_1 I_{it} + \frac{1}{\beta} s_{it} - \frac{1}{\beta} \zeta_{it} + \varepsilon_t \\
    \mu_{it} &= h_{it} - (\gamma_0 + \gamma_1 I_{it} + \gamma_2 s_{it})
\end{align*}
\]

which implies that the unobserved component equals:

\[
\mu_{it} = \varepsilon_{it} - \gamma_2 \zeta_{it} = h_{it} - (\gamma_0 + \gamma_1 I_{it} + \gamma_2 s_{it}) \tag{13}
\]

where \( \gamma_2 = 1/\beta \). This model can be rewritten in deviation of the means, which results:

\[
\begin{align*}
    \mu_{it} &= (h_{it} - \bar{h}) - [\gamma_1 (I_{it} - \bar{I}) + \gamma_2 (s_{it} - \bar{s})] \\
    &= h_{it} - [\gamma_1 i_{it} + \gamma_2 s_{it}]
\end{align*}
\]

which with some abuse of notation now lower case letters refer to deviation with respect to their mean. Thus the OLS estimates in this case can be derived from:

\[
\min \sum \mu_{it}^2 = \min \sum [h_{it} - (\gamma_1 i_{it} + \gamma_2 s_{it})]^2
\]

The FOC are:

\[
\begin{align*}
    \gamma_1 : \sum_i [h_{it} - (\bar{h}_1 i_{it} + \bar{h}_2 s_{it})] i_{it} &= 0 \tag{14} \\
    \gamma_2 : \sum_i [h_{it} - (\bar{h}_1 i_{it} + \bar{h}_2 s_{it})] s_{it} &= 0 \tag{15}
\end{align*}
\]

In matrix notation:

\[
x_t' h_t = x_t' x_t \gamma
\]

where:

\[
x_t' = \begin{pmatrix} i_{1t} & \ldots & i_{nt} \\ s_{1t} & \ldots & s_{nt} \end{pmatrix} = \begin{pmatrix} i_t \\ s_t \end{pmatrix}
\]

and

\[
h_t = \begin{pmatrix} h_{1,t} \\ \vdots \\ h_{n,t} \end{pmatrix}
\]

36
From the FOC (16), we get:

\[ \hat{\gamma} = (x_t^t x_t)^{-1} x_t^t h_t \]
\[ \hat{\gamma} = (x_t^t x_t)^{-1} x_t^t (x_t \gamma + \mu_t) \]
\[ \gamma - \gamma = (x_t^t x_t)^{-1} x_t^t \mu_t \]

Which implies that we can write the expected bias as:

\[ E(\hat{\gamma} - \gamma) = E\left[(x_t^t x_t)^{-1} x_t^t \mu_t \right] \]

\[ E(\hat{\gamma} - \gamma) = E\left[\left( \sum_A s_{it}^2 - \sum_A i_{it} s_{it} \right) \left( \sum_A i_{it} \mu_{it} \right) \right] \]

where \( A = (\sum s_{it}^2) (\sum i_{it}^2) - (\sum i_{it} s_{it})^2 \). We can re write the expected bias more compactly as:

\[ E(\hat{\gamma} - \gamma) = E\left[\left( \frac{s_{it}^2 x_t - i_{it} s_{it}}{A} \right) \left( \frac{i_{it} \mu_{it}}{s_{it}^2 x_t} \right) \right] \]

\[ E(\hat{\gamma} - \gamma) = E\left[\left( \frac{s_{it} x_t x_t - i_{it} x_t s_{it}}{A} \right) \left( \frac{i_{it} \mu_{it}}{s_{it} x_t x_t} \right) \right] \]

\[ E(\hat{\gamma} - \gamma) = \left( \frac{\sigma_{i\mu} - \sigma_{i\mu} \sigma_{i\gamma}}{\sigma_{i\gamma} \sigma_{i\gamma} - \sigma_{i\gamma}^2} \right) \]

(17)

Note that the OLS bias when we omit the unobserved component \( \eta_{it} \) is:

\[ E(\hat{\gamma}_{OLS}^1 - \gamma_1) = \frac{\sigma_{i\mu}}{\sigma_{ii}} = \frac{\sigma_{i\gamma} + \sigma_{i\epsilon}}{\sigma_{ii}} = \frac{\sigma_{i\gamma}}{\sigma_{ii}} \]

Using the suggested proxy variable and assuming that all the components of the error term are orthogonal to current inputs; i.e., \( \sigma_{i\mu} = 0 \), we can re write (17) as:

\[ \gamma_1 : E(\hat{\gamma}_1 - \gamma_1) = \frac{-\sigma_{i\mu} \sigma_{i\gamma}}{\sigma_{ii} \sigma_{i\gamma} - \sigma_{i\gamma}^2} \]

\[ \gamma_2 : E(\hat{\gamma}_2 - \gamma_2) = \frac{\sigma_{i\mu} \sigma_{i\gamma}}{\sigma_{ii} \sigma_{i\gamma} - \sigma_{i\gamma}^2} \]

In this particular example we can see that using savings as a proxy for ability provides estimates with a smaller bias than OLS, even when the relation between savings and ability is not determinist. To see this, consider the bias in \( \hat{\gamma}_1 \):

\[ \gamma_1 : E(\hat{\gamma}_1 - \gamma_1) = \frac{-\sigma_{i\mu} \sigma_{i\gamma}}{\sigma_{ii} \sigma_{i\gamma} - \sigma_{i\gamma}^2} \]
\[ \gamma_1 : E(\hat{\gamma}_1 - \gamma_1) = \frac{-\sigma_{s\mu} \sigma_{zz}^{\text{OLS}} \sigma_{ii}^{\text{OLS}}}{\sigma_{s\mu} \sigma_{ii}} \]

Consider first \( \sigma_{s\mu} \):
\[
\sigma_{s\mu} = E \left[ \sum (\beta \eta_{it} + \zeta_{it}) \mu \right] = E \left[ \sum (\beta \eta_{it} + \zeta_{it}) \left( \varepsilon_{it} - \frac{1}{\beta} \zeta_{it} \right) \right]
\]
\[
\sigma_{s\mu} = E \left[ \beta \sum \eta_{it} \varepsilon_{it} + \sum \zeta_{it} \varepsilon_{it} - \sum \eta_{it} \zeta_{it} - \frac{1}{\beta} \sum \zeta_{it} \zeta_{it} \right]
\]
\[
\sigma_{s\mu} = E \left[ -\frac{1}{\beta} \sum \zeta_{it}^2 \right] = -\frac{1}{\beta} \sigma_{\zeta}^2
\]

\[\frac{\sigma_{i\eta}}{\sigma_{ii}}\] can be expressed in terms of the bias of the OLS estimator when we do not use savings as a proxy for unobserved ability to learn. To see this more clearly:
\[
\frac{\sigma_{i\eta}}{\sigma_{ii}} = \frac{E \left[ \sum \iota (\beta \eta_{it} + \zeta_{it}) \right]}{\sigma_{ii}} = E \left[ \beta \sum \eta_{it} \right]
\]

since \( E(\eta) = 0 \), \( E(\zeta) = 0 \), and \( \sum I \zeta_{it} = 0 \) by assumption. Then
\[
\frac{\sigma_{i\eta}}{\sigma_{ii}} = \frac{\beta \sigma_{i\eta}}{\sigma_{ii}} = \beta E(\hat{\gamma}_1^{\text{OLS}} - \gamma_1)
\]

Replacing back in \( E(\hat{\gamma}_1 - \gamma_1) \):
\[
E(\hat{\gamma}_1 - \gamma_1) = \frac{-\frac{1}{\beta} \sigma_{\zeta}^2}{\sigma_{s\mu} \sigma_{ii}^{\text{OLS}} \sigma_{ii}} = \frac{\sigma_{i\eta}^2}{\sigma_{ii}} = \frac{\sigma_{\zeta}^2 E(\hat{\gamma}_1^{\text{OLS}} - \gamma_1)}{\sigma_{s\mu} \sigma_{ii}^{\text{OLS}} \sigma_{ii}}
\]
\[
= \frac{\sigma_{\zeta}^2}{\sigma_{s\mu} \sigma_{ii}^{\text{OLS}} \sigma_{ii}} E(\hat{\gamma}_1^{\text{OLS}} - \gamma_1)
\]

To show that \( E(\hat{\gamma}_1 - \gamma_1) < E(\hat{\gamma}_1^{\text{OLS}} - \gamma_1) \) we need to show \( \frac{\sigma_{\zeta}^2}{\sigma_{s\mu} \sigma_{ii}^{\text{OLS}} \sigma_{ii}} < 1 \), or \( \sigma_{\zeta}^2 < \sigma_{s\mu} \sigma_{ii}^{\text{OLS}} \sigma_{ii} \). The first term of \( \sigma_{s\mu} \sigma_{ii}^{\text{OLS}} \sigma_{ii} \) can be rewritten as:
\[
\sigma_{s\mu} = E \left[ \sum \eta_{it}^2 \right] = E \left[ \sum (\beta \eta_{it} + \zeta_{it})^2 \right]
\]
\[
= E \left[ \beta^2 \sum \eta_{it}^2 + 2 \beta \sum \eta_{it} \zeta_{it} + \sum \zeta_{it}^2 \right]
\]
\[
= \beta^2 \sigma_{\eta}^2 + \sigma_{\zeta}^2
\]
where we use the fact that \( E(\eta) = E(\zeta) = 0 \). We can decompose the second term as:

\[
\sigma^2_{si} = E \left[ \sum i (\beta \eta_{it} + \zeta_{it}) \right]^2 \\
= E \left[ \sum i \zeta_{it} + \sum i \beta \eta_{it} \right]^2 \\
= E \left[ \sum i \beta \eta_{it} \right] = \beta^2 E \left( \sum i \eta_{it} \right)^2
\]

This implies:

\[
\sigma_{s\eta \zeta} - \frac{\sigma^2_{is \zeta}}{\sigma_{ii}} = \beta^2 \sigma^2_\eta + \sigma^2_\zeta - \beta^2 \frac{E \left( \sum i \beta \eta_{it} \right)^2}{\sigma_{ii}}
\]

\[
= \beta^2 \left( \sigma^2_\eta - \frac{\sigma^2_{i \eta_{it}}}{\sigma_{ii}} \right) + \sigma^2_\zeta
\]

therefore, we can write the expected bias as:

\[
E(\tilde{\gamma}_1 - \gamma_1) = \frac{\sigma^2_\zeta}{\beta^2 \left( \sigma^2_\eta - \frac{\sigma^2_{i \eta_{it}}}{\sigma_{ii}} \right) + \sigma^2_\zeta} E(\tilde{\gamma}_1^{OLS} - \gamma_1)
\]

Then for \( \sigma^2_\zeta < \left[ \sigma_{s\eta \zeta} - \frac{\sigma^2_{i \eta_{it}}}{\sigma_{ii}} \right] \) to hold, we need \( \beta^2 \left( \sigma^2_\eta - \frac{\sigma^2_{i \eta_{it}}}{\sigma_{ii}} \right) > 0 \). Note that this last condition can be written in terms of the correlation coefficient between \( \eta \) and \( i \):

\[
\beta^2 \left( \sigma^2_\eta - \frac{\sigma^2_{i \eta_{it}}}{\sigma_{ii}} \right) = \beta^2 \sigma^2_\eta \left( 1 - \frac{\sigma^2_{i \eta_{it}}}{\sigma^2_\eta \sigma_{ii}} \right)
\]

\[
= \beta^2 \sigma^2_\eta \left( 1 - r^2_{i \eta_{it}} \right)
\]

where \( r^2_{i \eta_{it}} \) is the square of the correlation coefficient which is bounded by one. It follows that \( \beta^2 \left( \sigma^2_\eta - \frac{\sigma^2_{i \eta_{it}}}{\sigma_{ii}} \right) > 0 \) will always hold. Therefore, the bias of the contribution of inputs on the production of achievement when we use savings as a proxy for ability to learn will never exceed the bias if we ignore that proxy.
8 Appendix C: More Characteristics of NELS:88

NELS:88 is a nationally representative sample of eighth-graders that were first surveyed in the spring of 1988. The original sample employed a two-stage sampling design, selecting first a sample of schools and then a sample of students within these schools. In the first stage the sampling procedure set the probabilities of selection proportional to the estimated enrollment of eighth grade students. In the second stage 26 students were selected from each of those schools, 24 randomly and the other two were selected among hispanic and Asian Islander students, resulting in approximately 25,000 students. A sample of these respondents (18,221) were then resurveyed through four follow-ups in 1990, 1992, 1994, and 12,144 were interview again in 2000. Along with the student survey, NELS:88 included surveys of parents, teachers, and school administrators. By beginning with the 8th-grade, NELS:88 was able to capture the population of early dropouts—those who left school prior to spring term of 10th grade—as well as later dropouts (who left after spring of 10th grade). The study was designed not only to follow a cohort of students over time but also to “freshen” the sample at each of the first two follow-ups, and thus to follow multiple grade-defined cohorts over time. Thus, 10th grade and 12th grade cohorts were included in NELS:88 in the first follow-up (1990) and the second follow-up (1992), respectively. In late 1992 and early 1993, high school transcripts were collected for sample members, and, in the fall of 2000 and early 2001, postsecondary transcripts were collected, further increasing the analytic potential of the data.

Next the characteristics of each of the data collection years are summarize (See National Center for Education Statistics (2002) for a complete description):

**Base-Year Study.** The base-year survey for NELS:88 was carried out during the 1988 spring semester. The study employed a clustered, stratified national probability sample of 1,052 public and private 8th-grade schools. Almost 25,000 students across the United States participated in the base-year study. Questionnaires and cognitive tests were administered to each student in the NELS:88 base year. The student questionnaire covered school experiences, activities, attitudes, plans, selected background characteristics, and language proficiency. School principals completed a questionnaire about the school; two teachers of each student were asked to answer questions about the student, about themselves, and about their school; and one parent of each student was surveyed regarding family characteristics and student activities.
First Follow-up Study. Conducted in 1990, when most sample members were high school sophomores, the first follow-up included the same components as the base-year study, with the exception of the parent survey. The study frame included 19,363 in-school students, and 18,221 sample members responded. Importantly, the first follow-up study tracked base-year sample members who had dropped out of school, with 1,043 dropouts taking part in the study. Overall, the study included a total of 19,264 participating students and dropouts. In addition, 1,291 principals took part in the study, as did nearly 10,000 teachers.

Second Follow-up Study. The second follow-up took place early in 1992, when most sample members were in the second semester of their senior year. The study provided a culminating measurement of learning in the course of secondary school and also collected information that facilitated the investigation of the transition into the labor force and post-secondary education. The NELS:88 second follow-up resurveyed students who were identified as dropouts in 1990, and identified and surveyed additional students who had left school since the previous wave. For selected subsamples, data collection also included the sample member's parents, teachers, school administrators, and academic transcripts.

Third Follow-up Study (NELS:88/94). The NELS:88 third follow-up took place early in 1994. By this time in their educational careers, most of the sample members had already graduated from high school, and many had begun postsecondary education or entered the workforce. The study addressed issues of employment and postsecondary access and was designed to allow continuing trend comparisons with other NCES longitudinal studies. The sample for this follow up was created by dividing the second follow-up sample in 18 groups based on their response history, dropout status, eligibility status, school sector type, race, test score, socioeconomic status and freshened status. Each group was assigned an overall selection probability. Cases within a group were selected such that the overall probability was met, and the probability of selection within the group was proportional to each sample member second follow-up weight. The final sample size was 15,875 individuals.

Fourth Follow-up Study (NELS:88/2000). The fourth follow-up to NELS:88 (NELS:88/2000) included interviews with 12,144 members of the three NELS:88 sample cohorts 12 years after the base-year data collection (For costs reasons the third follow-up sample was subsample to limit the numbers of poor and difficult respondents and those who were unlikely to be located (those who couldn’t be located during earlier follow-up interviews). From here 15,649
individuals were selected and 12,144 of them completed the survey). Because these data represent the period 6 years after the last contact with the sample, they will enable researchers to explore a new set of educational and social issues about the NELS:88 respondents. For example, in 2000, most of the participants from the various cohorts of NELS:88 had been out of high school for 8 years and were 26 years old. At this age, the majority of students who intend to enroll in postsecondary schools will already have done so. Thus, a large proportion of students have completed college; some completed graduate programs. Many of these young people are successful in the market place, while others have had less smooth transitions into the labor force.