Estimating the Technology of Children's Skill Formation

Francesco Agostinelli
University of Pennsylvania

Matthew Wiswall *
University of Wisconsin-Madison
NBER

February 2020

JEL Classification C38, J13
Keywords: Child Development; Skill Formation; Measurement Error; Latent Factor Models; Dynamic Factor Analysis; Production Technology

*We thank the Editor, Jim Heckman, and four anonymous referees for their comments and feedback. We also thank Lance Lochner, Morteza Saharkhiz, Greg Veramendi, and participants at seminars and conferences at Arizona State, Barcelona GSE Summer Forum, Brown, CEHD (UChicago), Collegio Carlo Alberto, Duke, Interaction Workshop 2018 (Northwestern), Minnesota, Penn, Rice, University of Bristol, UCL, University of Houston, and Western Ontario for helpful comments. We are responsible for all errors.

Abstract

In this paper we study the process of children's skill formation. The identification of this process is challenging because children's skills are observed only through arbitrarily scaled and imperfect measures. Using a dynamic latent factor structure, we provide new identification results which illuminate the key identification trade-offs between restrictions on the skill production technology and the measurement relationships. One of our contributions is to develop empirically grounded restrictions on the measurement process that allow identification of more general production technologies, including those exhibiting Hicks neutral total factor productivity (TFP) dynamics and free returns to scale. We then use our identification results to develop a sequential estimation algorithm for the joint dynamic process of latent investment and skill development. Using data for the United States, we estimate different versions of the skill formation model under various identifying assumptions. Although all of our estimated models suggest that investments are particularly productive during early childhood. Moreover, we find that the marginal productivity of early investments is substantially higher for children with *lower* existing skills, suggesting the optimal targeting of interventions to disadvantaged children. When we compare these estimates to those using models which restrict the technology or ignore measurement error, we estimate policy effects which are substantially smaller, indicating that the generalities we allow are important to answering key policy questions.

1 Introduction

The wide dispersion of measured human capital in children and its strong correlation with later life outcomes has prompted a renewed interest in understanding the determinants of skill formation among children (for a recent review, see Heckman and Mosso, 2014). However, the empirical challenges in estimating the skill formation process, principally the technology of child development, is hampered by the likely imperfect measures of children's skills we have available. While measurement issues exist in many areas of empirical research, they may be particularly salient in research about child development. There exists a number of different measures of children's skills, and each measure can be arbitrarily located and scaled, and provide widely differing levels of informativeness about the underlying latent skills of the child.[1] In the presence of these measurement issues, identification of the underlying latent process of skill development is particularly challenging, but nonetheless essential, because ignoring the measurement issues through ad hoc simplifying assumptions could bias the empirical conclusions.

This paper makes two contributions. First, building on the results in (Cunha and Heckman, 2007; Cunha et al., 2010; Cunha and Heckman, 2008) we show identification of a general dynamic skill production function when the skills are unobserved and there exists only imperfect and arbitrarily scaled and located measures. Second, we develop a simple multiple step estimator for this skill development process and estimate this model using data on the child development process for the US.

Our identification results start by formulating the production function as a non-parametric model with mis-measured dependent and independent variables. We lean on recent advances in the econometrics literature, combining identification concepts for non-parametric regression and measurement error models, for identification of a general function combining skill production function primitives and measurement parameters. We then show that researchers face a key identification trade-off: separate point identification of the production function and measurement models can only be achieved with some combination of restrictions on the production function and measurement model. One simple way to understand this trade-off is that when observing that average test scores are increasing with child age, one cannot tell whether older children are learning more or the tests are getting easier. Rather than simply offering a "one-size-fits-all" generic assumption, we analyze various classes of assumptions that we show are sufficient for identification. We then evaluate the empirical relevance of these classes of restrictions, and provide guidance to researchers

[1]For a recent analysis of how measurement issues can be particularly salient, see Bond and Lang (see 2013, 2018) who analyze the black-white test score gap.

to evaluate whether their model and the measures available to them in their datasets satisfy these assumptions.

One of our key insights is to show that several prior studies already implicitly restrict the production technology, making further restrictions on the measurement system unnecessary. We introduce the concept of production technologies that have a *known* location and scale, technologies which are implicitly restricted so that the location and scale is already known. These known location and scale (KLS) technologies include the CES production technologies considered in a number of previous papers. [2] Starting with this class of technologies, we show that standard assumptions non-parametrically identify the production function, up to a normalization on the initial conditions only, without requiring further restrictions on the measurement process or latent variables in subsequent periods. In this way, we show that the identification concepts in this dynamic context differ substantially from simply applying techniques developed for cross-sectional latent factor models (Anderson and Rubin, 1956; Jöreskog and Goldberger, 1975; Goldberger, 1972; Chamberlain and Griliches, 1975; Chamberlain, 1977a,b; Carneiro et al., 2003; Hagglund, 1982) to each period or age of the development process, that is "re-normalizing" the model each period.

We also show that identification can be achieved with assumptions on the measurement system only, without KLS restrictions on the production technology. We introduce additional restrictions on the measurement process which are sufficient for identification of more general production technologies, including those exhibiting Hicks neutral total factor productivity (TFP) dynamics and free returns to scale, which much of the prior literature has not considered.[3] Using standard assumptions, these more general technologies cannot be identified because the location and scale of the technology cannot be separately identified from the location and scale of the measures. These more general aspects of the skill development formation process are nonetheless potentially important as imposing restrictions on the technology can reduce the permissible skill dynamics and investment productivities, substantially changing our inferences about the child development process and our evaluation of policies.

Our paper is the first to provide identification results for these more general production function models. We introduce the concept of "age-invariant" measures, measures that allow the comparison of skill development as children age. These assumptions are certainly not appropriate for all measures, but at least some skill measures are purposely designed by psychometricians and education experts for this

[2] see for example (Cunha and Heckman, 2007; Cunha et al., 2010; Cunha and Heckman, 2008)

[3] One recent exception to this is (Pavan, 2015), where the author estimates a linear non-KLS technology of skill formation.

purpose. Indicating the inherent trade-off in this type research, we show that if these types of measures are available, then more general skill technologies can be identified.

In the second part of our paper, we estimate a flexible parametric version of our model using data from the US National Longitudinal Survey of Youth (NLSY). We examine the development of cognitive skills in children from age 5 to age 14, and estimate a model of cognitive skill development allowing for complementarties between parental investment and children's skills; endogenous parental investment responding to the stock of children's skills, maternal skills, and family income; Hicks neutral dynamics in TFP; free returns to scale; and unobserved shocks to the investment process and skill production. Following Cunha et al. (2010), our empirical framework treats not only the child's cognitive skills as measured with error, but investment and maternal skills as well.

We develop an estimator for both of the two cases considered in the identification analysis, with restrictions on the production technology (assuming a known location and scale technology) or restrictions on the measurement process (assuming age-invariant measures), and present estimation results for both types of models. Constructively derived from our identification analysis, we form a method of moments estimator. Our estimator is not only relatively simple and tractable, but also robust to parametric distributional assumptions on the distribution of latent variables and measurement errors, as is commonly imposed in the past empirical literature. We jointly estimate the technology of skill formation, the process of parental investments in children, and the adult distribution of completed schooling and earnings, allowing the parameters of the production technology and investment process to freely vary as the child ages. Because we use measures in the NLSY dataset (PIAT scores) which are designed to account for developmental changes in children's skills, we argue that our estimates of the more general technology imposing age-invariance are preferred. Our estimates in this model of high TFP and increasing returns to scale at early ages indicate that investments are particularly productive early in the development period. We also find that the marginal productivity of early investments is substantially higher for children with *lower* existing skills, suggesting the optimal targeting of interventions to disadvantaged children. This conclusion is quite different from some existing estimates which find that the marginal productivity of investment is increasing in the existing stock children's skills.

Our estimates of the dynamic process of investment and skill development allow us to estimate the heterogeneous treatment effects of some simple policy interventions. We show that even a modest transfer of family income to families at age 5 would substantially increase children's skills and completed schooling, with the effects larger for low income families. When we compare these estimates to those using models which restrict the technology or ignore measurement error, we estimate policy effects which are substantially smaller, indicating that the generalities we allow are important quantitatively to answering key policy questions.

The paper is organized as follows. In the next section, we develop the model of skill development and measurement analyze the identification of this model. In the remainder of the paper we develop our estimator, we discuss our estimates and our counterfactual results, and we conclude.

2 Model and Identification

In this section, we lay out our simple stylized model of skill development and analyze the identification of the model.

2.1 Child Development Production Technology

Child development takes place over a discrete and finite period, $t = 0, 1, \ldots, T$, where $t = 0$ is the initial period (say birth) and $t = T$ is the final period of childhood (say age 18). There is a population of children and each child in the population is indexed i. For each period, each child is characterized by a skill stock $\theta_{i,t}$ and a flow investment $I_{i,t}$. In what follows, we consider only a single scalar skill and scalar investment, but the Appendix analyzes the multiple skill and multiple investment case. For each child, the current stock of skills and current flow of investments produce next period's stock of skill according to the skill formation production technology:

$$\theta_{i,t+1} = h_t(\theta_{i,t}, I_{i,t}, \eta_{i,t}) \text{ for } t = 0, 1, \ldots, T-1 \tag{1}$$

where $\eta_{i,t}$ is a production shock. Equation (1) can be viewed as a dynamic state space model with $\theta_{i,t}$ the state variable for each child i. The production technology $h_t(\cdot)$ is indexed with t to emphasize that the technology can vary over the child development period. According to this technology, the sequence of investments and shocks and the initial stock of child skills $\theta_{i,0}$ produce the sequence of skill stocks for each child i: $\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,T}$.

There are several features of the technology which have particular relevance both to understanding the process of child development and in evaluating policy interventions to improve children's skills. First, a key question is the productivity of investments at various child ages. At what ages are investments in children particularly productive in producing future skills ("critical periods") and, conversely, at what ages is it difficult to re-mediate deficits in skill? Second, how does heterogeneity in children's skills, at any given period, affect the productivity of new investments in children? Complementarity in the production technology between current skill stocks and investments implies heterogeneity in the productivity of investments across children. Third, how do investments in children persist over time and affect adult outcomes? Do early investments have a high return because they increase the productivity of later investments (dynamic complementarities) or do early investments "fade-out" over time? These features of the technology of skill development then directly inform the optimal *timing* of policy interventions – the optimal investment portfolio across early and late childhood – and the optimal *targeting* of policy – to which children should scarce resources be allocated to, with the goal of using childhood interventions to affect eventual adult outcomes.

2.2 Measurement Model

The focus of this paper is estimating the technology determining child skill development (1), while accommodating the reality that researchers have at hand various arbitrarily constructed and imperfect measures of children's skills. Our framework follows the approach in the literature (see for example Cunha and Heckman, 2007; Cunha et al., 2010; Cunha and Heckman, 2008), and it recognizes that children's skills are not directly measured by a single measure, but there exists multiple measures which we hypothesize can have some relationship to the unobserved latent skill stock θ_t.

Each measure m for child i skills in period (age) t is given by

$$Z_{i,t,m} = g_{t,m}(\theta_{i,t}, \epsilon_{i,t,m}), \tag{2}$$

For period t, we have $M_t \in \{1, 2, \ldots\}$ measures for latent skills $\ln \theta_{i,t}$: $m = 1, 2, \ldots, M_t$. $Z_{i,t,m}$ are the measures. $\epsilon_{i,t,m}$ are the individual measurement errors. To focus attention on the critical identification trade-offs considered next, throughout the identification analysis in this Section we assume investment $I_{i,t}$ is measured without error and is independent of both the production shock and the measurement errors for latent skills. In several important respects, we relax these assumptions in the empirical model we estimate in the next section.

2.3 Identification Trade-offs

In this sub-section we discuss the trade-offs for the identification of the dynamics of a child's latent skills. Our analysis of identification proceeds in the following way.

We first state some results on the identification of the initial period skill distribution and measurement parameters. In particular, we identify the distribution of latent skills and investments in the initial period, together with the associated measurement parameters. Our identification of the initial conditions follows standard arguments used in the current literature (e.g.: Cunha et al., 2010), but for completeness we fully specify this first step of the identification analysis. We then provide some negative results for the general identification of the dynamics of children's skills if no restrictions are imposed on the model. Finally, we provide new identification results under various forms of assumptions/restrictions. We conclude by discussing the relevance of each of the different assumptions for the various empirical situations researchers could face.

We specify our framework based on the general model in (1)-(2). We maintain a general non-parametric specification of the technology of skill formation (with additive separable shocks), while we consider a linear factor model for the skill measurements as in the empirical model of Cunha and Heckman (2007); Cunha et al. (2010).[4] The model is as follows:

$$\ln \theta_{i,t+1} = \ln f_t(\theta_{i,t}, I_{i,t}) + \eta_{i,t} \qquad \text{for } t = 0, 1, \ldots, T-1 \tag{3}$$

$$Z_{i,t,m} = \mu_{t,m} + \lambda_{t,m} \ln \theta_{i,t} + \epsilon_{i,t,m} \qquad \text{for } t = 0, 1, \ldots, T \text{ and } m = 1, \ldots, M_t. \tag{4}$$

where, without loss of generality, we assume that $E(\eta_{i,t}) = 0$ for all t and $E(\epsilon_{i,t,m}) = 0$ for all t and m. The measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$ represent the location and scale of the measures, respectively. For the remainder of the analysis we drop the i subscript.

The main identification trade-off comes from separately identifying the location and scale of the measurement model in equation (4) from the location and scale of the technology $f_t(\theta_{i,t}, I_{i,t})$ in equation (3). Put more intuitively, from some change in average scores between $t+1$ and t, $E(Z_{i,t+1,m}) - E(Z_{i,t,m})$, we cannot identify whether this change is due to a change in the measurements or an actual change in average latent skills:

$$E(Z_{i,t+!,m}) - E(Z_{i,t+1}) = (\mu_{t+1,m} - \mu_{t,m}) + \lambda_{t+1,m} E(\ln \theta_{i,t+1}) - \lambda_{t,m} E(\ln \theta_{i,t})$$

where changes in the measurements are represented by changes in the measurement

[4] In the empirical exercise we consider a parametric translog technology of skill formation.

intercepts μ or loadings λ, and for example $(\mu_{t+1,m} - \mu_{t,m}) > 0$ would imply that the measures have become "easier."

2.3.1 Identification of Initial Conditions

Latent skill stocks θ_t have no natural scale and location. A normalization is then required to fix the scale and location of the latent skill stocks to a particular measure. We normalize the latent skill stock to one of the measures of initial period skills:

Normalization 1 *Initial period normalizations*

(i) $E(\ln \theta_0) = 0$

(ii) $\lambda_{0,1} = 1$

This normalization fixes the location and scale of latent skills θ_0 to a particular measure, $Z_{0,1}$, where the labeling of the normalizing measure as measure $m = 1$ is arbitrary. For the normalizing measure, we then have the following:

$$Z_{0,1} = \mu_{0,1} + \ln \theta_0 + \epsilon_{0,1},$$

where $\mu_{0,1} = E(Z_{0,1})$ given the normalization $E(\ln \theta_0) = 0$.

The first set of assumptions restricts the measurement model for initial period skills:

Assumption 1 *Initial Period Measurement Assumptions:*
(i) $Cov\,(\epsilon_{0,m}, \epsilon_{0,m'}) = 0$ *for all* $m \neq m'$
(ii) $Cov\,(\epsilon_{0,m}, \ln \theta_0) = 0$ *for all* m*.*

Assumption 1 (i) is that measurement errors are uncorrelated contemporaneously across measures. Assumption 1 (ii) is that measurement errors are uncorrelated with the latent skill variable. Although these assumptions are strong in some sense, they are common in the current literature.

Under Normalization 1, Assumption 1, and with at least 3 measures in the first period, $M_0 \geq 3$, we identify the $\lambda_{0,2}, \lambda_{0,3}, \ldots, \lambda_{0,M_0}$ factor loadings from ratios of measurement covariances:

$$\lambda_{0,m} = \frac{Cov(Z_{0,m}, Z_{0,m'})}{Cov(Z_{0,1}, Z_{0,m'})}, \tag{5}$$

for $m \neq m'$, $m \neq 1$, $m' \neq 1$, where measure $m = 1$ is the normalizing measure.

Further, under the normalization that $E(\ln\theta_0) = 0$ (Normalization 1), we identify the $\mu_{0,1}, \mu_{0,2}, \ldots, \mu_{0,M_0}$ intercepts from

$$\mu_{0,m} = E(Z_{0,m})\text{for all } m. \tag{6}$$

We then construct the following "residual" skill measures from the original raw measures:

$$\widetilde{Z}_{0,m} = \frac{Z_{0,m} - \mu_{0,m}}{\lambda_{0,m}} \equiv \ln\theta_0 + \widetilde{\epsilon}_{0,m}, \tag{7}$$

where $\widetilde{\epsilon}_{0,m} = \frac{\epsilon_{0,m}}{\lambda_{0,m}}$, a scaled version of the original measurement error. The measures $\widetilde{Z}_{0,m}$ can be thought as "error-contaminated" measurements of the latent skills.

2.3.2 Identification of the Technology (Period 1)

We now express a measure of the latent skill in period 1 $Z_{1,m}$ as a function of both the measurement parameters and the technology by substituting equation (3) into equation (4):

$$\begin{aligned} Z_{1,m} &= \mu_{1,m} + \lambda_{1,m} \ln f_0(\theta_0, I_0) + (\eta_0 + \lambda_{1,m}\epsilon_{1,m}) \\ &= q_0(\theta_0, I_0) + u_{0,m} \end{aligned} \tag{8}$$

where the combined residual $u_{0,m} = \eta_0 + \lambda_{1,m}\epsilon_{1,m}$ is mean-zero. The period 0 child's skill (θ_0) on the RHS is unobserved, but we have some "error-contaminated" measurements $(\widetilde{Z}_{0,m})$ derived from the previous identification step (see equation 7): $\widetilde{Z}_{0,m} = \ln\theta_0 + \widetilde{\epsilon}_{0,m}$.

Equation (8) can be thought as a non-parametric regression equation relating an observed measure of period 1 skills to a non-parametric function of unobserved period 0 skills and investment. The new error term in this equation $u_{0,m}$ has two parts: the production shock η_0 and the measurement error $\epsilon_{1,m}$. There are two identification challenges here: (i) the unobservability of the RHS skills (θ_0); and (ii) the potential endogeneity of these skills: the error term $u_{0,m}$ being correlated with $\ln\theta_0$. As noted in a recent paper by Adusumilli and Otsu (2018), estimating the model given in (8) relies on two long-standing and largely parallel econometric research programs on non-parametric IV models and models with errors-in-variables (e.g. Hausman et al., 1991; Schennach, 2004). Sufficient conditions for identification of q_0 are given in Adusumilli and Otsu (2018), and they rely on the existence of a relevant instrumental variable W_0 which satisfies two key conditions: (i) $E(u_{0,m}|W_0) = 0$ and (ii) $\widetilde{\epsilon}_{0,m} \perp$

W_0. In addition, a third requirement is (iii) $\widetilde{\epsilon}_{0,m} \perp \theta_0$, a stronger assumption than the no correlation condition Assumption 1.[5] Below we discuss possible candidate instruments for our child skill formation application.[6]

Given the non-parametric identification of the function q_0, the next step is unpacking the components of this function to provide identification for the dynamics of the next period skills. Without loss of generality, we write the first term of the production technology in (3) as

$$\ln f_0(\theta_0, I_0) = \ln A_0 + \psi_0 \ln K_0(\theta_0, I_0), \tag{9}$$

where $\ln A_0$ and ψ_0 are the location and scale of the technology, and $K_0(\theta_0, I_0)$ is a Known Location and Scale (KLS) function, which we define as follows:

Definition 1 *A production function $K_t(\theta_t, I_t)$ has* Known Location and Scale (KLS) *if for two non-zero input vectors (θ'_t, I'_t) and (θ''_t, I''_t), where the input vectors are distinct, the outputs $K_t(\theta'_t, I'_t)$ and $K_t(\theta''_t, I''_t)$ are both known (do not depend on unknown parameters), finite, and non-zero.*

A production technology with known location and scale implies that, for a change in inputs from (θ'_t, I'_t) to (θ''_t, I''_t), the change in output $K_t(\theta'_t, I'_t) - K_t(\theta''_t, I''_t)$ is known. Other points in the production possibilities set may be unknown, i.e. they depend on free parameters to be estimated. Writing the technology as in (9), we have intuitively separated out two parameters representing location and scale from the general function f_0, the parameters $\ln A_0$ and ψ_0. A leading example of a KLS function is the CES function:

$$\theta_{t+1} = (\gamma_t \theta_t^{\sigma_t} + (1 - \gamma_t) I_t^{\sigma_t})^{1/\sigma_t}$$

[5] Note that in our framework we have mis-measured dependent and independent variables, but using the measurement model, we can re-arrange the estimating equation so that there is only measurement error in the RHS variables.

[6] One candidate instrument is an alternative measure $(Z_{0,m'})$ of initial skills (θ_0), and we discuss conditions under which this is a valid instrument below. In addition, there are many other potential instruments which researchers could use (e.g. Cunha et al. (2010) assume some exclusion restrictions that makes lagged household income a valid instrument). See Section 2.5 for an example of identification and estimation using a parametric form for q_0. Note that we previously assumed, and maintain throughout this Section, that investment is observed and independent of the production shock and measurement errors, hence $E(u_{0,m}|I_0) = 0$ and $\widetilde{\epsilon}_{0,m} \perp I_0$. We relax this assumption in the empirical model we take to data.

with $\gamma_t \in (0,1)$ and $\sigma_t \in (-\infty, 1]$. In this case it is easy to show that, for all pairs (θ_t, I_t) such that $\theta_t = I_t$, the output is known: $\theta_{t+1} = \theta_t = I_t$.[7]

Next, substituting equation (9) into the main equation (8), we have

$$Z_{1,m} = (\mu_{1,m} + \ln A_0 \lambda_{1,m}) + (\lambda_{1,m}\psi_0) \ln K_0(\theta_0, I_0) + u_{0,m} \tag{10}$$

At this point, we cannot separately identify the period 1 measurement parameters $(\mu_{1,m}, \lambda_{1,m})$ from production function parameters $(\ln A_0, \psi_0)$. That is, we cannot separately identify the location and scale of the measurement function from the location and scale of the production function. We consider identification under one of two prototypical restrictions:

Assumption 2 *Measurement Function Restriction* $\mu_{t,m} = \mu_{0,m}$ *and* $\lambda_{t,m} = \lambda_{0,m}$ *for all* $t > 0$ *and this* m

Assumption 3 *Production Function Restriction* $\ln A_t = 0$ *and* $\psi_t = 1$ *for all* t

Under *either* set of restrictions, we identify all of the parameters of interest. Let (θ_0', I_0') and (θ_0'', I_0'') be the two input vectors for which K_0 is, by definition, known and non-zero. Then we have a system of two equations:

$$q_0(\theta_0', I_0') = (\mu_{1,m} + \ln A_0 \lambda_{1,m}) + (\lambda_{1,m}\psi_0) \ln K_0(\theta_0', I_0')$$

$$q_0(\theta_0'', I_0'') = (\mu_{1,m} + \ln A_0 \lambda_{1,m}) + (\lambda_{1,m}\psi_0) \ln K_0(\theta_0'', I_0'')$$

where the q_0 function (defined in equation 8) is identified given the arguments above. The two parameters $\beta_0 = (\mu_{1,m} + \ln A_0 \lambda_{1,m})$ and $\beta_1 = (\lambda_{1,m}\psi_0)$ are just-identified. Under the measurement function restriction Assumption 2, we identify $\ln A_0$ and ψ_0, in addition to the other parameters. Or, under the production function restriction Assumption 3, we identify $\mu_{1,m}$ and $\lambda_{1,m}$, in addition to the other parameters.

Importantly, one does not need to assume *both* a production function and measurement function restriction. Either assumption is sufficient for identification, and

[7] This result follows from the constant return to scale property of the CES. Suppose $\theta_t = I_t = a$, we then have:

$$(\gamma_t \theta_t^{\sigma_t} + (1-\gamma_t) I_t^{\sigma_t})^{1/\sigma_t} = (\gamma_t a^{\sigma_t} + (1-\gamma_t) a^{\sigma_t})^{1/\sigma_t} = (a^{\sigma_t})^{1/\sigma_t} = a\,.$$

imposing both is over-identifying. In fact, imposing one, allows testing of the other. There are other types of restrictions that would be sufficient for identification, but these two broad classes of restrictions help clarify the range of options.[8]

2.4 Sequential Identification

We can continue the identification for remaining periods sequentially. As shown above, using either Assumption 2 or Assumption 3, we recover the measurement parameters for period 1, $(\mu_{1,m}, \lambda_{1,m})$. Just as we did for period 0, this allows us to form an error-contaminated measure for period 1:

$$\widetilde{Z}_{1,m} = \frac{Z_{1,m} - \mu_{1,m}}{\lambda_{1,m}} \equiv \ln \theta_1 + \widetilde{\epsilon}_{1,m},$$

Note that here, unlike with the period 0 initial period, we do not need multiple measures to identify the measurement parameters: these parameters are identified from the restrictions imposed on the previous period t=0 relationships. We continue as above, and using the analysis above, the existence of a sequence of valid instruments $\{W_0, W_1, \ldots, W_{T-1}\}$ and Assumption 2 or Assumption 3, we identify the sequence of production technologies $f_t(\theta_t, I_t)$ for $t = 0, 1, \ldots, T-1$ and the measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$ for $t = 1, \ldots, T$.

2.5 Example: Cobb-Douglas Production Function

In this section we provide one example that highlights the trade-offs for the identification of the technology of skill formation. Because the production function we consider here is a simple parametric one, we do not need to rely on the non-parametric results from Adusumilli and Otsu (2018). We consider a Cobb-Douglas production function as an example:

$$\ln \theta_1 = \ln A_0 + \psi_0(\gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0) + \eta_0 \tag{11}$$

with $\gamma_0 \in (0, 1)$. Note that the function $K_0(\theta_0, I_0) = \gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0$ is KLS (Definition 1).

[8] Restrictions on latent variables directly can be understood as either restrictions on the technology or the measurement system. For example, assuming $E(\ln \theta_t) = 0$ for all t (not just for the initial period $t = 0$) implies that skill formation process is mean (log) stationary. If, for example, in addition one assumed the KLS CES production technology, this restriction (implying $E(\ln \theta_t) = E(\ln \theta_{t-1})$ for all $t > 0$) restricts the technology to the Cobb-Douglas special case ($\sigma_t \to 0$). See Agostinelli and Wiswall (2016).

Using the initial normalization and Assumption 1, we identify the set of measurement parameters $\{\mu_{0,m}, \lambda_{0,m}\}_m$ for the initial period ($t = 0$). We then proceed as above, and by using the measurement system, we define the "error-contaminated" measures for the initial period and the next period measures as:

$$\begin{aligned}\widetilde{Z}_{0,m} &= \ln \theta_0 + \widetilde{\epsilon}_{0,m} \\ Z_{1,m} &= \mu_{1,m} + \lambda_{1,m} \ln \theta_1 + \epsilon_{1,m}\end{aligned}$$

As in all of our analysis above, the measurement parameters $\mu_{1,m}$ and $\lambda_{1,m}$ are treated as free parameters. Substituting the production technology into the period 1 measurement equation, we have

$$\begin{aligned}Z_{1,m} &= (\ln A_0 + \lambda_{1,m}\mu_{1,m}) + \lambda_{1,m}\psi_0(\gamma_0 \ln \theta_0 + (1-\gamma_0)\ln I_0) + u_{0,m} \\ &= (\ln A_0 + \lambda_{1,m}\mu_{1,m}) + \lambda_{1,m}\psi_0(\gamma_0(\widetilde{Z}_{0,m} - \widetilde{\epsilon}_{0,m}) + (1-\gamma_0)\ln I_0) + u_{0,m} \\ &= \beta_{0,0} + \beta_{0,1}\widetilde{Z}_{0,m} + \beta_{0,2}\ln I_0 + \pi_{0,m} \qquad (12)\end{aligned}$$

where equation (12) is a reduced-form equation of the original technology, with $\pi_{0,m} = u_{0,m} - \lambda_{1,m}\psi_0\gamma_0\widetilde{\epsilon}_{0,m}$. Each reduced-form parameter maps into a combination of the original technological parameters and the measurement parameters as follows:

$$\beta_{0,0} = \ln A_0 + \lambda_{1,m}\mu_{1,m},$$

$$\beta_{0,1} = \lambda_{1,m}\psi_0\gamma_0,$$

$$\beta_{0,2} = \lambda_{1,m}\psi_0(1-\gamma_0)$$

As we discuss before, we now face an identification trade-off. In particular, the reduced-form parameters $\beta_{0,0}, \beta_{0,1}, \beta_{0,2}$ are combinations of unknown measurement parameters $\mu_{1,m}, \lambda_{1,m}$ and unknown production function parameters $\ln A_0, \psi_0, \gamma_0$.

Identification takes two steps. First, the identification of the reduced-form parameters (βs) faces the standard error-in-variables problem. In this case, the OLS estimand does not identify $\beta_{0,0}, \beta_{0,1}, \beta_{0,2}$. We can solve this problem using standard IV techniques with the existence of a valid and relevant instrument W_0.[9]

Second, even with $\beta_{0,0}, \beta_{0,1}, \beta_{0,2}$ identified, we are still faced with an under-identification problem as there are 5 unknown primitive parameters $\ln A_0, \psi_0, \gamma_0, \mu_{0,m}$,

[9] Here a valid instrument satisfies $E(\pi_{0,m}|W_0) = 0$. Given that the combined residual $\pi_{0,m}$ is an additively separable function of the random variables $\epsilon_{0,m}$, $\epsilon_{1,m}$, and $\eta_{0,m}$, the instrument is valid if the instrument is mean-independent of each of these separably.

$\lambda_{1,m}$, and there are only 3 identified reduced form parameters in (12). As previously discussed, some restriction is needed to achieve point identification. Sufficient conditions come from either imposing assumptions on the measurement function (Assumption 2) or assumptions on the production function (Assumption 3).

Under the measurement function restriction (Assumption 2), we have an exactly identified system given by:

$$\beta_{0,0} = \ln A_0 + \lambda_{0,m}\mu_{0,m}\,,$$

$$\beta_{0,1} = \lambda_{0,m}\psi_0\gamma_0\,,$$

$$\beta_{0,2} = \lambda_{0,m}\psi_0(1-\gamma_0)\,,$$

where the identified parameters are:

$$\psi_0 = \frac{\beta_{0,1}+\beta_{0,2}}{\lambda_{0,m}}\,,$$

$$\gamma_0 = \frac{\beta_{0,1}}{\lambda_{0,m}\psi_0}\,,$$

$$\ln A_0 = \beta_{0,0} - \lambda_{0,m}\mu_{0,m}\,.$$

Under the production function restriction (Assumption 3), we have a different exactly identified system:

$$\beta_{0,0} = \lambda_{1,m}\mu_{1,m},$$

$$\beta_{0,1} = \lambda_{1,m}\gamma_0,$$

$$\beta_{0,2} = \lambda_{1,m}(1-\gamma_0)$$

where the solution is

$$\lambda_{1,m} = \beta_{0,1} + \beta_{0,2},$$

$$\gamma_0 = \frac{\beta_{0,1}}{\lambda_{1,m}},$$

$$\mu_{1,m} = \frac{\beta_{0,0}}{\lambda_{1,m}}.$$

The production function restriction here can also be understood as a restriction in a traditional "reduced form" error-in-variables model (Chamberlain, 1977a). In this literature, identification is often achieved by proportionality restriction (e.g. linear regression parameters are assumed proportional to each other), that is restrictions imposed on the β parameters directly. In our case, the restrictions we consider come from restrictions on the primitive production function, which is intuitively appealing because we can understand the consequences of these restrictions on the primitive production relationships.

Moving to the next period, given the period 1 measurement parameters $\mu_{1,m}$ and $\lambda_{1,m}$ are identified, we construct the period 1 contaminated measure:

$$\widetilde{Z}_{1,m} = \frac{Z_{1,m} - \mu_{1,m}}{\lambda_{1,m}} \equiv \ln \theta_1 + \widetilde{\epsilon}_{1,m}.$$

We can then continue to identify the production technology for period 2, which written in reduced form is given by

$$Z_{2,m} = \beta_{1,0} + \beta_{1,1} \widetilde{Z}_{1,m} + \beta_{1,2} \ln I_1 + \pi_{2,m}.$$

As for the period 1 function, identification requires a relevant and valid instrument, and an assumption on the measurement or production function. We continue in this way to period $t = T$.

2.6 Age-Invariant Measures

We conclude this section with a discussion of measures which would satisfy these auxiliary assumptions. Unless the researcher has strong reasons to impose restrictions on the production function, it would be desirable to leave the production function general and look for measures which would satisfy the measurement restriction, allowing identification of the most general skill process. An extensive literature, principally in psychometrics and education, is concerned with designing skill measures that can be "equated" across children of different ages so that the development of children can be tracked using a single measure. Outside economics these are often referred to as "vertical scales." These types of measures primarily consist of tests that are designed to be applicable for children of various ages, and include a range of test items which show meaningful variation for both younger and older children. An example is a vocabulary test in which children are asked to define words of increasing difficulty. To the extent that one could administer this same test to children in a range of ages, the raw count of words defined correctly could be considered an age-invariant measure of latent skill in this domain.

In our framework, we formalize this idea by defining *age-invariant* measures:

Definition 2 *A pair of measures Z_t and Z_{t+1} is* age-invariant *if $E(Z_{t,m}|\theta_t = p) = E(Z_{t+1,m}|\theta_{t+1} = p)$ for all $p \in \mathbb{R}_{++}$.*

Age-invariant measures imply that two children of different ages t and $t+1$ would nonetheless have the same expected level of measured skill *if* the children have the same latent level of skill: $\theta_t = \theta_{t+1} = p$.[10] In this case, the younger child, aged t, could be considered "ahead" of her age group, and the older child, aged $t+1$, could be considered "behind" her age group. The age invariant measures $Z_{t,m}$ and $Z_{t+1,m}$ would report the same score (in expectation) for these two children. Definition 2 implies that age-invariant measures would satisfy Assumption 2, allowing identification of general technologies.[11]

Whether a given pair of measures is age-invariant depends on the measures and data available, and must be evaluated on a case-by-case basis. Using pairs of unrelated measures, such as counts of body parts a toddler can identify to measure skills at age 1 and SAT scores to measures skills at age 18, would not seem to constitute a pair of age-invariant measures as there is no reason to believe these measures would have a common location and scale. Other measures may be age-invariant, such as certain test score measures developed specifically to track development as children age.[12] Examples of these types of measures for the cognitive skill domain include the Peabody Individual Achievement Test (PIAT) and the Woodcock-Johnson tests, both of which are specifically designed to track child development over a wide range of child ages. These measures have have been used in a numerous studies, both in

[10]Age-invariant measures should not be confused with "age-standardized" measures, which are measures the researcher constructs to be mean 0 and standard deviation 1 at all ages for the particular sample at hand. The concept of age-invariant measures concerns the underlying unobserved primitive parameters of the measurement equations. Age-standardized measures would in fact not represent any growth in average skills or changes in the dispersion of skills as children age.

[11]Age-invariance implies the following restrictions on measurement parameters: $\mu_{t+1} + \lambda_{t+1} \ln p = \mu_t + \lambda_t \ln p$ for all p. Re-arranging, we have $(\mu_{t+1} - \mu_t) = \ln p \ (\lambda_t - \lambda_{t+1})$ for all p. This is the case if and only if $\mu_t = \mu_{t+1}$ and $\lambda_t = \lambda_{t+1}$. For our theoretical and empirical results, the necessary condition is that a particular measure is age-invariant at least for two consecutive periods.

[12]In practice, these types of age-invariant tests are often administered such that the questions are endogenously determined by the previous answers of the child. Therefore, while not all children are in fact answering the exact same test questions, their scores are determined in an age comparable way. The typical test includes a number of test items ranging from low difficulty to high difficulty questions. Testing begins by first establishing a baseline test item for each child. While the baseline is initially based on the child's age, the baseline adjusts downward (to less difficult questions) as the child is unable to answer questions correctly. Once the baseline is established, the test then progressively asks more difficult questions. Testing stops when the child makes a certain number of mistakes. The score is then determined as the number of correct answers before testing stops. Included in this number of correct answers are the lower difficulty test items prior to the baseline item because it is assumed the child would have answered these items correctly (given she was able to answer more the difficult items).

and outside economics. In our empirical application, we use the PIAT measures and discuss in more detail why we believe these measures are age-invariant.[13]

Our concept of age-invariance should not be confused with "age-standardized" or "age-equivalent" measures. The latter are two transformations of the raw data: age-standardized measures are constructed to be mean 0 and standard deviation 1 at each age, and age-equivalent measures are constructed to express raw scores relative to the typical development pattern using mean or median scores by age.[14] Neither transformation of the data guarantees that the resulting transformed measures are age-invariant. Age-invariant measures cannot be automatically constructed using ex post data transformations because the age-invariance property concerns the relationship between data and unobserved latent skills. It is possible however that certain transformations may result in measures which are age invariant whereas the raw measures are not, and vice versa. A leading example is age-standardized measures. If the raw measures Z are age-invariant and average latent skills are increasing by age, then necessarily the age-standardized measures $\tilde{Z} = (Z - \bar{Z})/\sigma$ are not age-invariant. To see this, notice that the age-invariance condition for age-standardized measures states that the following difference is 0 (at all p):

$$E(\tilde{Z}_{t+1}|\theta_{t+1} = p) - E(\tilde{Z}_t|\theta_t = p) = \frac{E(Z_{t+1}|\theta_{t+1} = p) - \bar{Z}_{t+1}}{\sigma_{t+1}} - \frac{E(Z_t|\theta_t = p) - \bar{Z}_t}{\sigma_t}.$$

While the age-invariant condition on the raw measures guarantees that $E(Z_{t+1}|\theta_{t+1} = p) = E(Z_t|\theta_t = p) = 0$, the rest of the right-hand side of the above equation can differ from zero because latent skills are increasing with age ($\bar{Z}_{t+1} > \bar{Z}_t$). Therefore the age-standardizing transformation of the age-invariant raw scores produces measures which are no longer age-invariant.[15]

[13] Several recent papers explicitly invoke age-invariance assumptions regarding their particular measures. Attanasio et al. (forthcoming) argue that the raw count of number of tasks completed from the Bayley cognitive scale is age-invariant. Using the same data, Attanasio et al. (2019a) argue that a transformed age-equivalent version of the Bayley is age-invariant. Attanasio et al. (2019b) use the Peabody Picture Vocabulary Test (PPVT) as an age-invariant measure. These studies also assume age-invariance for certain non-cognitive and health measures.

[14] Consider a simple example. The average score for 8 year olds on some test is 37, and 43 for 9 year olds. The standard deviation of raw scores for both ages is 10. If a 8 year old child has a raw score of 37, then this child's age-equivalent score is 8, and age-standardized score is (37 - 37)/10 = 0. On the other hand, if a 9 year old child has a raw score of 37, then this child's age-equivalent score is 8, and age-standardized score is (37 - 43)/10 = -0.6. The age-equivalent score can be alternatively calculated by using the regression of the child's age at the time of the test on the raw score of the test. The predicted outcome from this regression represents the age-equivalent converted test score.

[15] Note also that our concept of age-invariance is unrelated to the concept of "anchoring" (Cunha

3 Estimation

In this section we discuss the empirical model we take to the data, the estimation algorithm we develop based on the identification analysis of the preceding section, and briefly describe the data. Additional details about the data and sample are left for the Appendix.

3.1 Empirical Model

There are five parts to the empirical model: 1) a model of skill development where skills in the next period are produced by the stocks of existing skills and parental investments; 2) a model of parental investment where investment depends on household characteristics and the existing stock of skills; 3) a distribution of initial conditions of household characteristics and child skills; 4) a model of the relationship between final childhood skills and adult outcomes (schooling and earnings); and 5) a measurement model relating each of the latent model elements to observed data measures. Besides specifying particular functional forms for the production technology, the major distinction between the empirical model and the preceding identification analysis is that we assume parental investment is also measured with error and allow parental investment to be endogenously related to the stock of existing children's skill.

The timing of the model is as follows. There are five biannual periods of child development: ages 5-6 ($t = 0$), 7-8 ($t = 1$), 9-10 ($t = 2$), 11-12 ($t = 3$), 13-14 ($t = 4$). While it would be ideal to extend the model to even earlier ages (to birth or even to pre-natal periods), we face the common trade-off of assuming "too much" relative to the data we have available. We have chosen here to focus on the childhood period from age 5 to 14 where we have more skill measures, and plausibly age-invariant measures, and can judge the performance of the model and estimator in closer to ideal conditions.

3.1.1 Skill Production Technology

At each age t the current level of latent cognitive skills and investment produce the next period's $(t + 1)$ skills. The technology takes a stochastic translog form:

$$\ln \theta_{t+1} = \ln A_t + \gamma_{1,t} \ln \theta_t + \gamma_{2,t} \ln I_t + \gamma_{3,t} \ln I_t \cdot \ln \theta_t + \eta_{\theta,t}, \tag{13}$$

et al., 2010). Anchoring in that paper is a transformation of the latent variables in terms of adult outcomes (e.g. earnings), to allow an interpretation of primitive production function parameters in terms of these outcomes. Age-invariance and the related measurement parameter assumptions concern the relationship between skill measure data and latent variables.

where $\ln A_t$ is the TFP term, and $\eta_{\theta,t}$ is the stochastic production shock, which is assumed i.i.d. $\sim N(0, \sigma^2_{\theta,t})$ for all t, and is assumed independent of the current stock of skills and investment. The translog specification is a generalization of the Cobb-Douglas specification, where the special case $\gamma_{3,t} = 0$ is the typical Cobb-Douglas specification (with the addition of a TFP term and a stochastic shock). We use the translog specification because of its flexibility relative to the Cobb-Douglas and other CES functions. The translog function allows a non-constant elasticity of substitution between inputs, and it can be expanded with the inclusion of additional terms to provide a close approximation of any unknown production technology. The log-linear form of the function is also convenient, because it allows us to derive the estimator in closed-form expression, as detailed below. Our general translog function also allows free returns to scale. And, with $\gamma_3 \neq 0$, the elasticity of skill production with respect to investment depends on the current level of children's skills:

$$\frac{\partial \ln \theta_{t+1}}{\partial \ln I_t} = \gamma_{2,t} + \gamma_{3,t} \ln \theta_t,$$

where $\gamma_{3,t} > 0$ implies a higher return to investment for children with currently high levels of skill than for children with low levels of skill, a dynamic complementarity where past skills (and past investments which produced those skills) increase the productivity of current investments. A negative value for $\gamma_{3,t}$ implies a higher return on investment for the lower skill children, and would generally imply a higher return to targeting investments to low skill children.

3.1.2 Parental Investment

We specify a parametric policy function for parental investment. The parametric policy function depends upon the state variables, such as the current stock of the child's skills, mother's skills, and family income:

$$\ln I_t = \alpha_{1,t} \ln \theta_t + \alpha_{2,t} \ln \theta_{MC} + \alpha_{3,t} \ln \theta_{MN} + \alpha_{4,t} \ln Y_t + \eta_{I,t} \tag{14}$$

where $\sum_j \alpha_{j,t} = 1$ for all t, θ_{MC} is the mother's stock of cognitive skills, θ_{MN} is the mother's stock of non-cognitive skills, Y_t is household income, and $\eta_{I,t}$ is the investment shock, where $\eta_{I,t}$ i.i.d. $\sim N(0, \sigma^2_{I,t})$ for all t, and is assumed independent of latent skills and income. Our concept of investment represents both quantity and quality aspects, where we use measures of investments which capture quantity aspects of investment (time parents spent reading to children) and quality aspects (whether children are "praised" by their parents).

This specification of investment is an approximation of the parental behavior which is not derived from an explicit economic model of the household behavior. This approach follows Cunha et al. (2010); Attanasio et al. (2015a,b). The advantages of this approach are twofold. First, it provides a simple and tractable model of the investment process, which avoids the computational burden of solving and estimating a formal model of household behavior. Second, this approach has the potential to allow for some generality as our specification of the investment process can be consistent with multiple models of the households. Other recent work derives parental endogenous behavior from explicit models of the household, including explicit representations of household preferences, decision making, beliefs, and constraints (see for example Bernal, 2008; Del Boca et al., 2014a,b; Cunha, 2013; Cunha et al., 2013; Doepke and Zilibotti, 2017; Agostinelli, 2019; Doepke et al., 2019). The advantage of these latter approaches is that the counterfactual policy analysis incorporates well defined household responses to policy, see Del Boca et al. (2014b) for some discussion.

Given the investment function does not derive from an explicit model, we interpret the parameters in a more "reduced-form" way. The parameter $\alpha_{1,t}$ can be interpreted as reflecting whether parents "reinforce" existing skill stocks ($\alpha_{1,t} > 0$) or "compensate" for low skill stocks ($\alpha_{1,t} < 0$). The parameters $\alpha_{2,t}$ and $\alpha_{3,t}$ reflect the extent to which the mother's skills relate to the quantity and quality of her parental investment as in the case where more skilled mothers read to their children more or provide higher quality interactions. Finally, the parameter $\alpha_{4,t}$ reflects the influences that household resources have on the extent of parental investments, and it includes the combined effects of constraints the household faces (such as credit market constraints), as well as the household's preferences for investing scarce resources in children (see Caucutt et al., 2015).

Finally, to close the investment model, we assume that log family income ($\ln Y_t$) follows an AR(1) process which allows for life-cycle trends in income:

$$\ln Y_{t+1} = \mu_Y + \delta_Y \cdot t + \rho_Y \ln Y_t + \eta_{Y,t} \tag{15}$$

where the innovation is $\eta_{Y,t}$ i.i.d. $\sim N(0, \sigma_Y^2)$ and is assumed independent of all latent variables. Initial family income Y_0 is allowed to be correlated with mother's and children's initial skills, and hence our model captures important correlations between household resources and the skills of parents and children.

3.1.3 Initial Conditions

The initial conditions consist of the child's initial (at age 5-6) stock of skills $\theta_{C,0}$, the mother's cognitive and non-cognitive skills (θ_{MC} and θ_{MN}), which are assumed to be time invariant over the child development period, and the level of family income at birth (Y_0).[16] Define the vector of initial conditions as

$$\Omega = (\ln\theta_0, \ln\theta_{MC}, \ln\theta_{MN}, \ln Y_0)$$

We assume a parametric distribution for the initial conditions:

$$\Omega \sim N(\mu_\Omega, \Sigma_\Omega)$$

where $\mu_\Omega = [0, 0, 0, 0, \mu_{0,\ln Y}]$. $\mu_{0,\ln Y}$ is the mean of log household income when children are 5-6 years old. The means of the remaining variables are set to zero as a normalization. Σ_Ω is the variance-covariance matrix for the initial conditions.

3.1.4 Adult Outcome

In order to provide a more meaningful metric to evaluate policy interventions in our model, we relate adult outcomes to the stock of children's skills in the final period of the child development process (period $T = 4$ or age 13-14). Each adult outcome Q is determined by

$$Q = \mu_Q + \alpha_Q \ln\theta_T + \eta_Q, \tag{16}$$

where η_Q is independent of $\ln\theta_T$. We use years of schooling measured at age 23 and log earnings at age 29 as adult outcomes. Schooling is an attractive adult outcome to use because it explains a large fraction of adult earnings and consumptions, is largely determined at an early point in adulthood and, unlike realized labor market earnings, does not suffer from a censoring issue due to endogenous labor supply.

3.1.5 Measurement

The final piece of our model is the model of measurement relating latent variables to observed data. Children's skills, parental investment, and mother's skills are all assumed to be measured with error. There are 4 latent variables: $\omega \in \{\theta, \theta_{MC}, \theta_{MN}, I\}$.

[16]We assume mother's cognitive and non-cognitive skills to be time invariant only because of data limitation. During the first waves of interviews, different measures were collected as part of the original NLSY79 dataset, such as the Armed Services Vocational Aptitude Battery (ASVAB), the Rotter-Locus of Control Scale, as well as the Rosenberg Self-Esteem index.

There are in general multiple measures for each latent variable. As in the preceding Section, each measure is assumed to take the following form:

$$Z_{\omega,t,m} = \mu_{\omega,t,m} + \lambda_{\omega,t,m} \ln \omega_t + \epsilon_{\omega,t,m}$$

where m indexes the measures for each latent variable.

We assume a generalized version of Assumption 1 appropriate for this more general empirical model. All measurement errors are assumed independent of each other (across measures and over time), and all measurement errors are assumed independent of the latent variables, household income, and the "structural" shocks $(\eta_{I,t}, \eta_{\theta,t}, \eta_Q)$. These assumptions are strong, and weaker assumptions, for example of mean-independence, are sufficient for identification of the parametric model. On the other hand, we make no other restrictions on the distribution of measurement error (e.g. we do not assume $\epsilon_{\omega,t,m}$ is distributed Normal), as is common in previous approaches in the literature. Our sequential estimator, described below, is therefore robust to misspecification of the marginal distributions of measurement errors.

3.2 Estimation Algorithm

Our estimation algorithm is formed from the identification results presented above. Before describing the steps of the algorithm, consider several estimation options. One approach – a "brute force" approach – is to implement a simulation-based estimator, where we would simulate the full sequence of latent variables and measures from some guessed primitive parameters based on explicit assumptions about the distribution of measurement errors (e.g. assume they are Normally distributed). The simulated empirical distribution will be used to compute a likelihood function or a set of moments to form the basis of an estimator. We do not prefer this approach because it requires additional assumptions about the distribution of measurement errors which are not required for identification. Moreover, this approach may also involve a tremendous amount of computationally costly simulation given the non-linear nature of the model.

A second alternative estimation approach is to use the measures directly to simulate the distribution of model's variables by assuming a particular distribution for the latent variables. One then could estimate the production function in a second step from the simulated distribution of latent variables. This approach used in Cunha et al. (2010) and Attanasio et al. (2015a,b) assumes that the latent variables are distributed according to a mixture of two Normal distributions. Below, we describe our estimation method, which is robust to misspecification of the assumed parametric distributions for latent skills and measurement noises.

Our estimation approach directly follows our identification approach in treating the measurement parameters as nuisance parameters which can be computed sequentially along with the primitive parameters of the model generating the latent variables. Following the estimation of the initial conditions using standard techniques, we sequentially estimate for each age the investment and production functions, followed by the measurement parameters for the measures used for that age. The sequential algorithm we develop has the advantage of tractability because our estimator does not require the simulation of the full model; the primitives of the production technology and investment functions can be estimated directly from data. In addition, another advantage of our approach over a joint estimation approach is by breaking the estimator into steps, we make the identification assumptions as transparent as possible. Of course, the disadvantage of our approach is a potential loss of efficiency from not estimating the parameters jointly and exploiting "cross-step" restrictions.[17]

We present the estimation algorithm for two different models based on either Assumption 2 or Assumption 3.

Model 1 (Measurement Function Restrictions: Age-Invariance): Assume that Assumption 2 holds, we let $\ln A_t$ be free and $\sum_{j=1}^{3} \gamma_{j,t}$ be free.

Model 2 (Production Function Restrictions: Known Location and Scale): Assume that Assumption 3 holds, we fix $\ln A_t = 0$ for all t (no TFP dynamics) and $\sum_{j=1}^{3} \gamma_{j,t} = 1$ for all t (restricted returns to scale).

For exposition, we start by presenting the estimation algorithm for the second version of the model, using the restricted technology. The estimator for the more general technology (Model 1) is described second.

3.2.1 Estimation using Production Function Restrictions (Model 2)

Step 0 (Estimate Initial Conditions and Initial Measurement Parameter)

First, we estimate the measurement parameters at the initial period (age 5-6), $\lambda_{\omega,0,m}, \mu_{\omega,0,m}$ for all measures m, for both children's and mother's skills. To estimate these measurement parameters, we use ratios of covariances and measurement

[17] It should be noted that to compute counterfactual simulations, we do simulate the full model forward from the estimated initial conditions, using the estimated model primitives. Simulating counterfactuals does not require assuming anything about the marginal distribution of measurement errors or latent variables.

means as outlined above (5) and (6). We choose one measure for children's cognitive skills, mother's cognitive skills, and mother's non-cognitive skills as the normalizing measure (which we label $m = 1$, without loss of generality) and normalize the factor loading for this measure to be 1: $\lambda_{\theta,0,1} = 1$, $\lambda_{MC,0,1} = 1$, $\lambda_{MN,0,1} = 1$.[18] We estimate the remaining factor loadings using the average of the covariances between all of the remaining measures, where each factor loading is computed from

$$\lambda_{\omega,0,m} = \frac{Cov(Z_{\omega,0,m}, Z_{\omega,0,m'})}{Cov(Z_{0,\omega,1}, Z_{\omega,0,m'})} \; \forall m \neq m' \text{ and } \forall \, \omega \in \{\theta, MC, MN\}.$$

Given the normalization that log skills are mean 0 in the initial period, we compute the initial measurement intercepts as

$$\mu_{\omega,0,m} = E(Z_{\omega,0,m}) \; \forall m \text{ and } \forall \, \omega \in \{\theta, MC, MN\}$$

With the factor loading estimates in hand, we then estimate the initial period variance-covariance matrix Σ_Ω using variances and covariances in measures of skills and family income (assumed measured without error). This step provides estimates of the initial joint distribution of children's skills, mother's skills, and family income. In this initial step, we also estimate the parameters of the income process (15) using a regression of log family on lagged log family income and an age trend.

Finally, given the estimates of the measurement parameters for children and mother skills, we form the following "residual" measures:

$$\widetilde{Z}_{\omega,0,m} = \frac{Z_{\omega,0,m} - \mu_{\omega,0,m}}{\lambda_{\omega,0,m}} \; \forall m \text{ and } \forall \, \omega \in \{\theta, MC, MN\}$$

We are now ready to estimate the investment function for period $t = 0$, where the investment in this first period depends on the initial child's skills and household characteristics (mother's skills and family income).

Step 1 (Estimate Investment Function Parameters):

Following the errors-in-variables formulation described above, substitute a "raw" measure for investment $Z_{I,0,m}$ and a "residual" measure for each of the latent skills ($\widetilde{Z}_{\theta,0,m}$, $\widetilde{Z}_{MC,0,m}$, $\widetilde{Z}_{MN,0,m}$) into the model of investment defined in terms of primitives (14):

[18] Note that while investment is a latent variable as well, we do not need to normalize the scale and location of latent investment because investment already has a scale and location specified by the KLS investment equation (14).

$$\frac{Z_{I,0,m} - \mu_{I,0,m} - \epsilon_{I,0,m}}{\lambda_{I,0,m}} = \alpha_{1,0}(\widetilde{Z}_{\theta,0,m} - \widetilde{\epsilon}_{\theta,0,m}) + \alpha_{2,0}(\widetilde{Z}_{MC,m} - \widetilde{\epsilon}_{MC,m})$$

$$+\alpha_{3,0}(\widetilde{Z}_{MN,m} - \widetilde{\epsilon}_{MN,m}) + \alpha_{4,0} \ln Y_0 + \eta_0$$

Re-arranging, we have

$$\begin{aligned} Z_{I,0,m} &= \mu_{I,0,m} + \lambda_{I,0,m}\alpha_{1,0}\widetilde{Z}_{\theta,0,m} + \lambda_{I,0,m}\alpha_{2,0}\widetilde{Z}_{MC,m} + \lambda_{I,0,m}\alpha_{3,0}\widetilde{Z}_{MN,m} + \lambda_{I,0,m}\alpha_{4,0}\ln Y_0 \\ &+ \epsilon_{I,0,m} + \lambda_{I,0,m}(\eta_0 - \widetilde{\epsilon}_{\theta,0,m} - \widetilde{\epsilon}_{MC,m} - \widetilde{\epsilon}_{MN,m}) \\ &= \beta_{0,0,m} + \beta_{1,0,m}\widetilde{Z}_{\theta,0,m} + \beta_{2,0,m}\widetilde{Z}_{MC,m} + \beta_{3,0,m}\widetilde{Z}_{MN,m} + \beta_{4,0,m}\ln Y_0 + \pi_{0,m} \qquad (17) \end{aligned}$$

where $\beta_{j,0,m} = \lambda_{I,0,m}\alpha_{j,0}$ for all j and

$$\pi_{0,m} = \epsilon_{I,0,m} + \lambda_{I,0,m}(\eta_0 - \alpha_{1,0}\widetilde{\epsilon}_{\theta,0,m} - \alpha_{2,0}\widetilde{\epsilon}_{MC,m} - \alpha_{3,0}\widetilde{\epsilon}_{MN,m}).$$

Estimation of (17) by OLS would yield inconsistent estimates of the $\beta_{j,0,m}$ coefficients because the measures are correlated with their measurement errors (included in the residual term $\pi_{0,m}$). Here the structure of the model affords the researcher several possible strategies to consistently estimate the $\beta_{j,0,m}$ coefficients. We use an instrumental variable estimator with the vector of excluded instruments composed of alternative measures of skills: $[Z_{\theta,0,m'}\ ,\ Z_{MC,0,m'}\ ,\ Z_{NC,0,m'}]$. Under our previous assumptions about the measurement error, these instruments are valid because the alternative measures are uncorrelated with all of the components of $\pi_{0,m}$. Using this IV strategy, we obtain consistent estimators for the $\beta_{j,t,m}$ coefficients. The primitive parameters of the investment function are then recovered from

$$\alpha_{j,0} = \frac{\beta_{j,0,m}}{\sum_{j=1}^{4}\beta_{j,0,m}} \quad \forall j \in \{1, \ldots, 4\}$$

Step 2 (Compute Measurement Parameters for Latent Investment):

After estimating the primitive parameters of the investment function, we recover the scale and location for the investment equation without further re-normalizations on the measurement equation parameters. The intercept and factor loading for the investment measure are given by

$$\mu_{0,m} = \beta_{0,0,m}\,,$$

and

$$\lambda_{0,m} = \sum_{j=1}^{4} \beta_{j,0,m}\,.$$

With these consistent estimates for the measurement parameters for investment, we form the "residual" measures for investment in period $t = 0$:

$$\widetilde{Z}_{I,0,m} = \frac{Z_{I,0,m} - \mu_{I,0,m}}{\lambda_{I,0,m}} \equiv \ln I_0 + \widetilde{\epsilon}_{I,0,m}\,.$$

Step 3 (Estimate Skill Production Technology)

Next, we use a similar technique to estimate the production technology. Substituting the residual measures into the production technology (13), we have

$$\frac{Z_{\theta,1,m} - \mu_{\theta,1,m} - \epsilon_{\theta,1,m}}{\lambda_{\theta,1,m}} = \gamma_{1,0}(\widetilde{Z}_{\theta,0,m} - \widetilde{\epsilon}_{\theta,0,m}) + \gamma_{2,0}(\widetilde{Z}_{I,0,m} - \widetilde{\epsilon}_{I,0,m})$$
$$+ \gamma_{3,0}(\widetilde{Z}_{\theta,0,m} - \widetilde{\epsilon}_{\theta,0,m})(\widetilde{Z}_{I,0,m} - \widetilde{\epsilon}_{I,0,m}) + \eta_{\theta,0}\,.$$

With some algebra, we can re-write this as:

$$Z_{\theta,1,m} = \delta_{0,0,m} + \delta_{1,0,m}\widetilde{Z}_{\theta,0,m} + \delta_{2,0,m}\widetilde{Z}_{I,0,m} + \delta_{3,0,m}\widetilde{Z}_{\theta,0,m} \cdot \widetilde{Z}_{I,0,m} + \pi_{\theta,0,m}\,, \qquad (18)$$

where the new error term $\pi_{\theta,0,m}$ is:

$$\pi_{\theta,0,m} = \epsilon_{\theta,1,m} + \lambda_{\theta,1,m}[\eta_{\theta,0} - \gamma_{1,0}\epsilon_{\theta,0,m} - \gamma_{2,0}\epsilon_{I,0,m} - \gamma_{3,0}(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m} + \widetilde{Z}_{I,0,m}\epsilon_{\theta,0,m} - \epsilon_{\theta,0,m}\epsilon_{I,0,m})]\,,$$

while the rest of the reduced-form parameters (δs) are: $\delta_{0,0,m} = \mu_{\theta,0,m}$, $\delta_{j,0,m} = \lambda_{\theta,1,m}\gamma_{j,0}$ for any $j \in \{1, 2, 3\}$.

As with the investment function, estimation of 18 using OLS would lead to inconsistent estimates. We use the same IV approach as above using instruments formed from alternative measures $[Z_{\theta,0,m'}\,,\, Z_{I,0,m'}\,,\, Z_{\theta,0,m'} \cdot Z_{I,0,m'}]$. Under our measurement assumptions these instruments are uncorrelated the residual error term $\pi_{\theta,0,m}$.[19] With consistent estimates of δs in hand, we can then recover the structural parameters and for the production technology as:

$$\gamma_{j,0} = \frac{\delta_{j,0,m}}{\sum_{j=1}^{3} \delta_{j,0,m}} \ \forall \ j \in \{1,2,3\} \ .$$

Step 4 (Compute Measurement Parameters for Latent Skill):

The measurement parameters for the latent skill measure in period $t = 1$ ($Z_{\theta,1,m}$) can then be recovered from

$$\mu_{\theta,1,m} = \delta_{0,0,m},$$

$$\lambda_{\theta,1,m} = \sum_{j=1}^{3} \delta_{j,0,0}.$$

We then form the residual measure for latent skill as

$$\widetilde{Z}_{\theta,1,m} = \frac{Z_{\theta,1,m} - \mu_{\theta,1,m}}{\lambda_{\theta,1,m}} \ .$$

Step 5 (Estimate variance of Investment and Production Function Shocks):

The variances of both the investment shocks ($\sigma^2_{I,0}$) and of the production function shocks ($\sigma^2_{\theta,0}$) remain to be estimated. In order to estimate $\sigma_{I,0}$, we use the covariance between the residual term ($\pi_{0,m}$) in (17), and an alternative residual measure of investment $\widetilde{Z}_{I,0,m'} = \ln I_0 + \widetilde{\epsilon}_{I,0,m'}$ as follows:

$$Cov(\pi_{0,m}/\lambda_{I,0,m}, \widetilde{Z}_{I,0,m'}) = V(\eta_{I,0}) = \sigma^2_{I,0} \ .$$

To compute the residual measure $\widetilde{Z}_{I,0,m}$, we need to compute the measurement parameters for this measure. We do this by following the procedure explained in

[19] Perhaps the less obvious terms are terms such as this $E(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'})$. Under the assumption of independence of the errors, we have

$$E(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'}) = E(\widetilde{Z}_{\theta,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'})E(\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'})$$

given $\epsilon_{I,0,m}$ is independent of $\widetilde{Z}_{\theta,0,m}$. Given the independence assumption, the latter term is $E(\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'}) = E(\epsilon_{I,0,m}) = 0$. Therefore, $E(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'}) = 0$.

the Steps 1 and 2 above, with the alternative measure $Z_{I,0,m'}$. The variance of the production shock is estimated in the same way using an alternative measure of children's skills in period $t = 1$:

$$Cov(\pi_{\theta,1,m}/\lambda_{1,m}, \widetilde{Z}_{\theta,1,m'}) = V(\eta_{\theta,0}) = \sigma^2_{\theta,0} \, .$$

Remaining Steps

We repeat Steps 1-5 for the remaining periods until the final period of child development T. This algorithm produces estimates of the parameters for all child ages.

3.2.2 Estimation using Measurement Function Restrictions (Model 1)

The preceding algorithm restricted the production technology to have no TFP dynamics and restricted the returns to scale (Model 2). Following our previous results, identification of the more general technology can instead be accomplished with restrictions on the measurement parameters. We assume we have available at least one child skill measure which is age-invariant. Label the age-invariant measure to be measure m, and for this measure we have $\mu_{\theta,t,m} = \mu_{\theta,0,m}$ for all t and $\lambda_{\theta,t,m} = \lambda_{\theta,0,m}$ for all t.

With this age invariant measure, we repeat Step 3 (Estimate Production Technology). The "reduced-form" equation (18) and estimation of the $\delta_{j,0,m}$ parameters remains the same. To allow for free returns to scale we do not restrict the structural $\gamma_{j,0}$ parameters to sum to 1. The structural parameters are computed as

$$\gamma_{j,0} = \frac{\delta_{j,0,m}}{\lambda_{\theta,1,m}} \; \forall \; j \in \{1, 2, 3\} \; ,$$

where $\lambda_{\theta,1,m}$ is now known.

With the inclusion of the TFP term $\ln A_0$, the $\delta_{0,0,m}$ intercept from the reduced-form equation (18) is now

$$\delta_{0,0,m} = \mu_{\theta,1,m} + \lambda_{\theta,1,m} \ln A_0 \, .$$

Given the age-invariance assumption, both measurement parameters $\mu_{\theta,1,m}$ and $\lambda_{\theta,1,m}$ are known, and we can estimate $\ln A_0$.

With the addition of these computations to Step 3, the other steps in the algorithm remain the same. We can use this extended algorithm to compute the full sequence of parameters for the investment and production functions for all child ages.

3.2.3 Estimating the Adult Outcome Equation

Finally, after we have computed the full path of primitive parameters for the investment and production functions, we are able to estimate the adult outcome process (16). We focus on both final years of education at age 23 and log earnings at age 29. We use the same IV method as before to solve the measurement error issue. Substituting the measures for skills at age 13-14 ($t = 4$) in equation (16), we have:

$$Q = \mu_Q + \alpha_Q \widetilde{Z}_{\theta,4,m} + (\eta_Q - \alpha_Q \widetilde{\epsilon}_{\theta,4,m}) \tag{19}$$

We use a second measure for skills at age 13-14 as an IV to identify α_Q.

3.3 Data

We estimate the model using information about children and their families obtained from the National Longitudinal Study of Youth 1979 (NLSY). Descriptive statistics for the sample and additional data construction details are left for the Appendix.

The NLSY dataset is constructed by matching female respondents of the original dataset with their children who were part of the Children and Young Adults surveys, from 1986 to 2012. The dataset provides observations of the first period of the model (age 5-6) through adulthood. The total number of children in our sample is 11,509.

The NLSY dataset contains multiple measures of children's skills, mother's skills, and parental investments. The complete set of measures, their ranges and descriptive statistics for our sample are included in the Appendix. For children's skills we rely on different sub-scales of the Peabody Individual Achievement Test (PIAT) in Mathematics, Reading and Recognition, and the Peabody Picture Vocabulary Test (PPVT). Finally, we use information for children when they become young adults to link the children skills into a more meaningful metric to evaluate policy intervention: we use children's highest grade completed at age 23 or older and their earnings at age 29. The information about the educational attainment is measured as the highest grade completed as of date of last interview. We considered schooling information only for those young adults who were at least 23 years old or older in the last 2012 interview. Age 29 earnings is in real 2012 dollars.

For mother's cognitive skills we use sub-scales of the Armed Services Vocational Aptitude Battery (ASVAB), and for mother's non-cognitive skills we use the Rotter and Rosenberg indexes. For parental investments, we use the various HOME score measures from direct observation and interview with the mother. Family income includes all sources of income for the parents, including mother's and father's labor income, and any sources of non-labor income.

4 Results

In this section we discuss our parameter estimates, simulate the estimated model to describe the development of children's skills, and compute the effects of simple interventions to improve skills and adult outcomes. We begin by presenting estimates of Model 1 in which we allow for TFP dynamics and free returns to scale. For this case, we assume that child skill measures are age-invariant. Given the structure of the longitudinal aspect of PIAT tests, which administer the same test to children of various ages (given their ability level), we believe it is appropriate to assume the measurement intercepts and factor loadings for these measures of cognitive skills are age-invariant (Definition 2).

Throughout this section, we also discuss results using the alternative Model 2 (restricted production function), and we also show the model's estimates when we do not correct for measurement error. We briefly discuss the policy predictions of these models below, but, for brevity, we report estimates of these several alternative models in the Appendix.

Finally, because the parameter estimates of the production technology and investment equations are relative to the initial skill normalizations, the magnitudes of many of the parameters estimates are not directly interpretable in isolation. We conclude this section with a series of policy counterfactual experiments using the estimated model. These exercises provide necessary metrics to interpret the estimates with respect to adult outcomes, schooling and earnings.

4.1 Parameter Estimates

4.1.1 Initial Conditions

Table 2 reports estimates of the initial conditions variance-covariance matrix Σ_Ω and the associated correlation matrix. We normalize children's cognitive skills to the PIAT-Mathematics test, mother's cognitive skills to the ASVAB2 (Arithmetics reasoning) and mother's non-cognitive skills to the Self-Esteem 1 (Rosenberg Self-Esteem: "I am a person of worth") measure. The variances and covariances of the latent skills, and the investment and production function parameters, are interpreted relative to these normalizations. As expected, we estimate that children's skills, mother's cognitive and non-cognitive skills, and family income are all highly positively correlated. For space considerations, estimates of the dynamic family income process can be found in the Appendix.

4.1.2 Investment Function

Table 3 reports the estimates of the investment function specified in Section 3.1.2. At ages 5-6, we find that investment is increasing in children's skills, mother's skills, and family income. Because of the log-log form of the investment equation, we can interpret parameter estimates as elasticities. The parameter estimate of 0.230 on the log children's skills variable indicates that a 1 percent increase in children's skills raises investment by 0.23 percent, an inelastic response. The positive coefficient suggests that parents are "reinforcing" existing skills with further investments: children with higher skills are receiving even more investment than children with lower skills. Mother's cognitive skills and non-cognitive skills also increase investment at ages 5-6, with non-cognitive skills of the mother estimated to have a substantially higher elasticity than cognitive skills. These coefficients indicate that mothers with higher skills are providing higher quantities and qualities of investments in children. Turning to the importance of income to parental investments, we find that a 1 percent increase in family income raises investment by 0.34 percent. The response of investment with respect to mother's skills and family income reflects the combination of parental preferences and household constraints, which we cannot unfortunately separately distinguish using this reduced-form model of investment. Given that positive correlation between mother's initial skills, child's initial skills, and household income, taken together, these estimates of the investment function indicate that endogenous investment increases inequality in children's skills. The estimated variance on the investment shock reveals how much of the remaining variation in parental investments remains unexplained by this model, such as investments from schools, peers, and the child herself.

Comparing parameter estimates of the investment function over the development period reveals that the influence of the child's prior skills on investments becomes much smaller at later ages, indicating that parental investments become less reinforcing of existing skill stocks at older ages. As the child develops, we find that mother's non-cognitive skills becomes the dominant influence on investment. However, while the importance of family income falls somewhat from an elasticity of 0.34 at age 5-6 to 0.275 at age 11-12, income is still a significant and positive factor for parental investment even at later ages.

4.1.3 Production Function

Table 4 reports the parameter estimates for the technology of skill formation, as described in Section 3.1.1. We present measurement error corrected estimates of the two versions of the model: Model 1 (restricted measurement function) and Model

2 (restricted production function). Our preferred estimates are the measurement error corrected estimates of Model 1, which imposes the age-invariance restriction but allows for TFP dynamics and free returns to scale. We argue this age-invariance restriction is appropriate for the particular skill measures we use in the NLSY dataset. Unless otherwise specified, we present results using measurement error corrected estimates from Model 1.

At all ages, we find that skills are "self-productive" (next period's skills are increasing in existing skill stocks) and that skills are positively increasing in investment. For age 5-6 skill production, we estimate a significant negative complementarity between the stock of a child's skills and parental investments (the interaction term $\ln \theta_t \ln I_t$). This result highlights the importance of departing from the Cobb-Douglas/CES specifications.

The elasticities of skill production with respect to investment are heterogeneous, and we graph the skill elasticity for the age 5-6 production function in Figure 1 with respect to the existing stock of children's skill. The estimated negative coefficient on the interaction term indicates that the elasticity of skill production with respect to investment is decreasing in the child's current skill level. For low skill children, the elasticity approaches 1.4, indicating a that 1 percent increase in investment increases next period's skills by 1.4 percent. For already high skill children, the elasticity approaches 0.2, indicating that a 1 percent increase in investment raises future skills by only 0.2 percent.

The heterogeneous investment elasticities suggest that targeting interventions to improve children's skills would have the largest effect on skill disadvantage children. This result stands in contrast to the estimates from previous works in the literature, which were based on CES (or linear) technologies. These technology specifications restrict the heterogeneity of the investment productivity by assuming that the marginal productivity of investment must be *increasing* (or constant) with respect to a child's skills. Note also that unlike the CES (constant returns to scale) case, our unrestricted model allows investment elasticities to be larger than 1. Indeed, our estimates suggest that, at least for some children, skill production is relatively highly elastic with respect to investment.

The high TFP estimate for age 5-6 and the increasing returns to scale indicate that existing skills and investments at this initial age are very productive relative to later ages. These estimates of high returns to early investment will underlie the policy experiment results we discuss next. As children age, Table 4 indicates that skills and investment become generally less productive and skills less "malleable." We graph the estimated TFP at each age in Figure 2. Our estimate of TFP at age 11-12 falls to 1/6 the level at age 5-6, indicating a dramatic slowdown in the productivity of existing skills and investments in producing new skills. This feature of the technology is largely consistent with the evidence that cognitive skills are difficult to change as children after age 10.

Comparing these estimates for Model 1, which imposes age-invariance in the measures but leaves the production technology free, to the restricted production function estimates of Model 2 in Table 4 reveals that we clearly reject the restricted technology of Model 2. The estimated sum of the input coefficients far exceeds 1, with a value of 2.66 at age 5-6 that declines to 1.3 at older ages. In addition, the estimate of high positive TFP term also indicates that we clearly reject the assumption of a 0 log TFP imposed in Model 2. As discussed below, these differences in production function estimates imply very different investment and policy effects, with Model 2 estimates implying a much smaller effect of an income transfer on children's skill development than in Model 1 with an unrestricted technology.[20]

4.1.4 Adult Outcomes

Table 5 presents our estimates of the completed schooling outcome equation and log earnings equation. We estimate that a change of 1 standard deviation (with respect to log-skills at age 5-6) in children skills at age 13-14 leads to an increase of 0.15 years of school and an approximately 2.1 percent change in earnings at age 29. Below, we use these estimates to "anchor" our policy estimates to a meaningful adult outcome metric.

4.2 Estimated Child Development Path

We analyze the quantitative implications of the estimated model by simulating the dynamic model. Simulation of the model proceeds by drawing 100,000 children from the estimated initial conditions distribution and, for each child, forward simulating the path of income, investments, children's skills, and adult outcomes.

Figure 3 shows the estimated development path of mean log latent cognitive skills. Figures 4 and 5 show the dynamics in the distribution of latent skills. And, Figure 6 provides the estimated dynamics in the distribution of latent investment.

Perhaps not surprisingly, we find that children's mean latent skills grow substantially over this development period, from age 5 to 14, with the most rapid growth at early ages and growth slowing somewhat in the later period. As discussed above, key to the identification of the non-stationary change in children's skills is the use of

[20]In the Appendix, we report estimates of the technology using alternative sample definitions and adding additional covariates. The results are qualitatively similar.

age-invariant measures of children's skills. In addition to growth in mean skills, we estimate that the latent distribution of cognitive skills becomes more dispersed as children age (Figure 4). Inequality rises substantially as there are different rates of skill growth for children at different percentiles of the initial skill distribution. Figure 5 shows that skills for high skill children at the 90^{th} percentile grow 20% from age 5-6 to age 9-10 and grow 9% during the rest of the childhood. For low initial skill children at the 5^{th} percentile, growth is slower, with a 6 % growth rate from age 5-6 to age 7-8 and a 3 % growth rate from age 11-12 to age 13-14.

4.3 Policy Experiments

In this section, we explore implications of the estimated model by using the estimated model to predict the effect of income transfers on childhood skill development and adult outcomes. Although we do not have a fully micro-funded model of household choices, we argue that the experiments do at the very least provide a meaningful metric to understand the magnitude of the parameter estimates, and allow us to meaningfully compare the importance of various model features, such as restrictions on the skill production technology and measurement error correction.

4.3.1 Short and Long-Term Effects

Before we analyze the quantitative results for our particular counterfactuals, we first present a brief discussion of the effects of income transfers in our model. To allow for the possibility that an income transfer could have heterogeneous effects across households, we examine policy effects conditional on a vector of current state variables $\Omega_t = [\theta_t, \theta_{MC}, \theta_{MN}, Y_t]$, which includes the child's initial skills, the mother's skills, and initial family income. First, consider the expected *short-term* marginal effect of an increase in household income Y_t on the log of childhood skills in period $t+1$:

$$\begin{aligned}\Delta_{t+1,t}(\Omega_t) &= \frac{\partial \ln \theta_{t+1}}{\partial Y_t} \\ &= \frac{\partial \ln I_t}{\partial Y_t} \frac{\partial \ln \theta_{t+1}}{\partial \ln I_t},\end{aligned}$$

$\Delta_{t+1,t}(\Omega_t)$ is the product of the marginal change in parental investment and the marginal change in skill production. With our parametrization, this is given by

$$\Delta_{t+1,t}(\Omega_t) = \frac{\alpha_{4,t}}{Y_t}(\gamma_{2,t} + \gamma_{3,t} \ln \theta_t).$$

This short-term effect is heterogeneous by the level of family income and the existing stock of the child's skills. The marginal increase in investment is decreasing in the current level of income, as would be expected given the log form of the investment equation. The key parameter for the heterogeneity of the short-term effect is $\gamma_{3,t}$, with $\gamma_{3,t} > 0$ implying a higher return to investment for children with higher existing stocks of skills.

The dynamic model of skill development we estimate also allows us to consider the *long-term* effect of an income transfer at age t on outcomes beyond the immediate next period. The expected long-term effect of a marginal increase in income at period t on children's skills in period $t+2$ is given by

$$\begin{aligned}\Delta_{t+2,t}(\Omega_t) &= \frac{\partial \ln \theta_{t+2}}{\partial Y_t} \\ &= \Delta_{t+1,t}(\Omega_t)\frac{\partial \ln \theta_{t+2}}{\partial \ln \theta_{t+1}}(1 + \frac{\partial \ln I_{t+1}}{\partial \ln \theta_{t+1}})\end{aligned}$$

Note that we are analyzing the long-term effect of a one-time change in income at period t; income remains at baseline levels for all subsequent periods. With our parametrization, the long-term effect becomes

$$\Delta_{t+2,t}(\Omega_t) = \frac{\alpha_{4,t}}{Y_t}(\gamma_{2,t} + \gamma_{3,t} \ln \theta_t)(\gamma_{1,t+1} + \gamma_{3,t} \ln I_t)(1 + \alpha_{4,t+1}).$$

The short-term effect ($\Delta_{t+1,t}(\Omega_t)$) and the long-term effect ($\Delta_{t+2,t}(\Omega_t)$) can differ in general. Our model of skill and investment dynamics allows for the possibility that either short-term effects are higher than long-term effects (the effect of the policy "fades-out" as the child ages) or that long-term effects can exceed short-term effects (early interventions have a "multiplier effect" on later skill development).

4.3.2 Effects on Final Skills

We first consider a simple exercise designed to assess the optimal timing of the income transfer. In Figure 7 we show the average change in the latent children's log-skills at age 13-14 by the different timing (age) of income transfer:

$$E[\ln \theta'_T(a) - \ln \theta_T],$$

where $\theta'_T(a)$ is level of skill at age $t = T$ (age 13-14) with an income transfer of \$1,000 dollars (in 2012 \$) provided to the family at age a, and θ_T is level of skill at age 13-14 in the baseline model (no income transfer). The transfer is a one-time transfer and does not affect the future levels of income. The figure shows that a \$1,000 transfer given at age 5-6 increase the average stock of age 13-14 skills by about 1 percent. Providing the same transfer later in the childhood period has a smaller average effect. Providing a \$1,000 transfer at age 11-12 would increase the average skill stocks at age 13-14 by less than 0.4 percent. We estimate that providing transfers early in the development period would have a long-term effect that exceeds the short-term effect of providing a transfer in later childhood. This result reflects the high productivity of investment in the early periods and the high level of productivity of existing stocks of skill in producing future skills (limited fade-out).

4.3.3 Effects on Completed Schooling

Figure 8 displays the results of the same set of policy experiments as in Figure 7 but using completed schooling at age 23 as the outcome. In this Figure, we plot $E(S^{'}(a) - S)$, where $S^{'}(a)$ is the number of months of completed schooling at age 13-14 with an income transfer of \$1,000 given at age a, and S is the number of months of completed schooling at age 13-14 in the baseline model (no income transfer). These estimates provide a meaningful metric to evaluate the magnitude of the policy effects. We find that a \$1,000 transfer given at age 5-6 would increase the number of average months of completed schooling by about 1.80 months. Providing the same transfer at a later period would increase completed schooling by only 0.55 months.

4.3.4 Heterogeneous Treatment Effects

The previous results showed the average effect of policies providing transfers at different stages of the development process. Our modeling framework allows potentially important sources of heterogeneity by the child's initial skills, mother's skills, and initial family income levels; all of which could affect the individual level treatment effect. The model estimates allow us to directly estimate this heterogeneity in the policy treatment effects.

Figure 9 plots the heterogeneous effect of a \$1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income. This figure also plots the average treatment effect (ATE), the average effect over the income distribution; the same effect as reported above. While the ATE is about 1.8 months, the effect varies considerably depending on the child's initial level of income. For the children from poor households in the 9-10th income percentiles, the effect of the income transfer is to increase completed schooling by around 4 months, and for the children from the richest households, the effect is near 0. The large heterogeneous effects by family income stem from the estimated importance of family income in producing child investments and the estimated positive correlation of income with maternal skills and the child's initial skills. This heterogeneity in the effects by income mirrors the heterogeneity in income effects found in previous papers using alternative sources of identification (see Dahl and Lochner, 2012; Loken et al., 2012; Agostinelli and Sorrenti, 2018). Using the varied effects of the Norwegian oil boom to instrument for family income, Loken et al. (2012) report estimates on completed schooling which are smaller in magnitude than those reported here, but similar qualitatively in finding that the effects are substantially larger for low income Norwegian families

Figure 10 plots the heterogeneous effect of the same policy by the level of the child's initial (age 5-6) skill. The ATE plotted in this Figure is the same as in the previous figure as it is simply the effect averaged over the initial skill distribution. In this Figure, we also find evidence of heterogeneous treatment effects with low initial skill children benefiting more (about 3 months of additional schooling) from the policy intervention than high initial skill children (near 0 effect). But the importance of heterogeneity by initial skill is substantially less than by family income. This suggests that it is better to target the policy to low income households than low skill households, but of course it cannot be worse to target based on both criteria.

4.4 Comparing Model Predictions: Quantifying the Importance of Model Generality and Measurement Error

Our results presented thus far have been focused on the model with unrestricted production function (Model 1) and measurement error correction, estimated using what we believe are skill measures in the NLSY that are age-invariant. We next briefly discuss how the estimates of the primitive production technology would differ if we were to instead estimate the restricted model (Model 2) or ignore the measurement error issues. This analysis allows us to quantify how important measurement error and model generality are to our findings, using policy predictions on adult schooling as a meaningful metric for comparison.

Table 6 presents estimates for four versions of the model: Models 1 and 2, using both measurement error corrected and not corrected estimators. For each model and estimator, we re-estimate all parts of the model: the investment and technology process equations at each age and the final adult outcome equation. The estimates of the primitive parameters for these equations can be found in the Appendix; we present here only the implied policy effects.

In Panel A of Table 6, we present the average treatment effects (ATE) on adult schooling of the $1,000 income transfer at various ages. The first row shows the estimates for the model with unrestricted technology (Model 1): we estimate that

$1,000 income transfer at age 5-6 would increase average schooling by about 1.8 additional months. In comparison, using the restricted Model 2 (assuming restricted returns to scale and no TFP dynamics) would imply an estimated increase in average schooling of about one-quarter this effect, at 0.40 additional months. This shows that restrictions on the TFP dynamics and on the returns to scale would severely bias downward the implied effects of income transfers on children's skill development.

The next panel of Table 6 presents the estimated ATE using the same models but not correcting for measurement error. There is no clear theoretical prediction about the sign of the measurement error bias, given that our models are dynamic, non-linear, and consist of inter-related multiple equations. Using these uncorrected estimates, we estimate policy effects less than half the size of the measurement error corrected estimates of Model 1; a substantial reduction in the estimated effect of an income transfer. On the other hand, we find that the measurement error bias doubles the estimated impact of an income transfer for Model 2, although the results are still attenuated relative to the measurement error corrected estimates of Model 1.

Panel B of Table 6 repeats the analysis but focusing on the heterogeneity in the treatment effect at different parts of the family income distribution. Similar conclusions are evident here: restricting the returns to scale and the TFP dynamics, or ignoring measurement error, would substantially reduce the estimated policy effect of the income transfer. We see that ignoring measurement error would bias the estimated policy effect on low income families at the 9-10th percentile from an effect size of about 4 months to only 1.4 - 1.8 months.

4.5 Cost-Benefit Analysis

We have thus far shown that the estimated model implies that a policy intervention of providing income transfers to family would produce modest but positive gains in children's skills, with larger effects for poorer households. Would these gains be justified given the cost? We next present a simple cost-benefit analysis to answer this question.

Table 7 shows the effects of the income transfer policy, by children's age, on the present value of earnings. The Table also provides the associated cost of that policy, including the cost of additional schooling. In this analysis, we consider a median earner worker. The expected present value of her lifetime earnings when she is age 5-6 is calculated to be approximately $260,000 (in 2012 dollars).[21] The benefit of this policy is the comparison between the present value of worker's earnings

[21] The baseline present value of earnings is computed using data from the Bureau of Labor Statistics (BLS) for the fourth quarter of 2012 with a discount rate of 4 %.

with and without that policy during the childhood. In other words, we compute the counterfactual present value of earnings if the worker's family had received the income transfer when the worker was a child. The effect of the family income transfer to the growth in children earnings are computed using estimates in Table 5 under the assumption that the change in the growth rate due to the policy intervention is constant over the life-cycle. Table 7 suggests that, considering both the cost of the income transfer and the cost of additional education, the net benefit of the policy is positive for any age, and the effect is largest when implemented at age 5-6. The additional present value for the policy intervention at age 5-6 is slightly more than \$5,500 and the net benefit is around \$2,700.

5 Conclusion

This paper develops new identification concepts and associated estimators for the process of skill development in children. One of the key empirical challenges in this context is that the various measures of children's skills are in general imperfect and arbitrarily located and scaled. We introduce the concept of known location and scale production technologies, which are the class of technologies actually estimated in many previous papers, and show that for these technologies, standard measurement assumptions non-parametrically identify the production technology, up to the normalization of initial period skills. Importantly, we show non-parametric identification for these cases without re-normalizing latent skills each period, which when imposed can bias the production technology. For production functions which do not have a known location or scale, additional assumptions are necessary, and we provide empirically grounded assumptions which are sufficient for identification of these more general technologies. Our paper provides the first analysis of these crucial identification trade-offs, and hopefully will serve as a useful guide for future work.

Based on our identification results, we develop a robust method of moments estimator and show that it can be implemented using a sequential algorithm. Our estimator does not require strong assumptions about the marginal distribution of measurement errors or the latent factors. We estimate the skill production process using data for the United States and a flexible parametric model of skill development allowing for free returns to scale, dynamics in TFP, and for parental investment to endogenously depend on unobserved children's skills.

Our parameter estimates reveal that investments are more productive at early ages and in particular for disadvantaged children. Our findings of a positive return to income transfers at early ages, especially for poorer households, is largely consistent with prior evidence of a positive effect of income on a number of child outcomes (see

Dahl and Lochner, 2012; Loken et al., 2012; Agostinelli and Sorrenti, 2018) using different sources of identification. Our results suggest that family income is a better "target" than initial children's skills for children's skills. Lastly, our finding that the estimated policy effects would be substantially smaller if one estimated a restricted technology or ignored measurement error demonstrates the critical importance of allowing for general technologies and correcting estimates for measurement error.

References

Karun Adusumilli and Taisuke Otsu. Nonparametric instrumental regression with errors in variables. *Econometric Theory*, 34(6):1256–1280, 2018.

Francesco Agostinelli. Investing in childrens skills: An equilibrium analysis of social interactions and parental investments. *Working Paper*, 2019.

Francesco Agostinelli and Giuseppe Sorrenti. Money versus time: family income, maternal labor supply, and child development. *Working Paper 2018-017, HCEO, University of Chicago*, 2018.

Francesco Agostinelli and Matthew Wiswall. Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. *Working Paper*, 2016.

W. Anderson and Herman Rubin. Statistical inference in factor analysis. *Proc. Third Berkeley Symp. on Math. Statist. and Prob*, 5:111–150, 1956.

Orazio Attanasio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. Estimating the production function for human capital: Results from a randomized control trial in colombia. *NBER Working Paper*, (21740), 2015a.

Orazio Attanasio, Costas Meghir, and Emily Nix. Human capital development and parental investment in india. *NBER Working Paper*, (20965), 2015b.

Orazio Attanasio, Flvio Cunha, and Pamela Jervis. Subjective parental beliefs: Their measurement and role. *NBER Working Paper*, 2019a.

Orazio Attanasio, Costas Meghir, and Emily Nix. Human capital development and parental investment in india. Technical report, National Bureau of Economic Research, 2019b.

Orazio Attanasio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. Estimating the production function for human capital: Results from a randomized control trial in colombia. *American Economic Review*, forthcoming.

Raquel Bernal. The effect of maternal employment and child care on children's cognitive development. *International Economic Review*, 49(4):1173–1209, 2008.

Timothy N. Bond and Kevin Lang. The evolution of the black-white test score gap in grades k3: The fragility of results. *Review of Economics and Statistics*, 95(5): 1468–1479, 2013.

Timothy N. Bond and Kevin Lang. The black-white education-scaled test-score gap in grades k-7. *Journal of Human Resources*, Fall(53):891–917, 2018.

Pedro Carneiro, Karsten T Hansen, and James J Heckman. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. *International Economic Review*, 2003.

Elizabeth M. Caucutt, Lance Lochner, and Youngmin Park. Correlation, consumption, confusion, or constraints: Why do poor children perform so poorly? *NBER Working Papers*, (21023), 2015.

Gary Chamberlain. Education, income, and ability revisited. *Journal of Econometrics*, pages 241–257, 1977a.

Gary Chamberlain. An instrumental variable interpretation of identification in variance components and mimic models. *in Paul Taubman, ed., Kinometrics: Determinants of Socio-Economic SuccessWithin and Between Families (Amsterdam: North-Holland)*, 1977b.

Gary Chamberlain and Zvi Griliches. Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers. *International Economic Review*, 16:422–449, 1975.

Flavio Cunha. Investments in children when markets are incomplete. *Working Paper*, 2013.

Flavio Cunha and James Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, 2007.

Flavio Cunha and James J. Heckman. Formulating, identifying, and estimating the technology for the formation of skills. *Journal of Human Resources*, 43(4):738–782, 2008.

Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Flvio Cunha, Irma Elo, and Jennifer Culhane. Eliciting maternal expectations about the technology of cognitive skill formation. *NBER working paper*, (19144), 2013.

Gordon B. Dahl and Lance Lochner. The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, 102(5):1927–1956, 2012.

Daniela Del Boca, Christopher Flinn, and Matthew Wiswall. Household choices and child development. *Review of Economic Studies*, 81(1):137–185, 2014a.

Daniela Del Boca, Christopher Flinn, and Matthew Wiswall. Transfers to households with children and child development. *Working Paper*, 2014b.

Matthias Doepke and Fabrizio Zilibotti. Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85(5):1331–1371, 2017.

Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. The economics of parenting. *Annual Review of Economics*, 11(1):55–84, 2019.

L Dunn and L Dunn. The peabody picture vocabulary test-revised. *Circle Pines, MN: American Guidance Service*, 1981.

Arthur S. Goldberger. Structural equation methods in the social sciences. *Econometrica*, 40:979–1001, 1972.

Gosta Hagglund. Factor analysis by instrumental variables methods. *Psychometrika*, 47:209–22, 1982.

J. Hausman, H. Ichimura, W. Newey, and J Powell. Measurement errors in polynomial regression models. *Journal of Econometrics*, 50:271–295, 1991.

James J. Heckman and Stefano Mosso. The economics of human development and social mobility. *NBER Working Paper*, (19925), 2014.

Karl G. Jöreskog and Arthur S. Goldberger. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70:631–639, 1975.

Cory Koedel and Julian Betts. Value added to what? how a ceiling in the testing instrument influences value-added estimation. *Education Finance and Policy*, 5 (1):54–81, 2010.

Katrine V. Loken, Magne Mogstad, and Matthew Wiswall. What linear estimators miss: Re-examining the effects of family income on child outcomes. *American Economic Journal: Applied Economics*, 4(2):1–35, 2012.

Ronni Pavan. On the production of skills and the birth order effect. *Working Paper*, 2015.

Susanne M. Schennach. Estimation of nonlinear models with measurement error. *Econometrica*, (72):33–75, 2004.

Lijuan Wang, Zhiyong Zhang, John J. McArdle, and Timothy A. Salthouse. Investigating ceiling effects in longitudinal data analysis. *Multivariate behavioral research*, 43(3):476–496, 2009.

Table 1: Sample Descriptive Statistics

	Mean	Std
N Obs	19,070	
N of Mothers	3,199	
N of Children	4,941	
% Male Children	51.32	
% Female Children	48.68	
% Hispanic Children	21.44	
% Black Children	30.44	
% Other races	48.12	
Mom Education	12.59	2.63
Family Income	61,657.88	47,527.85
Children Final Years of Education	13.30	2.36

Notes: This table shows the main descriptive statistics of the CNLSY79 sample we use to estimate the model. Children's Completed Education is the child's completed years of education at age 23. The variable "other races" represents all children which are not black neither Hispanic (i.e. it includes white, non-Hispanic children). Income is in $2012 USD.

Table 2: Estimates for Initial Conditions

	Log Child Skills at age 5	Log Mother Cognitive Skills	Log Mother Noncognitive Skills	Log Family Income
		Variance-Covariance Matrix		
Log Child Skills at age 5	4.947 (0.471)	6.254 (0.479)	0.122 (0.031)	0.668 (0.065)
Log Mother Cognitive Skills	6.254 (0.479)	30.190 (1.032)	0.593 (0.137)	2.588 (0.099)
Log Mother Noncognitive Skills	0.122 (0.031)	0.593 (0.137)	0.046 (0.017)	0.058 (0.012)
Log Family Income	0.668 (0.065)	2.588 (0.099)	0.058 (0.012)	0.780 (0.018)
		Correlation Matrix		
Log Child Skills at age 5	1.000 (-)	0.512 (0.026)	0.256 (0.029)	0.340 (0.027)
Log Mother Cognitive Skills	0.512 (0.026)	1.000 (-)	0.504 (0.025)	0.533 (0.015)
Log Mother Noncognitive Skills	0.256 (0.029)	0.504 (0.025)	1.000 (-)	0.307 (0.022)
Log Family Income	0.340 (0.027)	0.533 (0.015)	0.307 (0.022)	1.000 (-)

Notes: This table shows the estimated variance-covariance matrix (Σ_Ω) and associate correlation matrix of the initial conditions at age 5-6. The initial conditions estimates do not depend on the modeling assumptions for the skill production technology or any other remaining parts of the model. Standard errors in parenthesis are computed using a clustered bootstrap at the family level.

Table 3: Estimates for Investment (Model 1: Age-Invariance)

Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.230	0.027	0.020	0.018
	(0.059)	(0.009)	(0.009)	(0.009)
	[0.14, 0.33]	[0.01, 0.04]	[0.01, 0.04]	[0.01, 0.03]
Log Mother Cognitive Skills	0.071	0.004	0.012	-0.005
	(0.022)	(0.009)	(0.015)	(0.013)
	[0.04, 0.12]	[-0.01, 0.02]	[-0.01, 0.04]	[-0.02, 0.02]
Log Mother Noncognitive Skills	0.359	0.742	0.694	0.712
	(0.131)	(0.060)	(0.084)	(0.088)
	[0.11, 0.54]	[0.64, 0.82]	[0.52, 0.81]	[0.54, 0.82]
Log Family Income	0.341	0.227	0.274	0.275
	(0.076)	(0.056)	(0.076)	(0.087)
	[0.25, 0.48]	[0.15, 0.33]	[0.17, 0.43]	[0.17, 0.44]
Variance Shocks	1.186	1.019	0.868	1.087
	(0.232)	(0.148)	(0.236)	(0.296)
	[0.96, 1.53]	[0.83, 1.29]	[0.66, 1.33]	[0.82, 1.64]

Notes: This table shows the measurement error corrected estimates for the investment equation for Model 1 (Measurement Function Restriction: Age-Invariance). Each column shows the coefficients of the investment equation at the given ages. The dependent variable is (log) investment in period t, determined by the RHS variables at time t. For example, the first column shows the coefficients at age 5-6 parental investments determined by age 5-6 child's skill and family income. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level.

Table 4: Estimates for Skill Technology (Model 1 and Model 2)

	Model 1 (Measurement Function Restrictions)				Model 2 (Production Function Restrictions)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	1.966	1.086	0.897	1.065	0.739	0.816	0.833	0.910
	(0.153)	(0.036)	(0.027)	(0.029)	(0.087)	(0.072)	(0.105)	(0.096)
	[1.69, 2.21]	[1.03, 1.15]	[0.84, 0.93]	[1.01, 1.11]	[0.61, 0.88]	[0.69, 0.93]	[0.71, 1.02]	[0.76, 1.07]
Log Investment	0.799	0.695	0.713	0.252	0.300	0.187	0.170	0.087
	(0.262)	(0.339)	(0.404)	(0.541)	(0.077)	(0.069)	(0.097)	(0.095)
	[0.41, 1.23]	[0.15, 1.24]	[-0.10, 1.25]	[-0.53, 1.20]	[0.18, 0.42]	[0.08, 0.32]	[-0.01, 0.30]	[-0.07, 0.23]
(Log Skills *	-0.105	-0.005	-0.003	0.003	-0.040	-0.004	-0.003	0.003
Log Investment)	(0.066)	(0.019)	(0.013)	(0.010)	(0.026)	(0.015)	(0.014)	(0.009)
	[-0.22,-0.03]	[-0.04, 0.03]	[-0.02, 0.02]	[-0.02, 0.02]	[-0.09,-0.01]	[-0.03, 0.02]	[-0.03, 0.02]	[-0.02, 0.01]
Return to scale	2.660	1.776	1.606	1.320	1.000	1.000	1.000	1.000
	(0.225)	(0.317)	(0.398)	(0.535)	(-)	(-)	(-)	(-)
	[2.30, 3.02]	[1.25, 2.31]	[0.79, 2.14]	[0.58, 2.25]	[-,-]	[-,-]	[-,-]	[-,-]
Variance shocks	5.612	4.519	3.585	4.019	2.110	1.279	0.944	0.903
	(0.174)	(0.184)	(0.181)	(0.247)	(0.178)	(0.144)	(0.163)	(0.165)
	[5.37, 5.93]	[4.27, 4.89]	[3.27, 3.88]	[3.70, 4.46]	[1.88, 2.44]	[1.09, 1.57]	[0.78, 1.32]	[0.74, 1.33]
Log TFP	13.067	14.747	11.881	2.927	0.000	0.000	0.000	0.000
	(0.295)	(0.367)	(0.541)	(0.957)	(-)	(-)	(-)	(-)
	[12.67,13.61]	[14.22,15.47]	[11.17,13.00]	[1.38, 4.65]	[-,-]	[-,-]	[-,-]	[-,-]

Notes: This table shows the measurement error corrected estimates for the technology of skill formation for both Model 1 (Measurement Function Restriction: Age-Invariance) and Model 2 (Production Function Restriction: Known Location and Scale). Each column shows the coefficients of the technology of skill formations at the given age. The dependent variable is log skills in the next period $t+1$, and the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level.

Table 5: Estimates for Adult Outcome Equation (Model 1)

	Schooling	Log Wage
Constant	7.088	9.444
	(0.399)	(0.121)
	[6.56, 7.71]	[9.26, 9.64]
Log Children Skills at age 13-14	0.151	0.021
	(0.010)	(0.003)
	[0.14, 0.16]	[0.02, 0.03]
Variance Shock	4.333	0.246
	(0.143)	(0.012)
	[4.07, 4.56]	[0.22, 0.26]

Notes: This table shows the estimates for two adult outcome equation specifications: schooling and log earnings. In both cases the estimates are for Model 1 (Measurement Function Restriction: Age-Invariance) and they are corrected for measurement error. The dependent variable is either the years of completed education for the child at age 23 or log earnings at age 29. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level.

Table 6: Estimated Policy Effects under Different Modeling Assumptions

Panel A: ATE by Age of Income Transfer

Measurement Error Corrected

Model	Age of Income Transfer ($ 1000) Age 5-6	Age 7-8	Age 9-10	Age 11-12
Model 1	1.818	0.799	1.025	0.574
	[0.93, 2.56]	[0.29, 1.33]	[-0.05, 2.15]	[-0.39, 1.74]
Model 2	0.404	0.179	0.229	0.128
	[0.22, 0.64]	[0.07, 0.32]	[-0.02, 0.42]	[-0.10, 0.36]

Not Corrected for Measurement Error

Model	Age of Income Transfer ($ 1000) Age 5-6	Age 7-8	Age 9-10	Age 11-12
Model 1	0.687	0.220	0.210	0.251
	[0.48, 0.90]	[0.09, 0.36]	[0.07, 0.34]	[0.06, 0.47]
Model 2	0.846	0.271	0.259	0.309
	[0.62, 1.06]	[0.12, 0.44]	[0.09, 0.41]	[0.08, 0.55]

Panel B: ATE at age 5-6 by Family Income

Measurement Error Corrected

Low Income Families (10^{th} Income Percentile)		High Income Families (90^{th} Income Percentile)	
Model 1	4.11	Model 1	0.313
Model 2	0.91	Model 2	0.070

Not Corrected for Measurement Error

Low Income Families (10^{th} Income Percentile)		High Income Families (90^{th} Income Percentile)	
Model 1	1.465	Model 1	0.158
Model 2	1.806	Model 2	0.194

Notes: Panel A shows the average treatment effects on additional months of completed education by age of policy intervention ($ 1000 income transfer) for different model specifications – Model 1 (Measurement Function Restriction: Age-Invariance) and Model 2 (Production Function Restriction: Known Location and Scale – and different estimators – controlling for measurement error or not. The 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level. Panel B shows the ATE respect to family income for the different model specifications and different estimators.

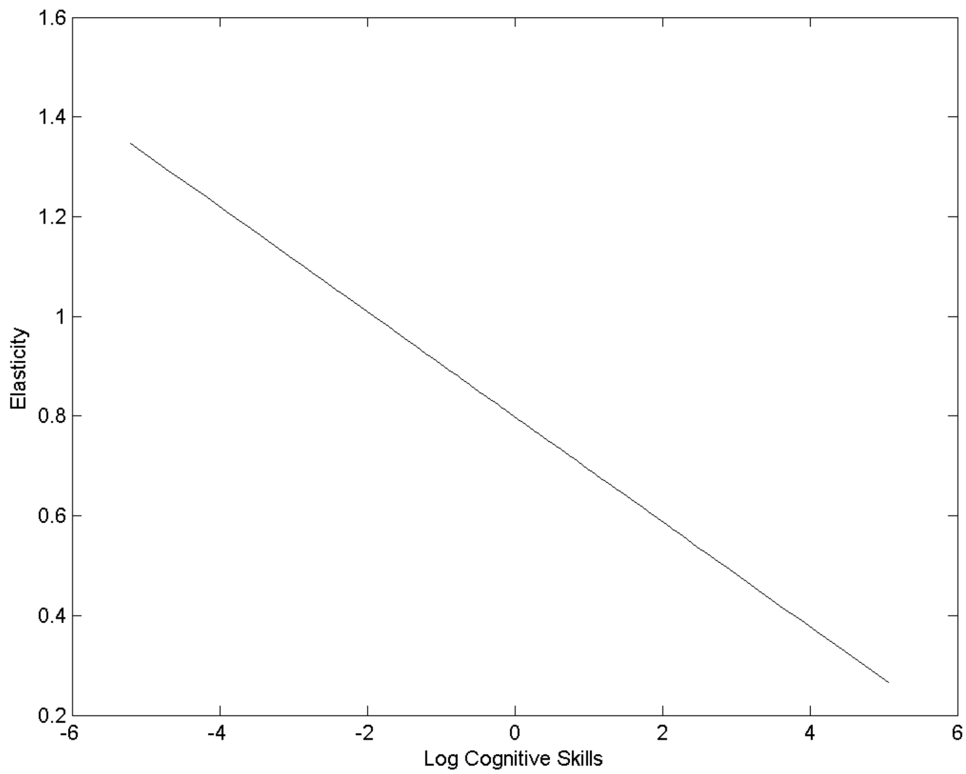
Table 7: Average Effect of an Income Transfer by Age of Transfer (Outcome: PV of Earnings)

Panel A: Benefit-Cost Analysis by Age

Age of Intervention	Benefit on PV Earnings ($)	Direct Cost (Income Transfer) ($)	Cost of Education ($)	Net Benefit ($)
Age 5-6	5549	1000	1818	2730
Age 7-8	2437	1000	799	638
Age 9-10	3128	1000	1025	1103
Age 11-12	1750	1000	574	177

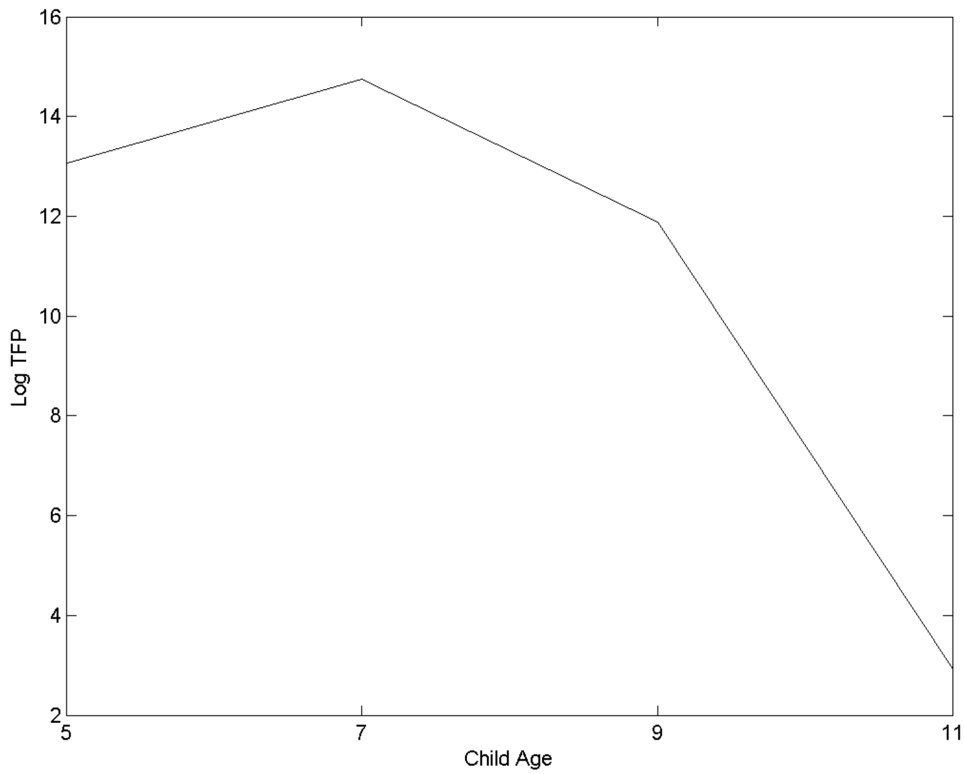
Notes: This table shows the benefit-cost analysis for a 1000 dollars transfer to family of a future median earner workers with 12 years of completed education. The benefit on the PV of earnings is the difference between the present value of earnings with and without that transfer when worker was age 5-6. The effect of family income transfer on earnings growth is computed adjusting for the increased earning growth implied by estimates in Table 5. The cost of that policy takes into account both the direct transfer and the discounted cost of additional education that the policy induces. We use a yearly cost of school of 12,000 dollars as approximately estimated from the National Center for Education Statistics.

Figure 1: Estimates of Skill Production Elasticity with Respect to Investment at Age 5-6 (Model 1)



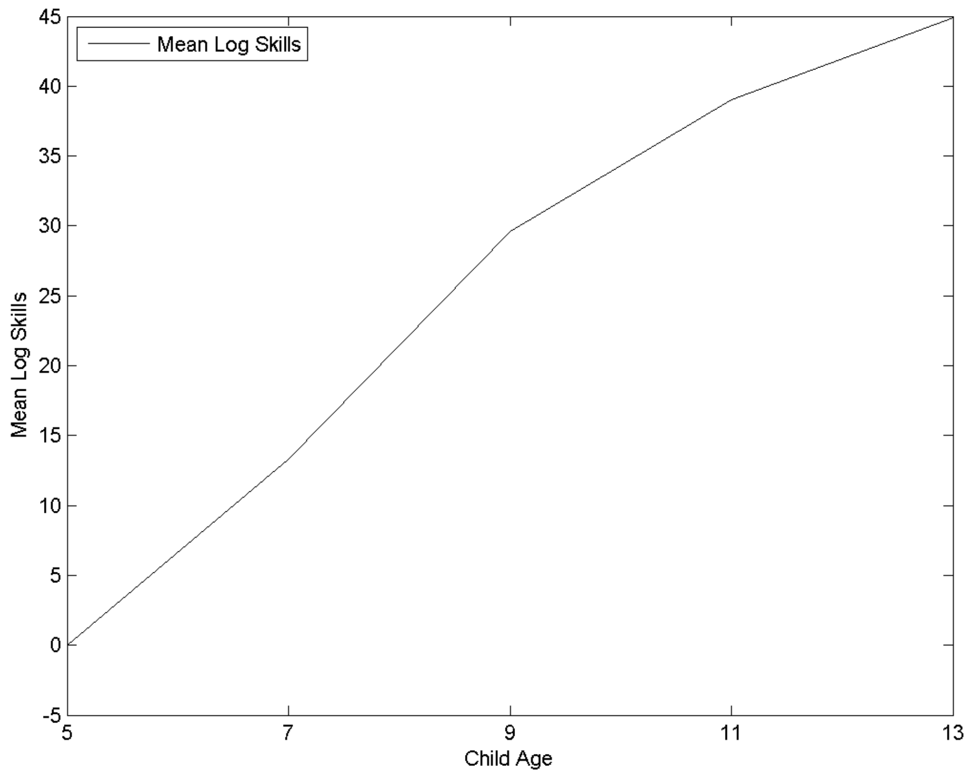
Notes: This figure shows the measurement error corrected estimates of the elasticity of children's skills at age 7-8 (θ_1) with respect to parental investments at age 5-6 (I_0) for Model 1 (Measurement Function Restriction: Age-Invariance): $\frac{\partial \ln \theta_1}{\partial \ln I_0} = \gamma_{2,0} + \gamma_{3,0} \ln \theta_0$.

Figure 2: Total Factor Productivity (TFP) Estimates (Model 1)



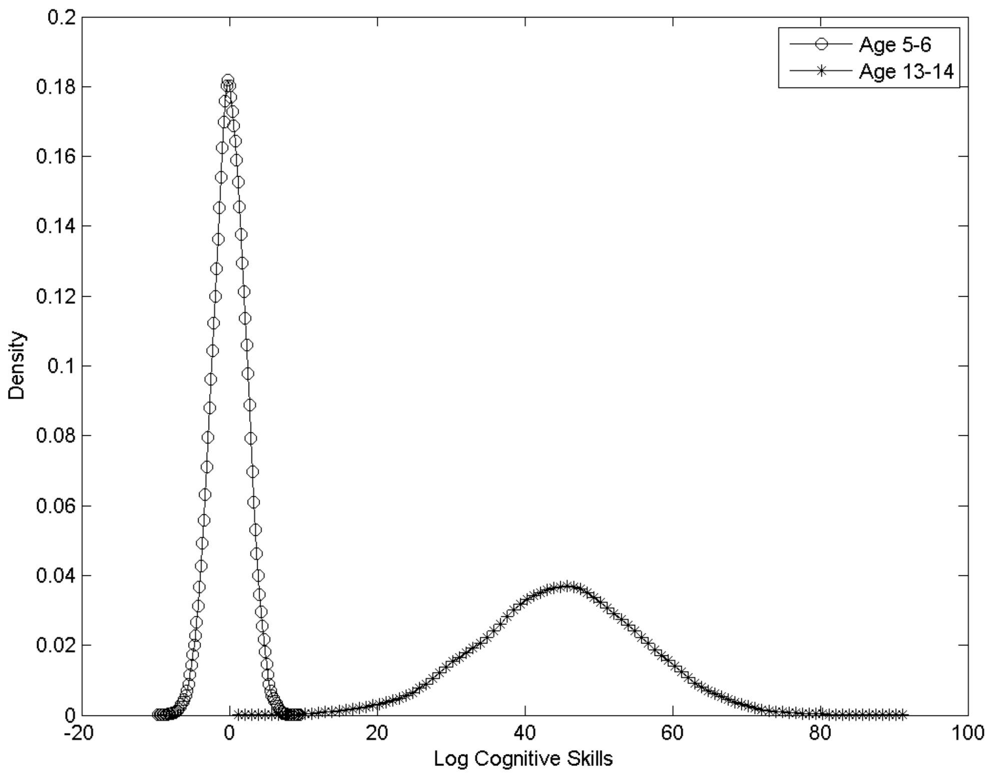
Notes: This figure shows the estimated log TFP (correcting for measurement error) for Model 1 (Measurement Function Restriction: Age-Invariance). The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on.

Figure 3: Estimated Mean of Log Latent Skills (Model 1)



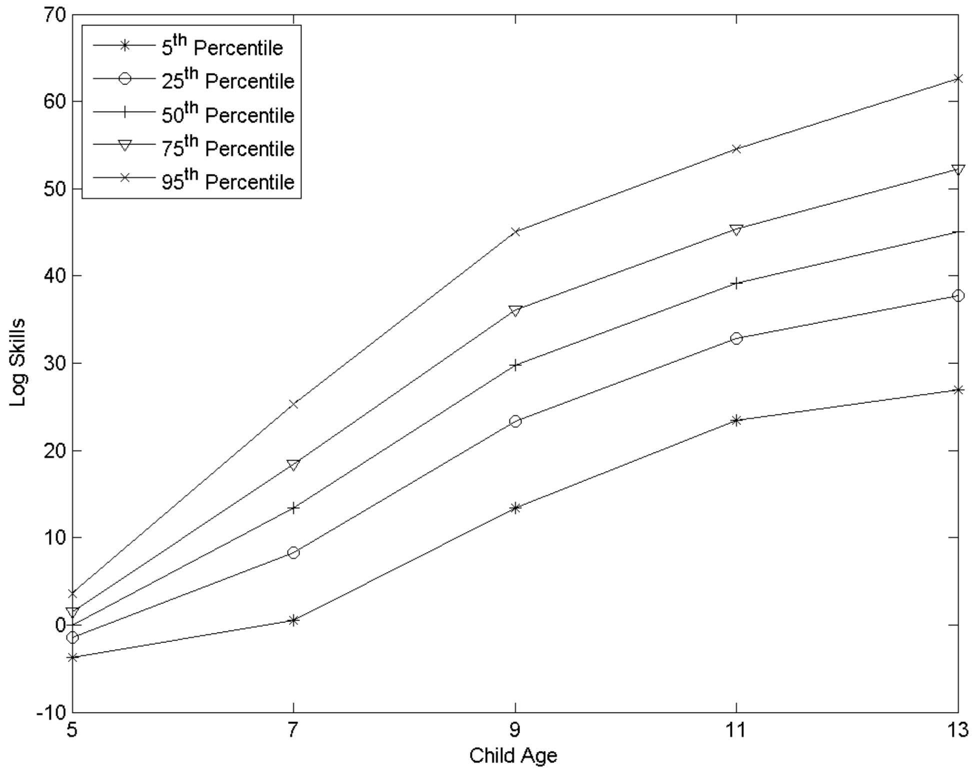
Notes: This figure provides the mean log latent skills ($E(\ln \theta_t)$) predicted by the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error . The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on. Log latent skills at age 5-6 are normalized to be mean 0.

Figure 4: Estimated Distribution of Log Cognitive Latent Skills at Age 5-6 and Age 13-14 (Model 1)



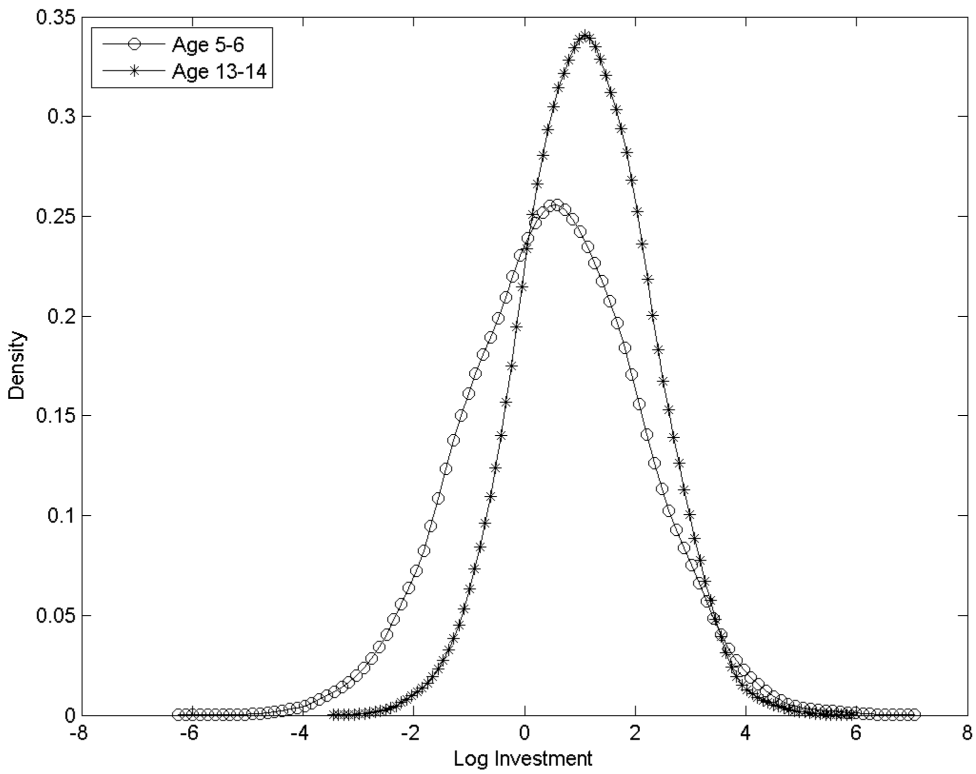
Notes: This figure shows the distribution of log latent skills at age 5-6 and at age 13-14 simulated from the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error. Log latent skills at age 5-6 are normalized to be mean 0.

Figure 5: Estimated Dynamics in the Latent Skills Distribution (Model 1)



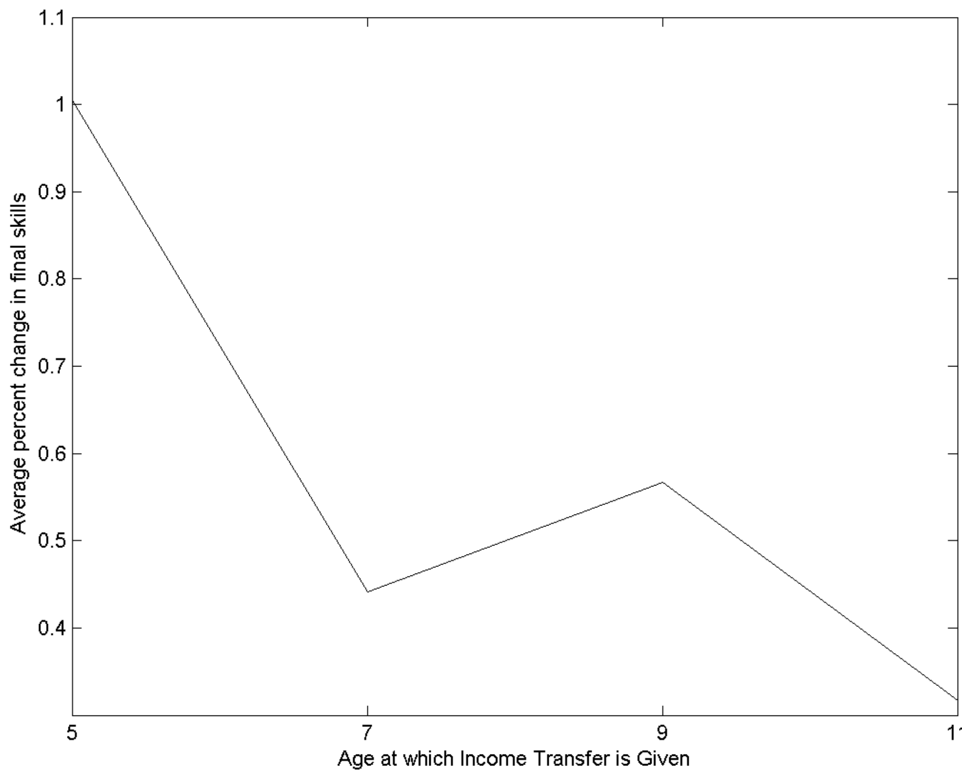
Notes: This figure shows the dynamics in the distribution of the log latent skill distribution for the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error. Log latent skills at age 5-6 are normalized to be mean 0.

Figure 6: Estimated Distribution of Log Investments at Age 5-6 and Age 13-14 (Model 1)



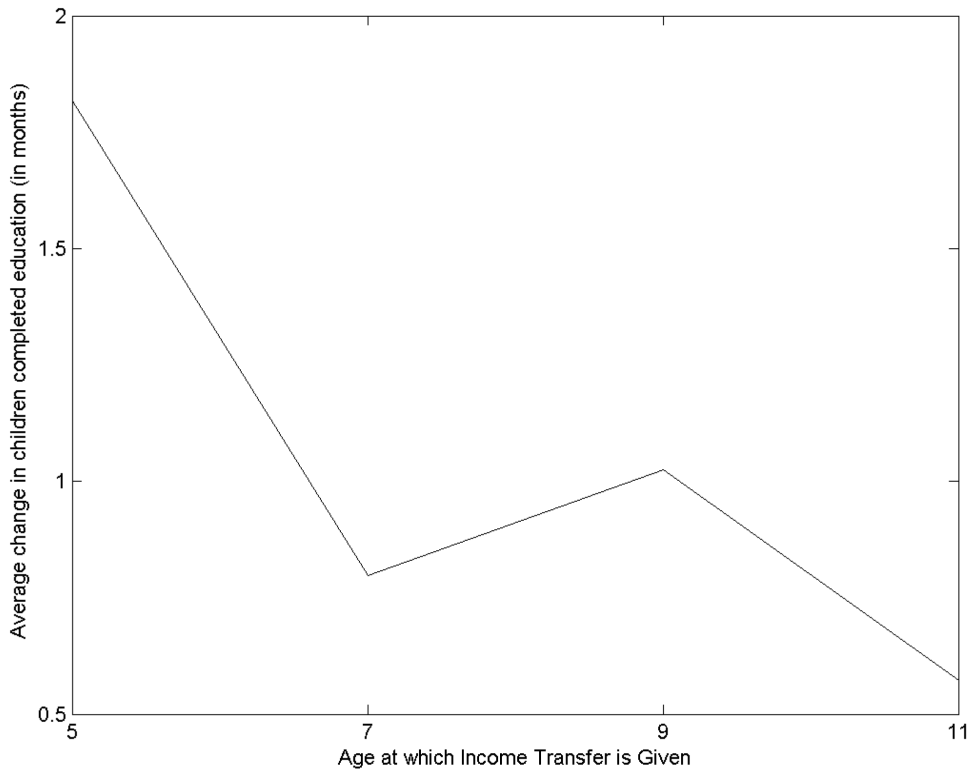
Notes: This figure shows the distribution of log latent investments at age 5-6 and at age 13-14 simulated from the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error.

Figure 7: Average Effect of Income Transfer by Age of Transfer (Outcome: Final Period θ_T Skills)



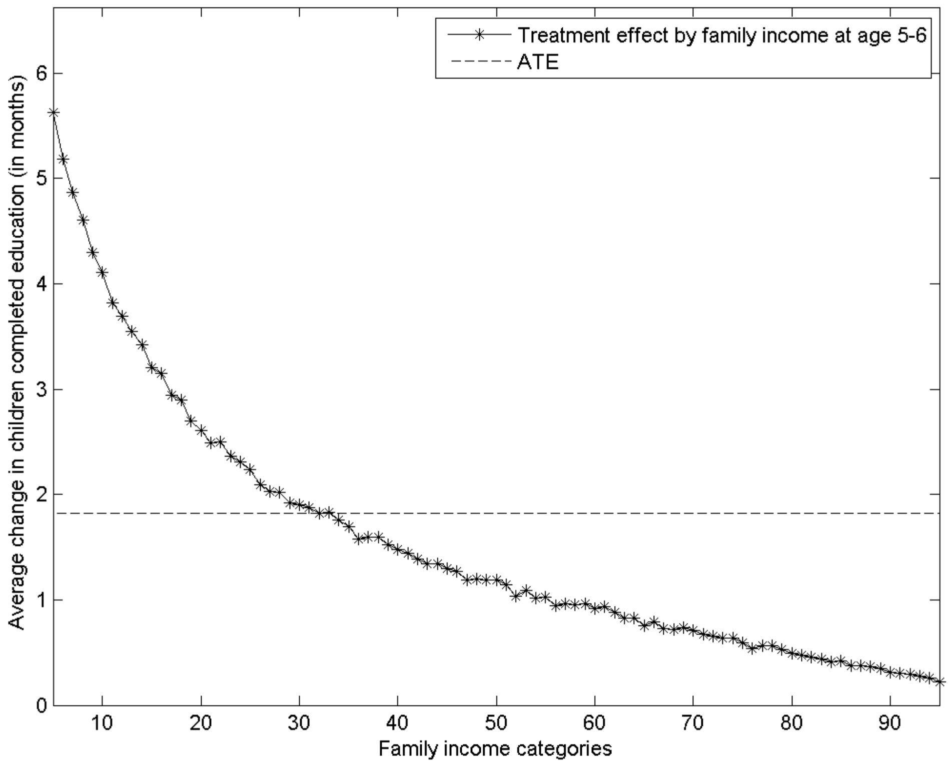
Notes: This figure shows the average percent change in the level of latent children's skills at age 13-14 by the different timing (age) of income transfer for the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error. The transfer is $1,000 in family income at some age t. We report $100 \cdot E(\frac{\theta_T^{'}(a)-\theta_T}{\theta_T})$, where $\theta_T^{'}(a)$ is level of skill at age 13-14 with an income transfer of $1,000 dollars provide to the family at age a and θ_T is level of skill at age 13-14 in the baseline model (no income transfer).

Figure 8: Average Effect of an Income Transfer by Age of Transfer (Outcome: Schooling at Age 23)



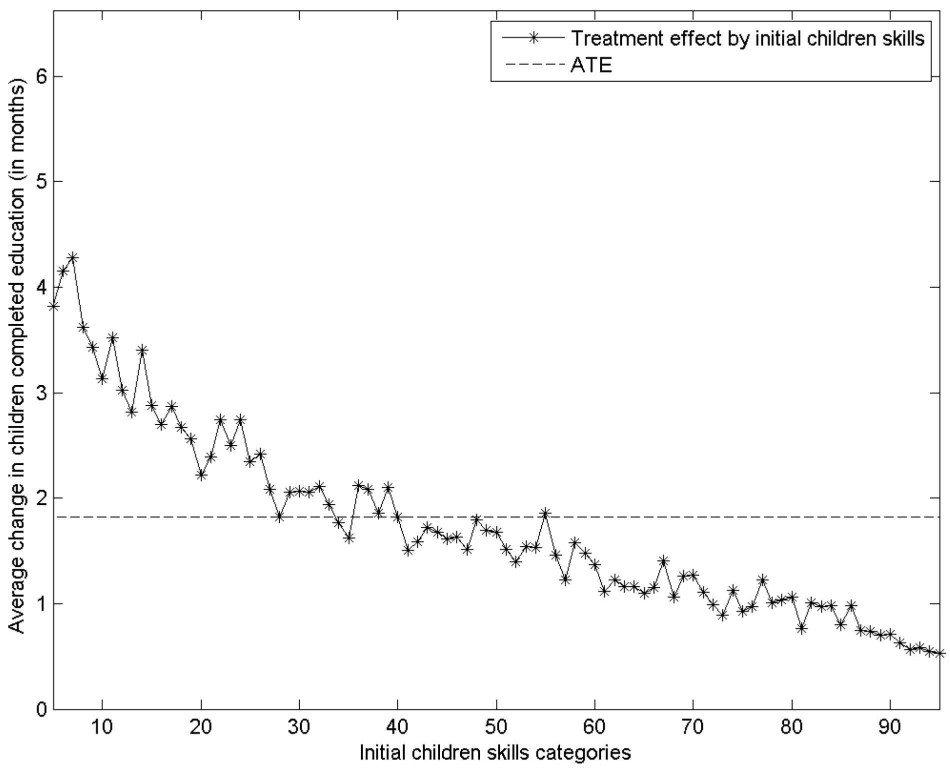
Notes: This figure shows the average change in the number of months of completed schooling at age 23 by different timing (age) of income transfer for the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error. We report $E\left[S^{'}(a) - S\right]$, where $S^{'}(a)$ is the number of months of completed schooling at age 23 with an income transfer of \$1,000 given at age a while S is the number of months of completed schooling in baseline model (no income transfer). This figure reports the results of the same policy experiment as Figure 7 but with a different outcome measure.

Figure 9: Heterogeneity in Policy Effects by Age 5-6 Household Income (Outcome: Schooling at Age 23)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

Figure 10: Heterogeneity in Policy Effects by Age 5-6 Children's Skills (Outcome: Schooling at Age 23)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of the child's initial (age 5-6) skill for the estimated Model 1 (Measurement Function Restriction: Age-Invariance), controlling for measurement error. Each initial skills category includes the children contained between n^{th} and the $n-1^{th}$ of the percentiles of the skills distribution. For example, skill category 10 is the children between the 9^{th} and 10^{th} percentile of the initial skills distribution. This figure also plots the average effect over the initial skill distribution.

ONLINE APPENDIX

A Technologies and Output Elasticities

One rationale for the choice of a technology specification with free returns to scale is the flexibility this specification offers with respect to the implied output elasticity. We consider the output elasticity with respect to investment defined as

$$\epsilon_I \equiv \frac{\partial \ln \theta_{t+1}}{\partial \ln I_t}$$

This elasticity is key to quantifying the effects of policy interventions.

In the general CES case, with technology given by

$$\theta_{t+1} = \left[\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}\right]^{\frac{\psi_t}{\phi_t}},$$

the output elasticity is given by

$$\epsilon_I = \frac{\psi_t}{\phi_t}\left[\gamma \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}\right]^{\frac{\psi_t}{\phi_t}-1} \phi(1-\gamma_t) I_t^{\phi_t - 1} \cdot \frac{I_t}{\left[\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}\right]^{\frac{\psi_t}{\phi_t}}}$$

$$= \frac{\psi_t (1-\gamma_t) I_t^{\phi_t}}{\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}} \in [0, \infty)$$

In the special case of constant returns to scale (CRS), $\psi_t = 1$, and $\epsilon_I \in (0,1)$. CRS implies this elasticity is bounded from above by 1. The general free returns to scale case allows a larger than unit elastic response.

Similarly, the general translog technology,

$$\ln \theta_{t+1} = \alpha_{1t} \ln \theta_t + \alpha_{2t} \ln I_t + \alpha_{3t} \ln \theta_t \ln I_t$$

with elasticity

$$\epsilon_I = \alpha_{1t} + \alpha_{3t} \ln \theta_t$$

also allows higher than unit elastic elasticities.

The main insight we want to underline is that the CES technology with constant returns to scale restricts the output elasticity to be between 0 and 1: a one percent change in investment leads to a less than one percent change in next period skills. This prediction is independent of data, hence it can potentially be very restrictive in the context of child development and skill formation.

B Multiple Skills and Multiple Investments

In this Appendix, we generalize the model to the case where there are multiple skills and investments. The identification results from the scalar case carry through essentially unchanged to this case where the skills and investments are vectors.

B.1 Model

Let $\Theta_{i,t} = \{\theta_{i,t,1}, \ldots, \theta_{i,t,K}\}$ be a vector K skills. Let $\mathbf{I}_{i,t} = \{I_{i,t,1}, \ldots, I_{i,t,L}\}$ be a vector of L investments. The skill formation technology for each of the $k = 1, \ldots, K$ skills is then

$$\theta_{i,t+1,k} = h_t(\Theta_{i,t}, \mathbf{I}_{i,t}, \eta_{i,t,k}) \tag{B-1}$$

All K skill stocks and all L investments potentially produce next period's k skills.

Each measure m for child i's skills in k in period (age) t is given by

$$Z_{i,t,m,k} = g_{t,m,k}(\theta_{i,t,k}, \epsilon_{i,t,m,k}) \tag{B-2}$$

That is, there is a dedicated measure for each skill.

Substituting as in the main text, we have

$$\begin{aligned} \ln \theta_{i,t+1,k} &= \ln f_{t,k}(\Theta_{i,t}, \mathbf{I}_{i,t}) + \eta_{i,t,k} \\ Z_{i,t,m,k} &= \mu_{t,m,k} + \lambda_{t,m,k} \ln \theta_{i,t,k} + \epsilon_{i,t,m,k} \quad . \end{aligned} \tag{B-3}$$

with all of the associated conditions.

B.2 Initial Conditions

Dropping the i subscript, as in the main text. The generalized normalization for the initial period is that for each $k = 1, \ldots, K$, we have

Normalization B-1 *Initial period normalizations*

(i) $E(\ln \theta_{0,k}) = 0$

(ii) $\lambda_{0,1,k} = 1$

The generalized measurement assumption is that for each $k = 1, \ldots, K$, we have

Assumption B-1 *Initial Period Measurement Assumptions:*
(i) $Cov\left(\epsilon_{0,m,k}, \epsilon_{0,m',k}\right) = 0$ *for all* $m \neq m'$
(ii) $Cov\left(\epsilon_{0,m,k}, \ln \theta_{0,k}\right) = 0$ *for all* m*.*

As above, we can use these conditions to identify the joint distribution of initial latent skills and the initial period measurement parameters.

B.3 Identification of the Technology

As in the scalar case, we have for each $k = 1, \ldots, K$

$$\begin{aligned} Z_{1,m,k} &= \mu_{1,m,k} + \lambda_{1,m,k} \ln f_{0,k}(\Theta_0, \mathbf{I}_0) + (\eta_{0,k} + \lambda_{1,m,k}\epsilon_{1,m,k}) \\ &= q_{0,k}(\Theta_0, \mathbf{I}_0) + u_{0,m,k} \end{aligned} \tag{B-4}$$

where the combined residual $u_{0,m,k} = \eta_{0,k} + \lambda_{1,m,k}\epsilon_{1,m,k}$ is mean-zero.

Following the results from the scalar case, a sufficient condition for identification of each $q_{0,k}$ is (a) an instrument $W_{0,k}$ which satisfies (i) $E(u_{0,m,k}|W_{0,k}) = 0$ and (ii) $\widetilde{\epsilon}_{0,m,k} \perp W_{0,k}$, and (b) $\widetilde{\epsilon}_{0,m,k} \perp \theta_{0,k}$, for all k.

Without loss of generality, we write the first term of the production technology as

$$\ln f_{0,k}(\Theta_0, \mathbf{I}_0) = \ln A_{0,k} + \psi_{0,k} \ln K_{0,k}(\Theta_0, \mathbf{I}_0), \tag{B-5}$$

$K_{0,k}(\Theta_0, \mathbf{I}_0)$ is a Known Location and Scale (KLS) function, which we define for the multiple skills and multiple investment case as follows:

Definition B-1 *A production function* $K_{t,k}(\Theta_t, \boldsymbol{I}_t)$ *has* Known Location and Scale (KLS) *if for two non-zero input vectors* $(\Theta'_t, \boldsymbol{I}'_t)$ *and* $(\Theta''_t, \boldsymbol{I}''_t)$*, where the input vectors are distinct, the outputs* $K_{t,k}(\Theta'_t, \boldsymbol{I}'_t)$ *and* $K_{t,k}(\Theta''_t, \boldsymbol{I}''_t)$ *are both known (do not depend on unknown parameters), finite, and non-zero.*

Next, substituting we have for each k:

$$Z_{1,m,k} = (\mu_{1,m,k} + \ln A_{0,k}\lambda_{1,m,k}) + (\lambda_{1,m,k}\psi_{0,k}) \ln K_{0,k}(\Theta_0, \mathbf{I}_0) + u_{0,m,k} \tag{B-6}$$

Generalizing to the multiple skills and investment case, we consider identification under one of two prototypical restrictions applied to each $k = 1, \ldots, K$ skills

Assumption B-2 *Measurement Function Restriction* $\mu_{t,m,k} = \mu_{0,m,k}$ *and* $\lambda_{t,m,k} = \lambda_{0,m,k}$ *for all* $t > 0$ *and this* m

Assumption B-3 *Production Function Restriction* $\ln A_{t,k} = 0$ *and* $\psi_{t,k} = 1$ *for all* t

Under *either* set of restrictions, we identify all of the parameters of interest, following the same proof as in the scalar case.

C Additional Tables and Figures

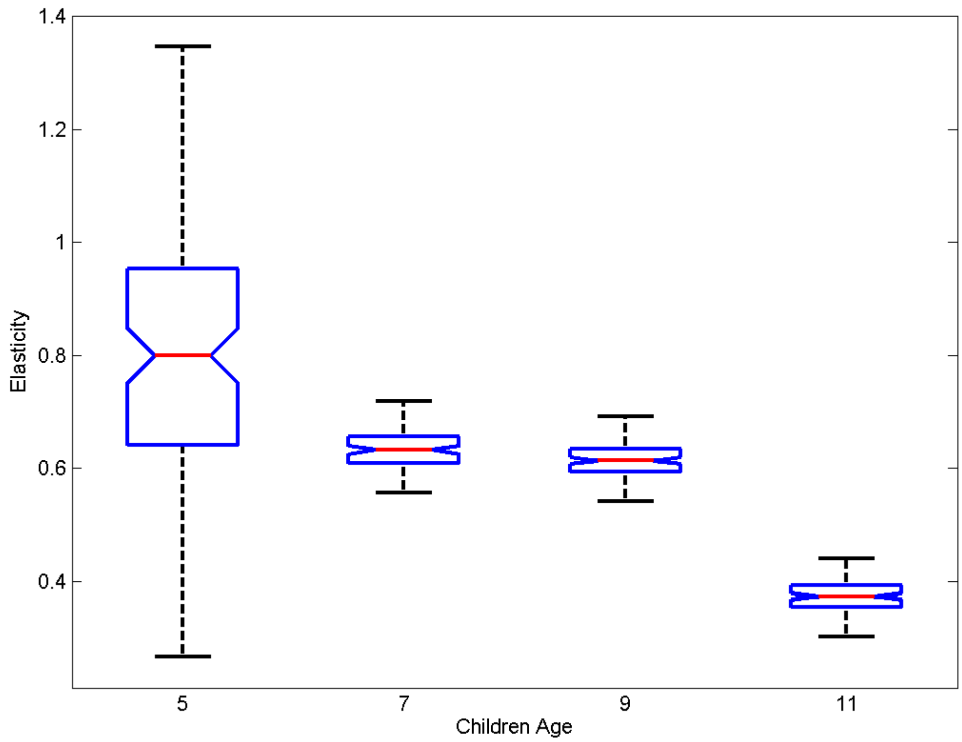
C.1 Additional Tables for Model 1 Corrected for Measurement Error

Table C.1-1: Estimates for Income Process

Constant	0.377
	(0.013)
Log Family Income t-1	0.753
	(0.008)
Variance Innovation	0.579
	(0.008)

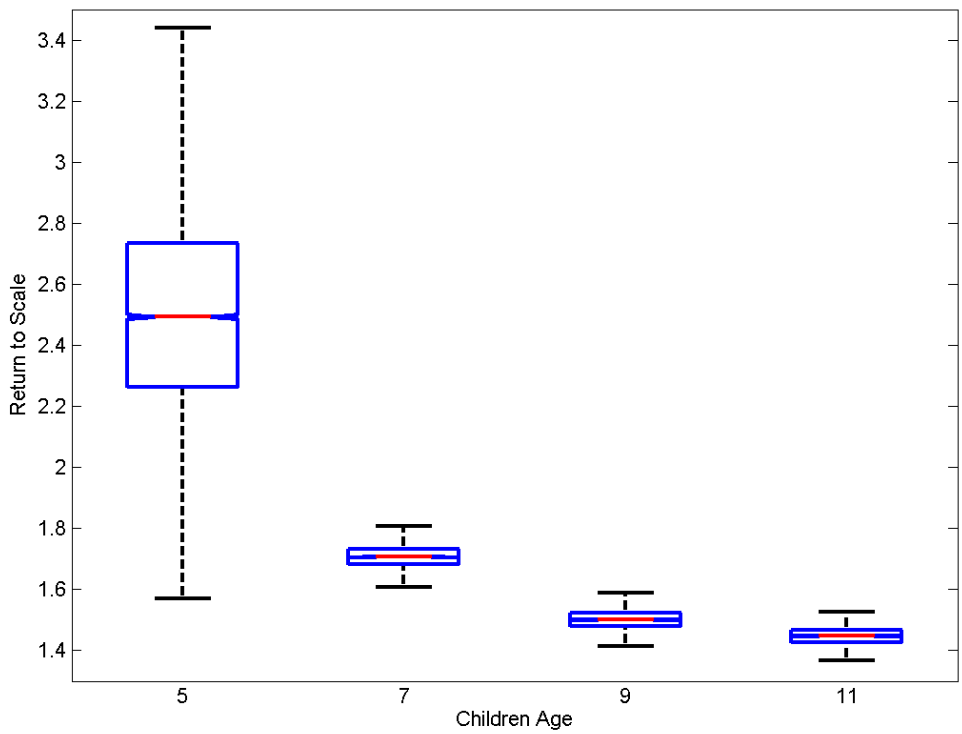
Notes: This table shows the estimates for the income process. The dependent variable is log family income at time t. Log Family Income $t-1$ is log family income two years prior. Standard errors in parenthesis are computed using a clustered bootstrap at the family level.

Figure C.1-1: Distribution of Elasticity of Next Period Skills with respect to Investment by Age



Notes: This figure shows the box plot for the elasticity of next period skills with respect to investment by different ages in the estimated Model 1 controlling for measurement error. The box plot is constructed as follow: the "central box" represents the central 50% of the data. Its lower and upper boundary lines are at the 25^{th} and 75^{th} quantile of the data. The central line indicates the median of the data while the two extreme lines (the top and the bottom ones) represents the 5^{th} and 95^{th} percentiles.

Figure C.1-2: Distribution of Technology Return to Scale by Age



Notes: This figure shows the box plot for the technology "return to scale" by different ages in the estimated Model 1 controlling for measurement error. The box plot is constructed as follow: the "central box" represents the central 50% of the data. Its lower and upper boundary lines are at the 25^{th} and 75^{th} quantile of the data. The central line indicates the median of the data while the two extreme lines (the top and the bottom ones) represents the 5^{th} and 95^{th} percentiles.

C.2 Descriptive Statistics

Table C.2-1: Children's Skills Measures

Measures	Range Values	Age Range	Scoring Order
(The Peabody Individual Achievement Test):			
Math	0-84	5-14	Positive
Recognition	0-84	5-14	Positive
Comprehensive	0-84	5-14	Positive

Notes: This table shows the features of children cognitive measures. The first column indicate each type of children skills measure we use to estimate our model. The second column shows the minimum and maximum value that each measure takes. The third column shows the minimum and maximum children age at which each measure is available. The last column indicates whether the measure is ordered positively (the higher score tend to reveal higher skills) or negatively (the lower score tend to reveal higher skills).

Table C.2-2: Mothers Cognitive Skills Measures

Measures	Range Values	Scoring Order
Arithmetics	0-30	Positive
Word Knowledge	0-35	Positive
Paragraph Composition	0-15	Positive
Numeric Operations	0-50	Positive
Coding Speed	0-84	Positive
Math Knowledge	0-25	Positive

Notes: This table shows the features of mother cognitive measures. The first column indicate each type of mother cognitive skills measure we use to estimate our model. The second column shows the minimum and maximum value that each measure takes. The last column indicates whether the measure is ordered positively (the higher score tend to reveal higher skills) or negatively (the lower score tend to reveal higher skills).

Table C.2-3: Mothers Noncognitive Skills Measures

Type of variables	Range Values	Label	Scoring Order
Mother Noncognitive Measures			
(Rosenberg indexes): I am a person of worth I have a number of good qualities I am able to do things as well as most other people I take a positive attitude toward myself	1-4	1= Strongly agree 2= Agree 3= Disagree 4= Strongly disagree	Negative
I am inclined to feel that I am a failure I felt I do not have much to be proud of I wish I could have more respect for myself I certainly feel useless at times At times I think I am no good at all	1-4	1= Strongly agree 2= Agree 3= Disagree 4= Strongly disagree	Positive
(Rotter Indexes): Rotter 1 (Life is in control or not)	1-4	1= In Control and closer to my opinion 2= In control but slightly closer to my opinion 3= Not in control but slightly closer to my opinion 4= Not in control and closer to my opinion	Negative
Rotter 2 (Plans work vs Matter of Luck)	1-4	1= Plans work and closer to my opinion 2= Plans work but slightly closer to my opinion 3= Matter of Luck but slightly closer to my opinion 4= Matter of Luck and closer to my opinion	Negative
Rotter 3 (Luck not a factor vs Flip a coin)	1-4	1= Luck not a factor and closer to my opinion 2= Luck not a factor but slightly closer to my opinion 3= Flip a coin but slightly closer to my opinion 4= Flip a coin and closer to my opinion	Negative
Rotter 4 (Luck big role vs Luck no role)	1-4	1= Luck big role and closer to my opinion 2= Luck big role but slightly closer to my opinion 3= Luck no role but slightly closer to my opinion 4= Luck no role and closer to my opinion	Positive

Notes: This table shows the features of mother noncognitive measures. The first column indicate each type of mother cognitive skills measure we use to estimate our model. The second column shows the minimum and maximum value that each measure takes. The third column shows the type of answers associated with each measure value. The last column indicates whether the measure is ordered positively (the higher score tend to reveal higher skills) or negatively (the lower score tend to reveal higher skills).

Table C.2-4: Descriptive Statistics about Children's Cognitive Skills Measures

Measures	Mean	Std	Min	Max	Number of Values
Age 5-6					
PIAT Math	11.858	4.278	0.000	37.000	32.000
PIAT Recognition	12.864	5.048	0.000	57.000	35.000
PIAT Comprehensive	12.770	4.930	0.000	49.000	35.000
Age 7-8					
PIAT Math	23.016	8.681	0.000	74.000	58.000
PIAT Recognition	25.748	8.774	0.000	80.000	67.000
PIAT Comprehensive	24.099	8.142	0.000	69.000	60.000
Age 9-10					
PIAT Math	38.720	10.832	0.000	84.000	71.000
PIAT Recognition	40.825	11.487	0.000	84.000	76.000
PIAT Comprehensive	37.540	10.231	0.000	78.000	64.000
Age 11-12					
PIAT Math	48.184	10.543	0.000	84.000	78.000
PIAT Recognition	51.079	13.278	0.000	84.000	74.000
PIAT Comprehensive	45.732	11.272	0.000	84.000	72.000
Age 13-14					
PIAT Math	53.767	11.387	0.000	84.000	78.000
PIAT Recognition	58.670	14.262	0.000	84.000	74.000
PIAT Comprehensive	51.015	12.229	0.000	84.000	74.000

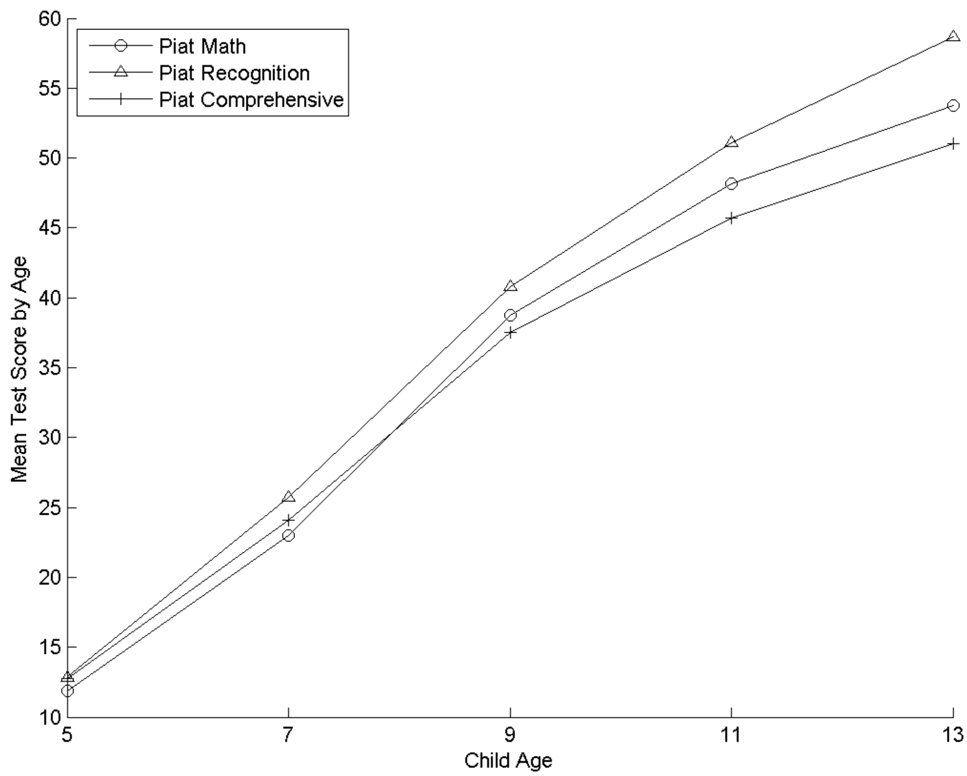
Notes: This table shows main sample statistics of children cognitive skills measures by children age.

Table C.2-5: Descriptive Statistics of Mother Cognitive and Noncognitive Skills Measures

Measures	Mean	Std	Min	Max	Number of Values
Mother Cognitive Skills					
Mom's Arithmetic Reasoning Test Score	13.946	6.603	0.000	30.000	31.000
Mom's Word Knowledge Test Score	21.773	8.562	0.000	35.000	36.000
Mom's Paragraph Composition Test Score	9.620	3.778	0.000	15.000	16.000
Mom's Numerical Operations Test Score	31.044	11.831	0.000	50.000	51.000
Mom's Coding Speed Test Score	42.953	17.468	0.000	84.000	85.000
Mom's Mathematical Knowledge Test Score	10.853	5.867	0.000	25.000	26.000
Mother Non Cognitive Skills					
Mom's Self-Esteem: "I am a person of worth"	2.461	0.549	0.000	3.000	4.000
Mom's Self-Esteem: " I have good qualities"	2.338	0.539	0.000	3.000	4.000
Mom's Self-Esteem: "I am a failure"	3.379	0.618	1.000	4.000	4.000
Mom's Self-Esteem: "I am as capable as others"	2.291	0.567	0.000	3.000	4.000
Mom's Self-Esteem: "I have nothing to be proud of"	3.360	0.669	1.000	4.000	4.000
Mom's Self-Esteem: "I have a positive attitude"	2.183	0.619	0.000	3.000	4.000
Mom's Self-Esteem: "I wish I had more self-respect"	2.796	0.817	1.000	4.000	4.000
Mom's Self-Esteem: "I feel useless at times"	2.650	0.770	1.000	4.000	4.000
Mom's Self-Esteem: "I sometimes think I am no good"	3.039	0.802	1.000	4.000	4.000
Mom's Rotter Score:"I have no control"	2.863	1.058	1.000	4.000	4.000
Mom's Rotter Score: "I make no plans for the future"	2.386	1.192	1.000	4.000	4.000
Mom's Rotter Score: "Luck is big factor in life"	3.205	0.856	1.000	4.000	4.000
Mom's Rotter Score: "Luck plays big role in my life"	2.594	1.024	1.000	4.000	4.000

Notes: This table shows main sample statistics of mother cognitive and non-cognitive skills measures.

Figure C.2-1: Descriptive Statistics: Mean of PIATs over the Childhood



Notes: This figure shows the mean Piat Math, Recognition and Comprehensive test scores by age. The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on.

C.3 Measurement Parameter Estimates

Table C.3-1: Measurement Parameter Estimates for Children's Cognitive Measures

Measures	μ	λ	Signal	Noise
Age 5-6				
PIAT Math	11.858	1.000	0.270	0.730
PIAT Recognition	12.864	2.238	0.972	0.028
PIAT Comprehensive	12.770	2.159	0.948	0.052
Age 7-8				
PIAT Math	11.858	1.000	0.757	0.243
PIAT Recognition	15.592	0.906	0.608	0.392
PIAT Comprehensive	15.014	0.802	0.554	0.446
Age 9-10				
PIAT Math	11.858	1.000	0.779	0.221
PIAT Recognition	10.297	1.136	0.894	0.106
PIAT Comprehensive	12.273	0.936	0.765	0.235
Age 11-12				
PIAT Math	11.858	1.000	0.803	0.197
PIAT Recognition	2.107	1.347	0.918	0.082
PIAT Comprehensive	6.129	1.089	0.833	0.167
Age 13-14				
PIAT Math	11.858	1.000	0.927	0.073
PIAT Recognition	8.556	1.195	0.845	0.155
PIAT Comprehensive	9.041	1.002	0.806	0.194

Notes: This table shows the measurement error parameters and associated statistics for children cognitive measures. The first two columns shows the measurement parameters (μ and λ) while the last two columns shows the signal and noise variance decomposition for the children cognitive measures.

Table C.3-2: Measurement Parameter Estimates for Mother Cognitive and Noncognitive Measures

Measures	μ	λ	Signal	Noise
Mother Cognitive Skills				
Mom‘s Arithmetic Reasoning Test Score	13.946	1.000	0.692	0.308
Mom‘s Word Knowledge Test Score	21.773	1.345	0.745	0.255
Mom‘s Paragraph Composition Test Score	9.620	0.584	0.722	0.278
Mom‘s Numerical Operations Test Score	31.044	1.720	0.638	0.362
Mom‘s Coding Speed Test Score	42.953	2.308	0.527	0.473
Mom‘s Mathematical Knowledge Test Score	10.853	0.854	0.639	0.361
Mother Non Cognitive Skills				
Mom‘s Self-Esteem: ”I am a person of worth”	2.461	1.000	0.152	0.848
Mom‘s Self-Esteem: ” I have good qualities”	2.338	1.263	0.252	0.748
Mom‘s Self-Esteem: ”I am a failure”	3.379	1.612	0.311	0.689
Mom‘s Self-Esteem: ”I am as capable as others”	2.291	1.127	0.181	0.819
Mom‘s Self-Esteem: ”I have nothing to be proud of”	3.360	1.746	0.312	0.688
Mom‘s Self-Esteem: ”I have a positive attitude”	2.183	1.474	0.260	0.740
Mom‘s Self-Esteem: ”I wish I had more self-respect”	2.796	2.080	0.297	0.703
Mom‘s Self-Esteem: ”I feel useless at times”	2.650	1.861	0.268	0.732
Mom‘s Self-Esteem: ”I sometimes think I am no good”	3.039	2.096	0.313	0.687
Mom‘s Rotter Score:”I have no control”	2.461	1.000	0.092	0.908
Mom‘s Rotter Score: ”I make no plans for the future”	2.338	1.263	0.140	0.860
Mom‘s Rotter Score: ”Luck is big factor in life”	3.379	1.612	0.118	0.882
Mom‘s Rotter Score: ”Luck plays big role in my life”	2.291	1.127	0.044	0.956

Notes: This table shows the measurement error parameters and associated statistics for mother cognitive and noncognitive measures. The first two columns show the measurement parameters (μ and λ) while the last two columns show the signal and noise variance decomposition for the mother measures.

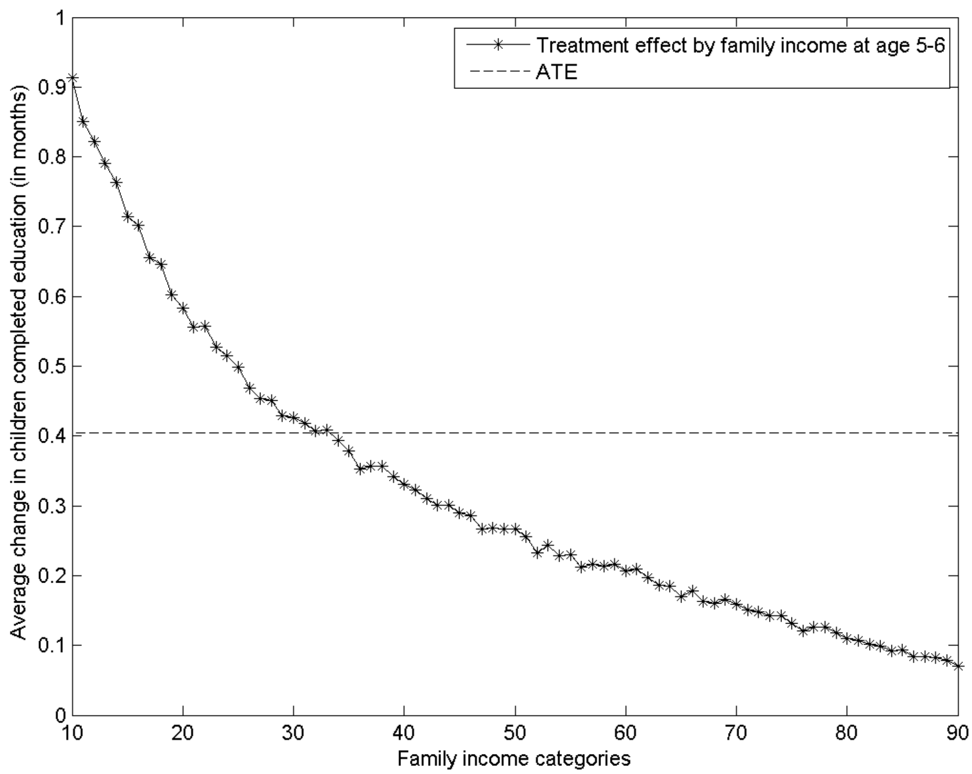
C.4 Estimates and Results for Model 2 with Measurement Error Corrected Estimator

Table C.4-1: Estimates for Investment (Model 2)

Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.230	0.069	0.068	0.065
	(0.059)	(0.021)	(0.029)	(0.030)
Log Mother Cognitive Skills	0.071	0.004	0.011	-0.005
	(0.022)	(0.009)	(0.014)	(0.012)
Log Mother Noncognitive Skills	0.359	0.711	0.660	0.678
	(0.131)	(0.059)	(0.084)	(0.084)
Log Family Income	0.341	0.217	0.261	0.262
	(0.076)	(0.054)	(0.072)	(0.082)
Variance Shocks	1.186	0.969	0.831	1.028
	(0.232)	(0.134)	(0.211)	(0.259)

Notes: This table shows the measurement error corrected estimates for the investment equation for Model 2. Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for parental investments and child's skill and family income at age 5-6 as well. Standard errors in parenthesis are computed using a clustered bootstrap at the family level.

Figure C.4-1: Heterogeneity in Policy Effects by Age 5 Household Income (Outcome: Schooling at Age 23)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 2, controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

C.5 Estimates and Results without Measurement Error Correction (Model 1 and Model 2)

Table C.5-1: Estimates for Investment (Model 1 and Model 2)

	Model 1 (Measurement Function Restrictions)				Model 2 (Production Function Restrictions)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.083	0.032	0.024	0.015	0.083	0.045	0.030	0.014
	(0.023)	(0.009)	(0.009)	(0.007)	(0.023)	(0.012)	(0.011)	(0.007)
Log Mother Cognitive Skills	0.082	0.010	0.010	-0.002	0.082	0.010	0.010	-0.002
	(0.019)	(0.011)	(0.014)	(0.011)	(0.019)	(0.011)	(0.014)	(0.011)
Log Mother Noncognitive Skills	0.248	0.454	0.442	0.553	0.248	0.448	0.440	0.553
	(0.093)	(0.073)	(0.098)	(0.074)	(0.093)	(0.073)	(0.098)	(0.074)
Log Family Income	0.587	0.504	0.524	0.434	0.587	0.498	0.521	0.435
	(0.074)	(0.070)	(0.095)	(0.077)	(0.074)	(0.069)	(0.095)	(0.078)
Variance Shocks	1.635	1.522	1.537	1.535	1.635	1.504	1.529	1.537
	(0.224)	(0.172)	(0.364)	(0.327)	(0.224)	(0.168)	(0.360)	(0.329)

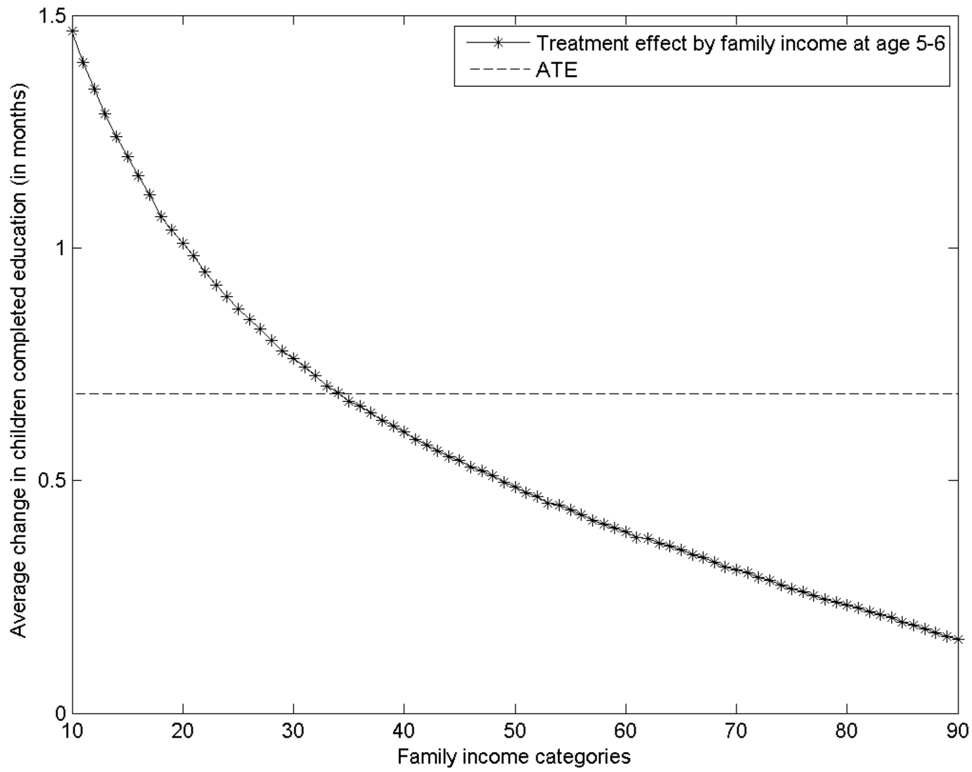
Notes: This table shows the estimates (not corrected for measurement error) for the investment equation for both Model 1 and Model 2. Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for parental investments and child's skill and family income at age 5-6 as well. Standard errors in parenthesis are computed using a clustered bootstrap at the family level.

Table C.5-2: Estimates for Skill Technology (Model 1 and Model 2)

	Model 1 (Measurement Function Restrictions)				Model 2 (Production Function Restrictions)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.875	0.771	0.669	0.770	0.625	0.868	0.897	0.880
	(0.057)	(0.022)	(0.017)	(0.018)	(0.047)	(0.039)	(0.039)	(0.052)
Log Investment	0.518	0.069	0.042	0.325	0.370	0.125	0.101	0.127
	(0.089)	(0.066)	(0.061)	(0.099)	(0.045)	(0.038)	(0.039)	(0.052)
(Log Skills *	0.006	0.007	0.002	-0.006	0.005	0.008	0.002	-0.007
Log Investment)	(0.012)	(0.003)	(0.002)	(0.002)	(0.009)	(0.004)	(0.002)	(0.003)
Return to scale	1.399	0.846	0.713	1.089	1.000	1.000	1.000	1.000
	(0.098)	(0.072)	(0.063)	(0.096)	(-)	(-)	(-)	(-)
Variance shocks	7.490	7.673	6.716	7.382	5.354	6.155	7.211	9.092
	(0.127)	(0.145)	(0.192)	(0.220)	(0.386)	(0.565)	(0.769)	(0.980)
Log TFP	12.789	18.491	18.477	14.011	0.000	0.000	0.000	0.000
	(0.215)	(0.299)	(0.444)	(0.690)	(-)	(-)	(-)	(-)

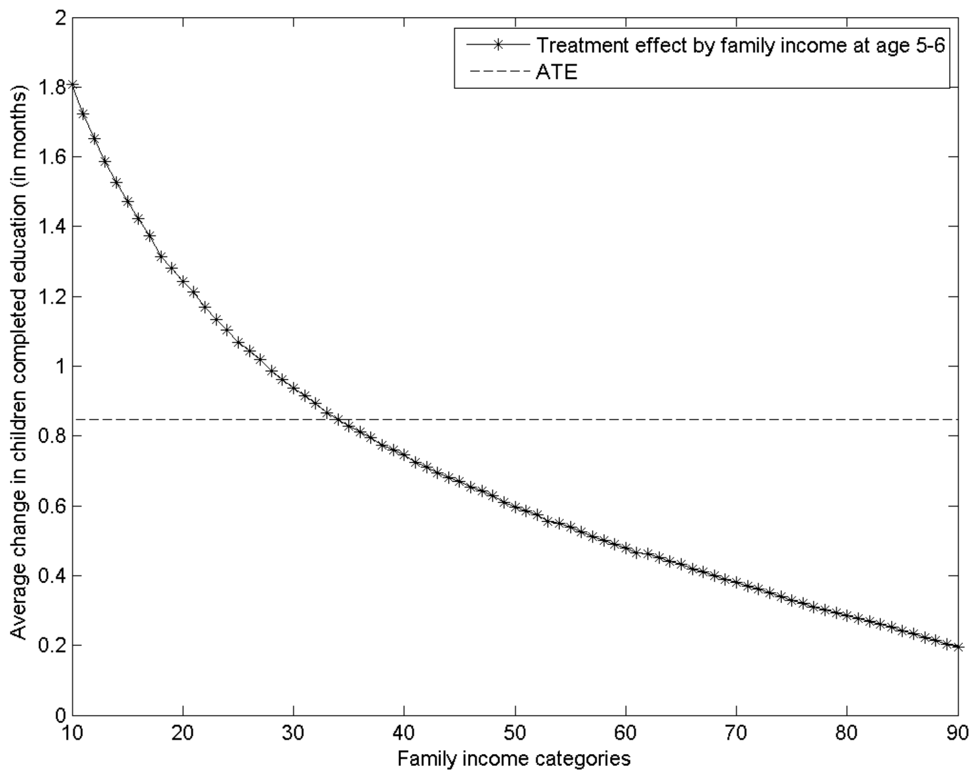
Notes: This table shows the estimates (not corrected for measurement error) for the technology of skills formation of both Model 1 and Model 2. Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t+1$ while the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8. Standard errors in parenthesis are computed using a clustered bootstrap at the family level.

Figure C.5-1: Heterogeneity in Policy Effects by Age 5 Household Income (Outcome: Schooling at Age 23, Model 1)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 1, not controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

Figure C.5-2: Heterogeneity in Policy Effects by Age 5 Household Income (Outcome: Schooling at Age 23, Model 2)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 2, not controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

Table C.5-3: Estimates for Skill Technology (Model 1 and Model 2) with CHS Sample

	Model 1 (Measurement Function Restrictions)				Model 2 (Production Function Restrictions)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	1.638	0.879	0.866	0.988	0.606	0.956	0.857	1.040
	(0.262)	(0.056)	(0.072)	(0.056)	(0.170)	(0.092)	(0.144)	(0.248)
	[1.17, 1.99]	[0.76, 0.94]	[0.73, 0.98]	[0.88, 1.08]	[0.35, 0.87]	[0.81, 1.09]	[0.64, 1.13]	[0.83, 1.52]
Log Investment	1.120	-0.141	-0.476	-0.602	0.415	0.026	0.117	-0.053
	(0.605)	(0.285)	(0.650)	(0.593)	(0.133)	(0.092)	(0.140)	(0.248)
	[0.40, 2.23]	[-0.67, 0.32]	[-1.26, 0.29]	[-1.44, 0.51]	[0.18, 0.62]	[-0.12, 0.17]	[-0.15, 0.35]	[-0.54, 0.17]
(Log Skills *	-0.057	0.014	0.025	0.012	-0.021	0.017	0.026	0.013
Log Investment)	(0.153)	(0.015)	(0.022)	(0.014)	(0.057)	(0.017)	(0.021)	(0.016)
	[-0.27, 0.24]	[-0.00, 0.04]	[-0.00, 0.06]	[-0.01, 0.03]	[-0.13, 0.07]	[-0.00, 0.05]	[-0.00, 0.05]	[-0.01, 0.03]
Return to scale	2.701	0.753	0.415	0.398	1.000	1.000	1.000	1.000
	(0.588)	(0.277)	(0.655)	(0.599)	(-)	(-)	(-)	(-)
	[2.04, 3.77]	[0.22, 1.18]	[-0.43, 1.22]	[-0.49, 1.47]	[-,-]	[-,-]	[-,-]	[-,-]
Variance shocks	5.459	3.684	3.536	3.624	2.021	1.453	1.312	1.386
	(0.303)	(0.332)	(0.359)	(0.365)	(0.413)	(0.321)	(0.362)	(0.628)
	[5.05, 6.04]	[3.27, 4.46]	[3.09, 4.29]	[3.33, 4.45]	[1.45, 2.81]	[1.06, 2.11]	[1.01, 2.09]	[0.97, 3.05]
Log TFP	14.057	17.928	12.825	7.214	0.000	0.000	0.000	0.000
	(0.665)	(0.726)	(1.916)	(1.891)	(-)	(-)	(-)	(-)
	[12.84,14.97]	[17.15,19.42]	[10.18,17.09]	[3.93,11.32]	[-,-]	[-,-]	[-,-]	[-,-]

Notes: This table shows the measurement error corrected estimates for the technology of skills formation for both Model 1 (Measurement Function Restriction: Age-Invariance) and Model 2 (Production Function Restriction: Known Location and Scale). Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t+1$, and the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level. We use the same estimating sample as in Cunha et al. (2010): firstborn white children.

Table C.5-4: Estimates for Skill Technology (Model 1 and Model 2) with Additional Controls

	Model 1 (Measurement Function Restrictions)				Model 2 (Production Function Restrictions)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	1.920	1.084	0.896	1.067	0.748	0.816	0.829	0.901
	(0.139)	(0.036)	(0.025)	(0.027)	(0.085)	(0.072)	(0.109)	(0.103)
	[1.67, 2.15]	[1.02, 1.15]	[0.85, 0.93]	[1.02, 1.11]	[0.60, 0.89]	[0.70, 0.94]	[0.71, 1.02]	[0.76, 1.08]
Log Investment	0.745	0.673	0.713	0.270	0.290	0.187	0.174	0.096
	(0.252)	(0.334)	(0.392)	(0.560)	(0.075)	(0.069)	(0.100)	(0.102)
	[0.39, 1.14]	[0.15, 1.28]	[-0.09, 1.18]	[-0.55, 1.25]	[0.17, 0.42]	[0.07, 0.31]	[-0.01, 0.29]	[-0.08, 0.25]
(Log Skills *	-0.098	-0.004	-0.003	0.003	-0.038	-0.004	-0.003	0.003
Log Investment)	(0.062)	(0.019)	(0.013)	(0.010)	(0.026)	(0.015)	(0.014)	(0.009)
	[-0.21,-0.03]	[-0.04, 0.03]	[-0.02, 0.02]	[-0.02, 0.02]	[-0.09,-0.01]	[-0.03, 0.02]	[-0.03, 0.02]	[-0.02, 0.01]
Return to scale	2.566	1.753	1.606	1.340	1.000	1.000	1.000	1.000
	(0.232)	(0.308)	(0.383)	(0.549)	(-)	(-)	(-)	(-)
	[2.22, 2.94]	[1.26, 2.33]	[0.80, 2.05]	[0.56, 2.30]	[-,-]	[-,-]	[-,-]	[-,-]
Variance shocks	5.612	4.519	3.585	4.019	2.187	1.328	0.977	0.923
	(0.174)	(0.184)	(0.181)	(0.247)	(0.195)	(0.154)	(0.171)	(0.175)
	[5.37, 5.93]	[4.27, 4.89]	[3.27, 3.88]	[3.70, 4.46]	[1.91, 2.53]	[1.12, 1.61]	[0.83, 1.33]	[0.75, 1.36]
Log TFP	13.420	15.060	12.105	3.133	0.000	0.000	0.000	0.000
	(0.304)	(0.433)	(0.570)	(0.947)	(-)	(-)	(-)	(-)
	[12.95,13.97]	[14.34,15.83]	[11.33,13.22]	[1.46, 4.72]	[-,-]	[-,-]	[-,-]	[-,-]

Notes: This table shows the measurement error corrected estimates for the technology of skills formation for both Model 1 (Measurement Function Restriction: Age-Invariance) and Model 2 (Production Function Restriction: Known Location and Scale) once we add additional controls in the investment equation in (14). In particular, we control for: a set of dummies for the maximum number of children ever observed in the household, a set of dummies for the mother's marital status, maternal hours worked, maternal hourly wage and a dummy for employment status (employed/non-employed). Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t+1$, and the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level.

Table C.5-5: Estimates for Investment (Model 1) with Additional Controls

Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.225	0.025	0.017	0.017
	(0.055)	(0.009)	(0.009)	(0.008)
	[0.13, 0.31]	[0.01, 0.04]	[0.01, 0.03]	[0.00, 0.03]
Log Mother Cognitive Skills	0.067	0.001	0.010	-0.006
	(0.024)	(0.009)	(0.016)	(0.012)
	[0.04, 0.12]	[-0.01, 0.02]	[-0.01, 0.04]	[-0.02, 0.02]
Log Mother Noncognitive Skills	0.353	0.707	0.667	0.697
	(0.147)	(0.070)	(0.100)	(0.088)
	[0.08, 0.55]	[0.54, 0.80]	[0.47, 0.80]	[0.54, 0.86]
Log Family Income	0.355	0.267	0.305	0.292
	(0.098)	(0.066)	(0.092)	(0.087)
	[0.23, 0.54]	[0.18, 0.42]	[0.18, 0.48]	[0.13, 0.45]
Variance Shocks	1.233	1.142	0.929	1.205
	(0.241)	(0.204)	(0.355)	(0.975)
	[0.99, 1.57]	[0.91, 1.49]	[0.69, 1.58]	[0.86, 1.93]

Notes: This table shows the measurement error corrected estimates for the investment equation for Model 1 (restricted measurement function) once we add additional controls in the investment equation. In particular, we control for: a set of dummies for the maximum number of children ever observed in the household, a set of dummies for the mother's marital status, maternal hours worked, maternal hourly wage and a dummy for employment status (employed/non-employed). Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for both contemporaneous parental investments and contemporaneous child's skill and contemporaneous family income. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a clustered bootstrap at the family level.

C.6 Skills measures in CNLSY79

Measures for Cognitive Skills

- **Peabody Picture Vocabulary Test**

  The Peabody Picture Vocabulary Test, revised edition (PPVT-R) ”measures an individual’s receptive (hearing) vocabulary for Standard American English and provides, at the same time, a quick estimate of verbal ability or scholastic aptitude” (see Dunn and Dunn, 1981). The PPVT was designed for use with individuals aged 3 to 40 years. The English language version of the PPVT-R consists of 175 vocabulary items of generally increasing difficulty. The child listens to a word uttered by the interviewer and then selects one of four pictures that best describes the word’s meaning. The PPVT-R has been administered, with some exceptions, to NLSY79 children between the ages of 3-18 years of age until 1994, when children 15 and older moved into the Young Adult survey. In the current survey round, the PPVT was administered to children aged 4-5 and 10-11 years of age, as well as to some children with no previous valid PPVT score.

  The first item, or starting point, is determined based on the child’s PPVT age. Starting at an age-specific level of difficulty is intended to reduce the number of items that are too easy or too difficult, in order to minimize boredom or frustration. The suggested starting points for each age can be found in the PPVT manual (see Dunn and Dunn, 1981).

  Testing begins with the starting point and proceeds forward until the child makes an incorrect response. If the child has made 8 or more correct responses before the first error, a “basal” is established. The basal is defined as the last item in the highest series of 8 consecutive correct answers. Once the basal is established, testing proceeds forwards, until the child makes six errors in eight consecutive items. If, however, the child gives an incorrect response before 8 consecutive correct answers have been made, testing proceeds backwards, beginning at the item just before the starting point, until 8 consecutive correct responses have been made. If a child does not make eight consecutive responses even after administering all of the items, he or she is given a basal of one. If a child has more than one series of 8 consecutive correct answers, the highest basal is used to compute the raw score.

  A “ceiling” is established when a child incorrectly identifies six of eight consecutive items. The ceiling is defined as the last item in the lowest series of eight consecutive items with six incorrect responses. If more than one ceiling is identified, the lowest ceiling is used to compute the raw score. The assessment is complete once both a basal and a ceiling have been established. The ceiling is set to 175 if the child never makes six errors in eight consecutive items.

A child's raw score is the number of correct answers below the ceiling. Note that all answers below the highest basal are counted as correct, even if the child answered some of these items incorrectly. The raw score can be calculated by subtracting the number of errors between the highest basal and lowest ceiling from the item number of the lowest ceiling.

- **The Peabody Individual Achievement Test (PIAT): Math**

  The PIAT Mathematics assessment protocol used in the field is described in the documentation for the Child Supplement (available on the Questionnaires page). This subscale measures a child's attainment in mathematics as taught in mainstream education. It consists of 84 multiple-choice items of increasing difficulty. It begins with such early skills as recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The child looks at each problem on an easel page and then chooses an answer by pointing to or naming one of four answer options.

  Administration of this assessment is relatively straightforward. Children enter the assessment at an age-appropriate item (although this is not essential to the scoring) and establish a "basal" by attaining five consecutive correct responses. If no basal is achieved then a basal of "1" is assigned (see PPVT). A "ceiling" is reached when five of seven items are answered incorrectly. The non-normalized raw score is equivalent to the ceiling item minus the number of incorrect responses between the basal and the ceiling scores.

- **The Peabody Individual Achievement Test (PIAT): Reading Recognition**

  The Peabody Individual Achievement Test (PIAT) Reading Recognition subtest, one of five in the PIAT series, measures word recognition and pronunciation ability, essential components of reading achievement. Children read a word silently, then say it aloud. PIAT Reading Recognition contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include matching letters, naming names, and reading single words aloud.

  The only difference in the implementation procedures between the PIAT Mathematics and PIAT Reading Recognition assessments is that the entry point into the Reading Recognition assessment is based on the child's score in the Mathematics assessment, although entering at the correct point is not essential to the scoring.

The scoring decisions and procedures are identical to those described for the PIAT Mathematics assessment.

- **The Peabody Individual Achievement Test (PIAT): Reading Comprehension**

  The Peabody Individual Achievement Test (PIAT) Reading Comprehension subtest measures a child's ability to derive meaning from sentences that are read silently. For each of 66 items of increasing difficulty, the child silently reads a sentence once and then selects one of four pictures that best portrays the meaning of the sentence.

  Children who score less than 19 on Reading Recognition are assigned their Reading Recognition score as their Reading Comprehension score. If they score at least 19 on the Reading Recognition assessment, their Reading Recognition score determines the entry point to Reading Comprehension. Entering at the correct location is, however, not essential to the scoring.

  Basals and ceilings on PIAT Reading Comprehension and an overall nonnormed raw score are determined in a manner identical to the other PIAT procedures. The only difference is that children for whom a basal could not be computed (but who otherwise completed the comprehension assessment) are automatically assigned a basal of 19. Administration instructions can be found in the assessment section of the Child Supplement.

D Alternative Measures

One of the characteristics of the data used to study child development is the rich variety skill measures. The previous sections considered identification where the skill measures are in a "raw" form: each measure is a linear function of the latent log skill. This measurement system, while commonly assumed in the prior literature, is in some respects a "best case."

In this section, we consider alternative forms of measures and re-examine whether we can identify the same types of production technologies using these alternative measures. We consider four classes of measures which are frequently encountered empirically: (i) *age-standardized* measures where the raw measures are transformed ex post to have mean 0 and standard deviation 1 for the sample at hand; (ii) *relative* measures where the measures reflect not the level of a child's skill but the child's skill relative to the population mean; (iii) *ordinal* measures which provide a discrete ranking of children's skills; and (iv) *censored* measures where the measures are truncated with a "floor" (finite minimum value) and/or a "ceiling" (finite maximum value). For each type of measure, we discuss which of our prior identification results still hold, if any, and what auxiliary assumptions would be sufficient to restore our identification results.

D.1 Age-Standardized Measures

Age-standardized measures are defined as the following transformation of raw measures $Z_{t,m}$:

$$Z^S_{t,m} = \frac{Z_{t,m} - E(Z_{t,m})}{V(Z_{t,m})^{1/2}}. \tag{D-1}$$

By construction, these measures are mean 0 and standard deviation 1 for all child ages.

Our identification results using standardized measures continues to hold if the technology of skill formation has known scale and location functions (KLS, Definition 1). To show this, we can re-write the standardized measures as a linear function of the latent variable:

$$Z^S_{t,m} = \mu^S_{t,m} + \lambda^S_{t,m} \ln\theta_t + \epsilon^S_{t,m}$$

where the measurement parameters and measurement error are

$$\mu^S_{t,m} = -\lambda^S_{t,m}(V(\ln\theta_t)) \cdot E(\ln\theta_t)$$

$$\lambda_{t,m}^{S} = \frac{\lambda_{t,m}}{V(Z_{t,m})^{1/2}} = \frac{\lambda_{t,m}}{(\lambda_{t,m}^{2} V(\ln \theta_t) + V(\epsilon_t))^{1/2}}$$

$$\epsilon_{t,m}^{S} = \frac{\epsilon_{t,m}}{V(Z_{t,m})^{1/2}} = \frac{\epsilon_{t,m}}{(\lambda_{t,m}^{2} V(\ln \theta_t) + V(\epsilon_t))^{1/2}}$$

These expressions show that the standardized measurement parameters are linear functions of the underlying moments of the latent skill distribution.

It is important to recognize that the use of standardized measures does not necessarily imply that any particular restriction on the underlying latent variables such as $E(\ln \theta_t) = 0$ or $V(\ln \theta_t) = 1$. The standardizations are necessarily in terms of the observed measures, not the unobserved latent variables. In addition, identification of the KLS production technologies is invariant to any increasing linear transformation of the original raw measures, say $Z'_{t,m} = a + bZ_{t,m}$ for $a \in R$ and $b \in R^{+}$.

One caveat deserves mention. Recall that because the initial conditions are normalized to a particular measure, using standardized rather than raw measures can affect the normalized location and scale of the latent skills, and in general affect the values of the production parameters which are identified up to the normalized initial period measure.

It is important to note that the use of age-standardized measures may not be cost free in the sense that age-standardized measures, which are constructed to be age-stationary in their first and second moments, contain no information about skill dynamics in these moments. For example, standardizing *age-invariant* measures, as defined in the previous section, so that the mean and variance of these measures is equal at all ages, would essentially "throw away" information regarding the average skill development of children across ages. This loss of information prevents the identification of the broader classes of technology of skills formation discussed above.

To see this point, recall that the identification of TFP or scaling parameter are based on additional information of the dynamics of measurement parameters. In the case of raw measures, those parameters are fully free parameters. On the other hand, when we use standardized measures, the new measurement parameters ($\mu_{t,m}^{S}$ and $\lambda_{t,m}^{S}$) are no longer free parameters but functions of the moments of the latent distribution. Hence, restricting the dynamics of the measurement parameters in this case is equivalent to restricting the dynamics of the latent skills, and can restrict the possible classes of technologies. While age-standardizing measures may provide some descriptive value, in the context of identifying dynamic production technologies, there is simply no point to transforming the measures in this way and throwing away potentially important identifying information.

D.2 Relative Measures

Some of the proxies used to measure children outcomes come from surveys where observers (often mothers, fathers, or other caregivers) provide assessments of the child. It can be plausible then that these observers are actually evaluating the child with respect to their perceptions of the average in the population. We call this type of measure a *relative measures*. In this case, these measures can be written as:

$$Z^R_{t,m} = \mu^R_{t,m} + \lambda^R_{t,m}(\ln \theta_t - E(\ln \theta_t)) + \epsilon^R_{t,m}. \tag{D-2}$$

where $(\ln \theta_t - E(\ln \theta_t)$ is the latent variable being measured by $Z^R_{t,m}$, which we model as the deviation of the actual level of the child's skill $\ln \theta_t$ relative to the mean value in the population $E(\ln \theta_t)$. Relative measures are not ordinal ranking measures (which we discuss below) but a continuous measure of skills relative to the population mean. As with the age-standardized measures, the relative measures are an increasing linear function of the underlying latent variable, and therefore our identification result continues to hold.

D.3 Ordinal Measures

We define ordinal measures the measures which are based on children rankings: this child has higher skills than another child. Let's assume that we observe in data children's skill rank. Let $Z_t = \{1, 2, \ldots, J\}$ be the child's human capital rank, with 1 highest level, and J lowest level. The observer (or us forming ranks from test scores) forms rank according to this ordinal model:

$$Z^O_{t,m} = \begin{cases} J \text{ if } \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} < \kappa_{J,t,m} \\ J-1 \text{ if } \kappa_{J,t,m} < \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} < \kappa_{J-1,t,m} \\ \vdots \\ 2 \text{ if } \kappa_{3,t,m} < \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} < \kappa_{2,t,m} \\ 1 \text{ if } \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} > \kappa_{2,t,m} \end{cases} \tag{D-3}$$

where the $\kappa_2, \ldots, \kappa_J$, with $\kappa_2 > \kappa_3, \ldots, \kappa_J$, are measurement parameters which provide the mapping from latent skills $\ln \theta_t$ and measurement error $\epsilon^O_{t,m}$ to the observed ordinal ranking values $Z^O_{t,m}$. The probability a child is ranked first $(j = 1)$ is then

$$\begin{aligned} pr(Z^O_{t,m} = 1) &= pr(\lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} > \kappa_{2,t,m}) \\ &= F_\epsilon(\lambda^O_{t,m} \ln \theta_t - \kappa_{2,t,m}) \end{aligned}$$

where F_ϵ is the distribution function for the measurement error $\epsilon^O_{t,m}$.

With ordinal ranking measures the non-parametric identification result no longer holds. There is no longer a one-to-one mapping between a child's latent skills θ_t and expected measures, as multiple values of θ_t are consistent with a child having a certain rank. Without additional assumptions, ordinal measures of skills do not allow non-parametric identification of the continuous skill production function.

If the researcher were to assume a particular known distribution for the measurement errors F_ϵ, then under this assumption for an ordinal measure of $t+1$ skills we would have:

$$F_\epsilon^{-1}(pr(Z^O_{t+1,m} = 1 | \ln\theta_t, \ln I_t)) = \lambda_{t+1,m} f_t(I_t, \theta_t) - \kappa_{2,t+1,m}$$

where $pr(Z^O_{t+1,m} = 1 | \ln\theta_t, \ln I_t)$ is the probability the child receives rank 1 at age $t+1$ given inputs θ_t, I_t at age t. This expression shows that with a known distribution for measurement errors, we can then apply our previous results to identify a KLS technology $f_t(I_t, \theta_t)$ up to this assumed distribution.

D.4 Censored Measures

Censored measures are defined as

$$Z^C_{t,m} = \begin{cases} \overline{Z} \text{ if } Z_{t,m} \geq \overline{Z} \\ Z_{t,m} \text{ if } \underline{Z} < Z_{t,m} < \overline{Z} \\ \underline{Z} \text{ if } Z_{t,m} < \underline{Z} \end{cases} \tag{D-4}$$

where $Z_{t,m} = \mu_{t,m} + \lambda_{t,m}\ln\theta_t + \epsilon_{t,m}$ is the "latent" measure, and $\overline{Z}$ ("ceiling") and $\underline{Z}$ ("floor"), with $\overline{Z} > \underline{Z}$, are the truncation points. Censoring occurs, for example, when a test score used as the measure has a maximum score (answering all questions correctly) and a minimum score (say answering none of the questions correctly). In practice, researchers can ascertain whether censoring is an important issue empirically by investigating what proportion of the sample actually has measured skills at the floor or ceiling points of the measure. Because censored measures do not have full support, our previous non-parametric identification results appear no longer to hold. As with the ordinal measures, auxiliary assumptions could be used to achieve identification up these additional assumptions (for a complete analyze of the problem, see Wang et al., 2009; Koedel and Betts, 2010).

E Monte Carlo Exercise for Model 1 and Measurement Error Correction

We implement a Monte Carlo exercise to examine the properties of our estimator. The true data generating process is assumed to be the estimated (measurement error corrected) Model 1 with some additional parametric assumptions about the measurement error process. In order to simulate the dataset, we use both the estimated measurement parameters and the joint distribution of children skills and investments. In addition, we assume that all the measurement noises are Normally distributed. We assume that the standard deviation of the error terms for all the skills measures are 0.5 (children and mothers) while we fix to 0.1 the standard deviation of the error terms for all the investment measures.

We generate a simulated longitudinal dataset of 10,000 children ranging from age 5-6 to age 13-14. In particular, the Monte Carlo analysis is performed estimating the model on 200 simulated data sets. In the following tables we show the mean estimates over the 200 estimates of the coefficients.

We focus only on estimates of skills technology, investment process and children's skills measurement parameters. Tables E-1-E-3 show true and mean estimated parameters. Overall, the estimator is able to recover the true parameters with minimal bias.

Table E-1: Monte Carlo Estimates for Investment Process

	True Parameters				Monte Carlo Estimates			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.230	0.027	0.020	0.018	0.249	0.026	0.020	0.018
Log Mother Cognitive Skills	0.071	0.004	0.012	-0.005	0.077	0.002	0.008	-0.011
Log Mother Noncognitive Skills	0.359	0.742	0.694	0.712	0.322	0.748	0.700	0.700
Log Family Income	0.341	0.227	0.274	0.275	0.352	0.224	0.272	0.292
Variance Shocks	1.186	1.019	0.868	1.087	1.263	0.993	0.827	1.103

Notes: This table shows both the true estimates (reported also in Table 3) and the mean Monte Carlo estimates for the investment equation. Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for parental investments and child's skill and family income at age 5-6 as well.

Table E-2: Monte Carlo Estimates for Skill Technology

Parameter	True Parameters Age 5-6	Age 7-8	Age 9-10	Age 11-12	Monte Carlo Estimates Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	1.966	1.086	0.897	1.065	1.955	1.091	0.897	1.071
Log Investment	0.799	0.695	0.713	0.252	0.759	0.700	0.839	0.502
(Log Skills * Log Investment)	-0.105	-0.005	-0.003	0.003	-0.092	-0.005	-0.005	-0.002
Return to scale	2.660	1.776	1.606	1.320	2.623	1.786	1.731	1.571
Variance shocks	5.612	4.519	3.585	4.019	5.613	4.520	3.586	4.018
Log TFP	13.067	14.747	11.881	2.927	13.060	14.689	11.801	2.594

Notes: This table shows both the true estimates (reported also in Table 4) and the mean Monte Carlo estimates for the technology of skills formation. Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t+1$ while the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8.

Table E-3: Monte Carlo Estimates for Measurement Parameters

Parameter	True Constant (μ) Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14	Monte Carlo Constant (μ) Estimates Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14
PIAT Math	11.858	11.858	11.858	11.858	11.858	11.858	11.858	11.858	11.858	11.858
PIAT Recognition	12.864	15.592	10.297	2.107	8.556	12.864	15.592	10.298	2.110	8.555
PIAT Comprehensive	12.770	15.014	12.273	6.129	9.041	12.770	15.013	12.270	6.132	9.040

Parameter	True Factor Loadings (λ) Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14	Monte Carlo Factor Loadings (λ) Estimates Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14
PIAT Math	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PIAT Recognition	2.238	0.906	1.136	1.347	1.195	2.238	0.905	1.136	1.347	1.196
PIAT Comprehensive	2.159	0.802	0.936	1.089	1.002	2.159	0.802	0.936	1.089	1.002

Notes: This table shows both the true estimates (reported also in Table C.3-1) and the mean Monte Carlo estimates for the measurement parameters of children skills measures equation. Each column shows the parameters at the given ages for each test score.