Estimating the Technology of Children's Skill Formation

Francesco Agostinelli
W.P. Carey School of Business,
Arizona State University

Matthew Wiswall *
W.P. Carey School of Business,
Arizona State University
University of Wisconsin-Madison
NBER

July, 2016

JEL Classification C38, J13

Keywords: Child Development; Skill Formation; Measurement Error; Latent Factor Models; Dynamic Factor Analysis; Production Technology

Online Appendix

*We thank Lance Lochner, Morteza Saharkhiz, Greg Veramendi, and participants at seminars at Arizona State, Rice, Brown, Minnesota, and Western Ontario for helpful comments. We are responsible for all errors.

Abstract

We develop a new estimator for the process of children's skill formation in which children's skills endogenously develop according to a dynamic latent factor structure. Rather than assuming skills are measured perfectly by a particular measure, we accommodate the variety of skills measures used in practice and allow latent skills to be measured with error using a system of arbitrarily located and scaled measures. For commonly estimated production technologies, which already have a known location and scale, we prove non-parametric identification of the primitive production function parameters. We treat the parameters of the measurement model as "nuisance" parameters and use transformations of moments of the measurement data to eliminate them, analogous to the data transformations used to eliminate fixed effects with panel data. We develop additional, 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 non-constant returns to scale.

We use our identification results to develop a sequential estimation algorithm for the joint dynamic process of investment and skill development, correcting for the biases due to measurement error in skills and investment. Using data for the United States, we estimate the technology of skill formation, the process of parental investments in children, and the adult distribution of completed schooling and earnings, allowing the production technology and investment process to freely vary as the child ages. Our estimates of high TFP and increasing returns to scale at early ages indicate that investments are particularly productive at these ages. 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. Our estimates of the dynamic process of investment and skill development allow us to estimate heterogeneous treatment effects of policy interventions. We show that even a modest transfer of family income to families at ages 5-6 would substantially increase children's skills, completed schooling, and adult earnings, with the effects largest for low income families.

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 severely bias our inferences.

In this paper, we develop a new method to estimate the skill formation process in children when skills are not observed directly but instead measured with error. Rather than assuming skills are measured perfectly by a particular measure, we accommodate the variety of skills measures used in practice and allow latent skills to be measured with error using a system of arbitrarily located and scaled skill measures. In our framework, we treat the parameters of the measurement model as "nuisance" parameters and use transformations of moments of the measurement data to eliminate them, analogous to the transformations used to eliminate fixed effects with panel data. We show non-parametric identification of the primitive parameters of the production technology, without assuming any particular values for the measurement process parameters or "re-normalizing" latent skills each period.

The heart of our identification analysis is a characterization of the classes of production technologies which can be identified given different assumptions about the measurement process. 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 (Cunha and Heckman, 2007; Cunha et al., 2010; Cunha and Heckman, 2008; Pavan, 2015). Starting with this class of technologies, we show that standard measurement error assumptions non-parametrically identify the primitive

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

production function parameters, up to a normalization on the initial conditions only. Importantly, identification is obtained without restrictions on the later skill measures as imposed in some previous papers, which can bias the production function estimates (see Agostinelli and Wiswall (2016) for a discussion).

Our identification analysis builds on previous work but offers a distinct approach to the empirical challenges. Previous approaches apply the 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) to the dynamic latent factor models describing the development of children's skills. In an influential paper applying latent factor modeling to child development, Cunha et al. (2010) identify the skill production technology by first "re-normalizing" the latent skill distribution at each period, treating the skills in each period as separate latent factors. While latent skills, which lack a meaningful location and scale, require some normalization (say at the initial period), repeated re-normalization every period is an unnecessary over-identifying restriction if the production function estimated already has a known location and scale, as is the case for the technology estimated by Cunha et al. (2010). We show that non-parametric identification of this class of KLS production functions is possible without these re-normalization restrictions, and our identification approach avoids imposing restrictions these restrictions because they can bias the estimation (Agostinelli and Wiswall (2016)).

In an important extension of our baseline results, we develop 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 non-constant returns to scale. 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 restricting the technology can reduce the permissible skill dynamics and productivity of investments, substantially changing our inferences about the child development process and our evaluation of policy. Our paper provides the first identification results for these more general models. Our analysis makes clear the key identification tradeoff researchers face: identification of restricted KLS technologies is possible with standard measurement assumptions, but identification of more general technologies requires stronger assumptions. We evaluate the empirical relevance of these additional assumptions, and provide guidance to researchers to evaluate whether the measures available to them satisfy these assumptions.

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; non-constant 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.

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 because it does not impose parametric distributional assumptions on the distribution of latent skills and measurement errors, as is commonly imposed in previous estimators. 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 production technology and investment process to freely vary as the child ages. Our estimates 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.

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 two sections, we develop the model of skill development and the measurement process. The next sections analyze the identification of this model, first under weak assumptions about the measurement process, and then under stronger assumptions about measurement which allows the identification of more general technology specifications, including those with TFP dynamics and non-constant returns to scale. The remainder of the paper develops our estimator and discusses our estimation results.

2 A Model of Skill Development in Children

In this section, we lay out our simple stylized model of skill development. In later sections, we develop a more detailed, and in many respects more general, empirical model which we take to the data.

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 stock of skills $\theta_{i,t}$, with $\theta_{i,t} > 0$ for all t and i, and a flow level of investments $I_{i,t}$, withe $I_{i,t} > 0$ for all i and t. For each child, the current stock of skills and current flow of investment produce next period's stock of skill according to the skill formation production technology:

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

where equation (1) can be viewed as dynamic state space model with $\theta_{i,t+1}$ the state variable for each child i. The production technology $f_t(\cdot)$ 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 the initial stock of child skills $\theta_{i,0}$ produces 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. We provide a more detailed analysis of policy interventions after the presentation of the full empirical model, but a few brief points are important to emphasize here. 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 as they are not reinforced by later investments? 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.

3 Measurement

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 scaled and imperfect measures of children's skills. Our framework 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.

3.1 Measurement Model

In our baseline case, we follow the literature and assume a commonly used (log) linear system of measures. In later sections,. we explore a variety of other measures and whether our identification results extend to these other types of measures. Each measure m for child i skills in period (age) t is given by

$$Z_{i,t,m} = \mu_{t,m} + \lambda_{t,m} \ln \theta_{i,t} + \epsilon_{i,t,m}, \tag{2}$$

For period t, we have $M_t \in \{1, 2, \ldots\}$ measures for each child i skills ($\ln \theta_{i,t}$): $m = 1, 2, \ldots, M_t$. $Z_{i,t,m}$ are the measures, $\mu_{t,m}$ are the measurement intercepts, and $\lambda_{t,m}$ are the measurement "factor loadings" or "scaling" parameters, with $\lambda_{t,m} > 0$ for all t and m. The $\mu_{t,m}$ and $\lambda_{t,m}$ measurement parameters allow the latent skills to be represented by arbitrarily located and scaled measures. Finally, $\epsilon_{i,t,m}$ are the individual measurement errors, with $E(\epsilon_{i,t,m}) = 0$ for all t, m (across children), which given the free intercept $\mu_{t,m}$, the assumption of mean zero $\epsilon_{t,m}$ errors is without loss of generality. To focus on the key identification issues, we assume investments I_t are observed without error. In the empirical model which we take to data, we allow for investments to also be measured with error and allow the investments to be endogenously determined by the existing skill stocks.

This measurement system has two important advantages over the alternative approach of using a single measure and assuming it perfectly measure skills, that is assuming $Z_{i,t,m} = \ln \theta_{i,t}$. First, the measurement system allows for noisy measures, in particular allowing measures to differ in their relative "noise" to "signal" ratio, $V(\epsilon_{i,t,m})/\lambda_{t,m}^2 V(\ln \theta_{i,t,m})$, thus allowing for the possibility that some measures have higher correlations to latent skills than others. Given this flexibility the researcher can then form estimators to take advantage of the greater signal some measures have available.

A second advantage is that the measurement parameters allow a kind of "arbitrariness" in the relationship between the measure and the latent skills. An ideal measurement system is one which can accommodate arbitrary changes in the location and scale of measures. Allowing the measures to have free measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$, which can vary by measure, allows the measurement model to capture the arbitrary location or scaling of particular measures.[2] We show below that the estimator of the primitive production function parameters we develop is robust to changes in the location and scale of the measures up to the initial normalization.

For the remainder of the paper, we omit the children's i subscript to reduce notational clutter. All expectations operations (E, Var, Cov, etc) are defined over the population of children (indexed i). For random variable $X_{i,t}$, we generically define $\kappa_t \equiv E(X_{i,t}) = \int X_{i,t} dF_t$, with F_t the distribution function for random variable $X_{i,t}$ in period t. For simplicity, we drop the i subscript and equivalently write this as $\kappa_t \equiv E(X_t)$.

3.2 Normalization

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 choice 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 latent skill stock θ_0 shares the scale of the normalizing measure in the sense that an 1 unit increase in log

[2]Measures, such as test scores, can be arbitrarily scaled and located in the sense that for any measure Z, we could create a new measures $Z' = a + bZ$, where a and $b > 0$ are some constants, and the new measure Z' therefore preserves at least the ordinal ranking of latent skills given by Z.

latent skills is equal to a 1 unit increase in the level of the normalized measure $Z_{0,1}$: $\frac{\partial Z_{0,1}}{\partial \ln \theta_0} = 1$, where, for intuition, we have treated the Z as a deterministic function. For symmetry with the latent skills, we also normalize log investment to be mean zero in the initial period $E(\ln I_0) = 0$.[3]

While the issue of model normalizations are typically trivial in most cases, in the case of dynamic models such as this, the type of assumed normalization is actually quite important. Our limited normalization for the initial period skills is quite different from the "re-normalization" approach used in much of the prior research (see Cunha and Heckman, 2007; Cunha et al., 2010; Attanasio et al., 2015a,b). In this approach, skills are re-normalized *every* period such that latent skills are assumed to be mean log stationery ($E(\ln \theta_t) = 0$ for all t) and latent skills "load" onto a different arbitrarily measure in each period ($\lambda_{t,1} = 1$ for all t). Agostinelli and Wiswall (2016) analyze the implications of the re-normalization approach and find that in many standard cases these assumptions are not necessary for point identification and can bias the estimates of the production technology.

We argue that our limited normalization is appropriate for the dynamic setting of child development we analyze. With our normalization for the initial period only, latent skills in *all* periods share a common location and scale with respect to the one chosen normalizing measure. This approach is analogous to deflating a nominal price series to a particular base year; that is, "normalizing" prices to some chosen base year (e.g. 2012 US Dollars).[4] As in the price normalization context, the choice of normalizing skill measure does affect the interpretation of the production function parameters, and we return to this issue when interpreting our particular estimates.

[3]In practice, if investments are truly observed without error, this can be accomplished by simply de-meaning the investment data so that the sample mean of $\ln I_0$ is zero. In the more general model we estimate, we assume investment is also observed with error and there are multiple measures of latent investments. For now, given we assume investment is observed, this normalization is merely for convenience.

[4]Given the normalizing measure we use is for young children, as children develop, their stock of skill may increase to the extent that the implied measure of skill using the initial normalizing measure $Z_{0,1}$ exceeds the sample maximum level of the measure. This is not an issue for the identification of the model since the measurement system assumes no floor or ceiling to the measures. The measures, and the normalization we use, fixes only the location and scale of the skills, but not the maximum or minimum values. We briefly discuss the issues of measurement floor and ceilings in the Appendix. For an example of an alternative measurement system which respects the discreteness, floor, and ceiling of a particular skill measure, see Del Boca et al. (2014a).

3.3 Ignoring Measurement Error

Before analyzing the identification of the model, it is helpful to motivate our analysis by briefly pausing to consider the consequences if we were to ignore measurement error. Consider a simple regression estimator in which we regress a measure of skills in period $t+1$ on a measure of skills in period t:

$$Z_{t+1,m} = \beta_0 + \beta_1 Z_{t,m} + \eta_{t,m}$$

The Ordinary Least Squares (OLS) estimand is

$$\beta_1(OLS) = \frac{Cov(Z_{t+1,m}, Z_{t,m})}{V(Z_{t,m})}$$

Assuming the measurement system above (2) and that the measurement errors $\epsilon_{t,m}$ are uncorrelated with latent skills $\ln\theta_t$ for all t, m and uncorrelated across time, we have

$$\beta_1(OLS) = \frac{\lambda_{t+1,m}\lambda_{t,m}Cov(\ln\theta_{t+1}, \ln\theta_t)}{\lambda_{t,m}^2 V(\ln\theta_t) + V(\epsilon_{t,m})}$$

This expression makes clear several problems in naively using observed measures to uncover latent production function relationships given by $Cov(\ln\theta_{t+1}, \ln\theta_t)$. First, the standard issue of attenuation bias: as the "noise" in the measure $V(\epsilon_{t,m})$ increases the OLS estimand goes to 0, biasing the inference of the relationship in latent skills given by $Cov(\ln\theta_{t+1}, \ln\theta_t)$. Second, the OLS estimand $\beta_1(OLS)$ is a combination of model primitives (production technology parameters) and measurement parameters, but we cannot directly separately identify them from the data. One common solution is simply to set $\lambda_{t+1,m} = 1$ and $\lambda_{t,m} = 1$ (a "single measure" approach). If this assumption is incorrect, then the resulting inference about latent production function relationships are biased.[5] The problem is even more severe if we consider regressions including higher order terms (with the goal of identifying some curvature or complementarities in the skill production process):

$$Z_{t+1,m} = \beta_0 + \beta_1 Z_{t,m} + \beta_2 Z_{t,m}^2 + \eta_{t,m}$$

[5] Other approaches include age standardizing the measures such that the measures have 0 mean and standard deviation 1 at each child age. However this approach does not imply $\lambda_{t+1,m} = 1$. Another approach is to re-normalize measures at all periods. This approach biases the resulting estimates. See Agostinelli and Wiswall (2016) for more discussion of these issues. Similar issues arise if we to examine conditional expectations, $E(Z_{t+1,m}|Z_{t,m})$, instead of covariances. In this case, the intercept of the measurement equations, $\mu_{t,m}$ and $\mu_{t+1,m}$, would also come into play.

where

$$Z_{t,m}^2 = (\mu_{t,m} + \lambda_{t,m} \ln \theta_t + \epsilon_{t,m})^2$$

In this case, $\lambda_{t,m}$ (factor loadings), $\mu_{t,m}$ (measurement intercepts), and in general the $\epsilon_{t,m}$ distribution need to be identified to uncover structural relationships between latent skills.

4 Identification

This section provides our main identification results. These identification results are constructive in the sense that they form the basis of our estimator of the skill development technology.

Our identification analysis proceeds in two steps. First, we identify the distribution of latent skills and investments in the initial period $G_0(\theta_0, I_0)$. 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. The second step of our identification analysis is to identify the production technology. This identification analysis is new.

We consider identification under the following assumptions about the joint distribution of latent skills ($\{\theta_t\}_t$), investments $\{I_t\}_t$, and measurement errors ($\{\epsilon_{t,m}\}_{t,m}$):

Assumption 1 *Measurement model assumptions:*
(i) $\epsilon_{t,m} \perp \epsilon_{t,m'}$ *for all* t *and* $m \neq m'$
(ii) $\epsilon_{t,m} \perp \epsilon_{t',m'}$ *for all* $t \neq t'$ *and all* m *and* m'
(iii) $\epsilon_{t,m} \perp I_{t'}$ *for all* t *and* t' *and all* m
(iv) $\epsilon_{t,m} \perp \theta_{t'}$ *for all* t *and* t' *and all* m

Assumption 1 (i) is that measurement errors are independent contemporaneously across measures. Assumption 1 (ii) is that measurement errors are independent over time. Assumption 1 (iii) and (iv) are that measurement errors in any period are independent of the latent stock of skills and parental investments in any period. While these assumptions are strong in some sense, they are common in the current literature.[6]

[6]Our assumption of full independence is sufficient, but not necessary, for at least some of our identification analysis. Below, we point out instances where weaker assumptions, allowing for some forms of dependence among measures and among measures and latent variable, can be used for identification.

4.1 Identification of Initial Conditions

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:

$$\lambda_{0,m} = \frac{Cov(Z_{0,m}, Z_{0,m'})}{Cov(Z_{0,1}, Z_{0,m'})}, \tag{3}$$

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

$$\mu_{0,m} = E(Z_{0,m}). \tag{4}$$

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

$$\widetilde{Z}_{0,m} = \frac{Z_{0,m} - \mu_{0,m}}{\lambda_{0,m}}, \tag{5}$$

where $\widetilde{Z}_{0,m}$ identifies the sum of the latent skill and a scaled version of the measurement error:

$$\widetilde{Z}_{0,m} = \ln \theta_0 + \frac{\epsilon_{0,m}}{\lambda_{0,m}}.$$

Applying the Kotalarski Theorem (Kotalarski 1964) to the $\{\widetilde{Z}_{0,m}\}_{m=1}^{M_0}$ residual measures, conditional on each level of investment I_0, we identify the distribution of θ_0 for any level of investment I_0. This then allows us to identify the joint distribution of latent skills and investment in the initial period $G_0(\theta_0, I_0)$, up to the normalizations given in Normalization 1.[7]

4.2 Identification of the Production Technology

With the initial distribution for latent skills and investments $G_0(\theta_0, I_0)$ identified in the first step, we next identify the process of child development given by the sequences of production technologies $f_0(\theta_0, I_0), \ldots, f_{T-1}(\theta_{T-1}, I_{T-1})$. Our identification analysis is sequential, and uses the production technology in period t, $f_t(\theta_t, I_t)$, to identify

[7] The key necessary condition for the Kotalarski theorem to hold in this case is that at least two of the residual measures in the set of measures $\{\widetilde{Z}_{0,m}\}_{m=1}^{M_0}$ have full support conditional on I_0, that is $\widetilde{Z}_{0,m} \in \mathbb{R}$ conditional on I_0.

the distribution of latent skills (and investments) in the next period, $G_{t+1}(\theta_{t+1}, I_{t+1})$. We first establish a general identification result for any periods t and $t+1$. We then conclude this section by describing the sequence of identification steps starting from the initial period $t = 0$.

4.2.1 From Measures to Latent Relationships

Given the generalities we have allowed, in which we do not assume that skills are measured perfectly in data, identification of the production technology now poses considerable challenges. The production technology in some period t would in principle be identified by the relationship between output θ_{t+1} and inputs θ_t, I_t. We do not directly observe latent skills θ_{t+1} or θ_t in data. Instead, we observe relationships among measures $Z_{t+1,m}$ and $Z_{t,m}$. Under Assumption 1, we have the following relationship between measures and latent variables:

$$E(Z_{t+1,m}|Z_{t,m}, \ln I_t) = \mu_{t+1,m} + \lambda_{t+1,m} E(\ln \theta_{t+1}|Z_{t,m}, \ln I_t)$$

This expression shows that $E(Z_{t+1,m}|Z_{t,m}, \ln I_t)$ does not identify a production function relationship directly, but instead a combination of latent skill relationships and measurement parameters.

In the following Lemma, we first show that we can identify dynamic production function relationships, $E(\ln \theta_{t+1}| \ln \theta_t, \ln I_t)$, from measures of latent skills in periods t and $t+1$, $Z_{t,m}$ and $Z_{t+1,m}$, up to the measurement parameters for the $t+1$ measure, $\mu_{t+1,m}$ and $\lambda_{t+1,m}$.[8]

Lemma 1 *Given i) $G_t(\theta_t, I_t)$ is known, ii) a pair of measures $Z_{t,m}$ and $Z_{t+1,m}$ which satisfy Assumption 1, and iii) measurement parameters for $Z_{t,m}$ ($\mu_{t,m}$, and $\lambda_{t,m}$) are known, $E(Z_{t+1,m}| \ln \theta_t = a, \ln I_t = \ell)$ is identified for some $(a,\ell) \in \mathbb{R}_2$ and is equal to $\mu_{t+1,m} + \lambda_{t+1,m} E(\ln \theta_{t+1}| \ln \theta_t = a, \ln I_t = \ell)$.*

Proof. See Appendix. ■

Lemma 1 establishes that while we cannot use realizations of measures $Z_{t,m} = z$ to identify particular values of the latent variable $\theta_t = p$, we can identify moments of the latent distribution.

[8] Note that the measure m for $t+1$, $Z_{t+1,m}$, which is used to measure $\ln \theta_{t+1}$, can be a completely different "kind" of measure from the measure $Z_{t,m}$ used to measure $\ln \theta_t$. The use of the same measure index m does not connote any relationship.

Substituting the production technology $\theta_{t+1} = f_t(\theta_t, I_t)$, Lemma 1 shows that measures $Z_{t+1,m}$ and $Z_{t,m}$ identify the following:

$$E(Z_{t+1,m} | \ln \theta_t = a, \ln I_t = \ell) = \mu_{t+1,m} + \lambda_{t+1,m} \ln f_t(e^a, e^\ell) \tag{6}$$

Note that the left-hand side of (6) is not directly observed in data (given the unobservability of $\ln \theta_t$) but is identified from observed measures (Lemma 1). The right-hand side of (6) is a combination of production function relationships and measurement parameters $\mu_{t+1,m}$ and $\lambda_{t+1,m}$.[9] If we do not know the measurement parameters $\mu_{t+1,m}$, $\lambda_{t+1,m}$, we cannot directly use $E(Z_{t+1} | \ln \theta_t, \ln I_t)$ to identify the production technology. One simple but problematic solution to this problem is to assume values for the $\mu_{t+1,m}$ and $\lambda_{t+1,m}$ parameters, and identification is trivially obtained.[10] However, if the assumptions on the measurement parameters are incorrect, then estimation under these assumptions can be biased.

4.2.2 Transformations to Eliminate Measurement Parameters

Our solution to this problem is to treat the measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$ for all $t > 0$ as "nuisance" parameters and use transformations of the moments (6) to eliminate them. Using four pairs of $(\ln \theta_t, \ln I_t) = \{(a_1, l_1), (a_2, l_2), (a_3, l_3), (a_4, l_4)\}$, we compute the following transformation of the conditional expectations:

$$\frac{E(Z_{t+1,m} | \ln \theta_t = a_1, \ln I_t = \ell_1) - E(Z_{t+1,m} | \ln \theta_t = a_2, \ln I_t = \ell_2)}{E(Z_{t+1,m} | \ln \theta_t = a_3, \ln I_t = \ell_3) - E(Z_{t+1,m} | \ln \theta_t = a_4, \ln I_t = \ell_4)} = \frac{\ln f_t(e^{a_1}, e^{\ell_1}) - \ln f_t(e^{a_2}, e^{\ell_2})}{\ln f_t(e^{a_3}, e^{\ell_3}) - \ln f_t(e^{a_4}, e^{\ell_4})} \tag{7}$$

where the values $a_k, \ell_k, k = 1, 2, 3, 4$ are such that $E(Z_{t+1,m} | \ln \theta_t = a_3, \ln I_t = \ell_3) \neq E(Z_{t+1,m} | \ln \theta_t = a_4, \ln I_t = \ell_4)$.

The left-hand side of (7) is a transformation of moments which are identified directly from the measures of skills for periods $t + 1$ and t (Lemma 1), and the

[9] Note that we have used the fact that in our stylized model, there are no stochastic elements to the production process, hence θ_{t+1} given θ_t, I_t is a constant for all t. We return to the topic of how to identify a shock to the production technology below. In brief, adding a mean 0 (log) shock to the production technology does not change the main identification analysis. Re-write the technology as $\theta_{t+1} = f_t(\theta_t, I_t) \exp(\eta_t)$, where $E(\eta_t) = 0$ and η_t is independent of θ_t, I_t, and $\epsilon_{t,m}, \epsilon_{t+1,m}$. Log skills are then $\ln \theta_{t+1} = \ln f_t(\theta_t, I_t) + \eta_t$, and mean log skills are $\ln f_t(\theta_t, I_t)$ as before.

[10] For example, the researcher could assume the values $\mu_{t,m} = 0$ and $\lambda_{t,m} = 1$ for all t and m, as in the case when all measures are assumed to be "classical" in the sense that $Z_{t,m} = \ln \theta_t + \epsilon_{t,m}$, and $\epsilon_{t,m}$ is simply a mean zero measurement error.

right-hand side is the corresponding transformation of the technology. The transformation in (7) has eliminated the measurement parameters $\mu_{t+1,m}$ and $\lambda_{t+1,m}$ without making any assumption about their values. This transformation is analogous to the transformation used in panel data analysis where differences at the observation level are used to eliminate common fixed effects.[11] As in the panel data literature, we exploit the particular form of the measurement equations and Assumption 1 to find an appropriate transformation to eliminate the nuisance measurement parameters. Other transformations can accomplish the same goal, and for convenience in some examples, we work with ratios of covariances, which already implicitly eliminate dependence on the measurement intercepts $\mu_{t,m}$.

4.2.3 Location and Scale of the Production Technology

Much of our analysis centers on the classes of production technologies which can be identified given that some inputs (latent skills) are measured with error. Crucial to our analysis is whether the production technology has a known location and scale or whether the location or scale is unknown in the sense that it depends on free parameters which need to be estimated. This concept is new to the production function identification literature, as far as we know. This concept is key to our analysis because our results below show that we can identify the production technologies up to location and scale, and can therefore point identify production technologies which already have a known location and scale.

We first define the concept of a production function with "known location and scale":

Definition 1 *A production function* $f_t(\theta_t, I_t)$ *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* ($\theta'_t \neq \theta''_t$ *or* $I'_t \neq I''_t$)*, the output* $f_t(\theta'_t, I'_t)$ *and* $f_t(\theta''_t, I''_t)$ *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 $f_t(\theta'_t, I'_t) - f_t(\theta''_t, I''_t)$ is known. Other points in the production possibilities set may be unknown, i.e. depend on free parameters to be estimated.

For example, consider the class of Constant Elasticity of Substitution (CES) skill production technologies, the class of technologies estimated in a number of previous

[11] Consider the model $y_{i,t} = \mu_i + X'_{i,t}\beta + \epsilon_{i,t}$, and the within transformation of the data $y_{i,t+1} - y_{it}$ eliminates the μ_i fixed effects.

studies (e.g.: Cunha et al., 2010).[12] The CES technology is

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{1/\phi_t}. \tag{8}$$

with $\gamma_t \in (0,1)$ and $\phi_t \in (-\infty, 1]$, and $\phi_t \to -\infty$ (Leontif), $\phi_t = 1$ (linear), $\phi_t \to 0$ (log-linear, Cobb-Douglas). The elasticity of substitution is $1/(1-\phi_t)$.

The production technology (8) satisfies Definition 1 because for inputs $I_t = \theta_t = \alpha > 0$, $\theta_{t+1} = \alpha$. That is, for inputs which are known to be equal at value α, we also know the output is α as well. This property of known location and scale is related to constant returns to scale property of this function, but constant returns to scale is not necessarily a sufficient property to satisfy Definition 1, as shown below. While the scale and location of the production function (8) are known, other points in the production possibilities set are determined by the free parameters γ_t and ϕ_t. Identifying these remaining parameters is the subject of the section.

Another example of KLS production technologies are those based on the translog function, a generalization of the Cobb-Douglas production technology which does not restrict the elasticity of substitution to be constant:

$$\ln \theta_{t+1} = \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} (\ln \theta_t)(\ln I_t) \tag{9}$$

with $\sum_{j=1}^{3} \gamma_{jt} = 1$. Consider the points $(\theta_t, I_t) = (1,1)$ and (e, e). For these points, the output of the production technology is known at $\ln \theta_{t+1} = 0$ and 1, respectively, and thus this function satisfies Definition 1.

In contrast, a class of technologies which does not satisfy the known location and scale property (Definition 1) is the following

$$\theta_{t+1} = A_t (\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{1/\phi_t} \tag{10}$$

with $A_t > 0$ representing Total Factor Productivity (TFP). The previous case (8) is a special case of (10) with $A_t = 1$. In the more general case, the addition of the unknown TFP process term A_t implies that the scale of the function is unknown. For example, for $\theta_t = I_t = \alpha > 0$, we have $f_t(\alpha, \alpha) = A_t \alpha$, where A_t is a free parameter. This class of technologies has constant returns to scale but does not have a known location and scale.[13]

[12] The functions estimated by Cunha et al. (2010) also include a mean zero production function shock. We consider identification of these functions below. In general, including a mean zero shock does not change the main results of the identification analysis, see Footnote 9.

[13] Similarly, a function where the factor share parameters did not sum to a known constant would also lack a known scale, for example $\theta_{t+1} = (\gamma_{1t} \theta_t^{\phi_t} + \gamma_{2t} I_t^{\phi_t})^{1/\phi_t}$, with $\gamma_{1t} + \gamma_{2t} \neq 1$. In this case, $\theta_t = \alpha$, $I_t = \alpha$, we have $f_t(\alpha, \alpha) = ((\gamma_{1t} + \gamma_{2t})\alpha)^{1/\psi_t}$.

Another class of technologies which does not satisfy Definition 1 is CES technologies without constant returns to scale:

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{\psi_t/\phi_t}. \tag{11}$$

where $\psi_t > 0$ is a returns to scale parameter, with $\psi_t = 1$ constant returns to scale, $\psi_t < 1$ decreasing returns to scale, and $\psi_t > 1$ increasing returns to scale. In this case, $f_t(\alpha, \alpha) = \alpha^{\psi_t}$. For this function, while we know the point $f_t(1,1) = 1$, and can identify the location of the function, we do not know a second point in the production possibilities set, and therefore cannot identify the scale of the function. Similarly, the translog function (9) with $\sum_{j=1}^{3} \gamma_{j,t}$ not equal to a known constant would not satisfy the KLS definition (Definition 1).

4.2.4 Per Period Identification of the Production Technology

We next proceed to the main identification result. We show that with the distribution of skills and investments in period t, $G_t(\theta_t, I_t)$ and the measurement parameters, $\mu_{t,m}$ and $\lambda_{t,m}$, known for period t, then a single measure of skills in period $t+1$, $Z_{t+1,m}$, with sufficient support, non-parametrically identifies a production technology $\theta_{t+1} = f_t(\theta_t, I_t)$ with known location and scale (satisfying Definition 1). The key aspect of the identification result is that we identify the production technology without knowledge of the period $t+1$ measurement parameters, $\mu_{t+1,m}$ and $\lambda_{t+1,m}$. We specify the exact conditions for identification in the following theorem.

Theorem 1 *If i) the distribution of skills $G_t(\theta_t, I_t)$ is known, ii) measurement parameters $\mu_{t,m}, \lambda_{t,m}$ are known, iii) there exists at least one measure $Z_{t+1,m}$ which satisfies Assumption 1, iv) the measure $Z_{t+1,m}$ has full support, $Z_{t+1,m} \in \mathbb{R}$, and v) the production technology $f_t(\theta_t, I_t)$ has known location and scale (Definition 1), then the production technology $f_t(\theta_t, I_t)$ is identified for all $(\theta_t, I_t) \in \mathbb{R}_+^2$.*

Proof. See Appendix. ■

Theorem 1 indicates that we can identify production technologies which have a known scale (Definition 1), such as the CES technologies (8) considered in much of the previous literature (see Cunha and Heckman, 2007; Cunha et al., 2010). The limitation of Theorem 1 is that we cannot apply it to more general production technologies which do not satisfy the known location and scale property. In the next section, we propose stronger assumptions on the measurement process which could allow for identification of more general production technologies.

4.2.5 Sequential Production Function Identification

Theorem 1 shows identification of the production technology for period t, $\theta_{t+1} = f_t(\theta_t, I_t)$, given measures $Z_{t+1,m}$ and $Z_{t,m}$. We now apply these per-period results to show how we can sequentially identify the full sequence of production technologies, $f_0(\theta_0, I_0), \ldots, f_{T-1}(\theta_{T-1}, I_{T-1})$, and hence the distribution of the sequence of skill stocks $(\theta_1, \ldots, \theta_T)$. The minimal data we require are at least 3 measures of latent skills for the initial period $Z_{0,1}, Z_{0,2}, Z_{0,3}$, and a single measure m of latent skills in the following periods, $Z_{1,m}, Z_{2,m}, \ldots, Z_{T,m}$.

The sequential identification proceeds as follows. First, using the measures for the initial period, following the analysis above, we identify the initial distribution of skills and investments $G_0(\theta_0, I_0)$, and initial measurement parameters, $\mu_{0,m}$ and $\lambda_{0,m}$, for some measure $Z_{0,m}$. Then, applying Theorem 1, we identify the production technology for period 0, $\theta_1 = f_0(\theta_0, I_0)$, where the technology is assumed to be of the known location and scale class (satisfying Definition 1). With the production function identified, we identify the distribution of period 1 skills from the production technology:

$$G_1(\theta_1|I_0) = pr(\theta_1 \leq \theta|I_1) = \int pr(f_0(\theta_0, I_0) \leq \theta) dG_0(\theta_0|I_0)$$

where $G_1(\theta_1|I_0)$ is the conditional distribution of latent skills, which given that investments are assumed observed, can then be used to identify the joint distribution of skills and investment.

We then proceed to identify the measurement parameters for measure $Z_{1,m}$ used to measure period 1 latent skills. The factor loadings can be identified from the across time correlation in measures of skills, $Cov(Z_{1,m}, Z_{0,m})$:

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} Cov(\ln\theta_1, \ln\theta_0)}.$$

where $Cov(\ln\theta_1, \ln\theta_0)$ is identified from the production technology and the initial distribution of skills:

$$\begin{aligned} Cov(\ln\theta_1, \ln\theta_0) &= Cov(\ln f_0(\theta_0, I_0), \ln\theta_0). \\ &= \int (\ln f_0(\theta_0, I_0) \ln\theta_0) dG(\theta_0, I_0) \end{aligned}$$

The measurement intercept for period 1 is then identified from

$$\mu_{1,m} = E(Z_{1,m}) - \lambda_{1,m} E(\ln\theta_1),$$

where as above $E(\ln\theta_1)$ is identified from the production technology, as above:

$$E(\ln\theta_1) = \int \ln f_0(\theta_0, I_0) dG(\theta_0, I_0).$$

This shows the identification of the technology $f_0(\theta_0, I_0)$ and the measurement parameters for $\mu_{1,m}$ and $\lambda_{1,m}$. We can continue to follow these steps, applying Lemma 1 and Theorem 1 sequentially, to identify the technology in the next periods, $f_1(\theta_1, I_1), \ldots, f_{T-1}(\theta_{T-1}, I_{T-1})$.

4.3 Intuition

Before we continue with examples and extensions to our identification concept, we pause to consider some simple intuition for our idea in a general setting. Consider a general production technology $Y = f(X_1, X_2)$, where Y is some latent unobserved output and X_1 and X_2 are some observed inputs. We have measure Z of the output, and the measure has error of the form we consider above: $Z = \mu + \lambda \ln Y + \epsilon$, where ϵ is uncorrelated with $\ln Y$, X_1, and X_2, and μ and λ are measurement parameters. The ratio of covariances of the measure of the output Z with the two inputs X_1, X_2 is

$$\begin{aligned}
\frac{Cov(Z, X_1)}{Cov(Z, X_2)} &= \frac{\lambda Cov(\ln Y, X_1)}{\lambda Cov(\ln Y, X_2)} \\
&= \frac{Cov(\ln Y, X_1)}{Cov(\ln Y, X_2)}.
\end{aligned}$$

The ratio of covariances has eliminated the "nuisance" measurement parameter λ. Working with covariances, rather than conditional expectations, has already eliminated dependence on the measurement intercept μ. This expression makes clear that even with output mis-measured in data and with free unknown measurement parameters allowing for arbitrary scale and location, we can still learn something about the production technology. For example, the ratio in this example is related to the relative marginal product of the two inputs X_1, X_2. Considering ratios of higher order covariances, such as $Cov(Z, X_1^2)/Cov(Z, X_1)$ and $Cov(Z, X_1 X_2)/Cov(Z, X_1)$, can similarly provide information about the "curvature" of the production function and the degree of complementarities between inputs. Our results above show identification in models which generalize this simple example, allowing for a dynamic production technology and mis-measured inputs as well.

4.4 Examples

We next proceed to demonstrate the identification results using simple two period models and commonly used production technologies.

Example 1 *Log-Linear (Cobb-Douglas) Technology*

There are two periods $T = 2$. Skills in $t = 1$ are given by the following log-linear (Cobb-Douglas) production technology:

$$\ln\theta_1 = \gamma_0 \ln\theta_0 + (1 - \gamma_0)\ln I_0 \tag{12}$$

where $\gamma_0 \in (0, 1)$ is the unknown production function parameter we would like to identify. Like the more general CES class to which it belongs, this production function has a known location and scale (Definition 1).

We have three measures of initial period skills: $Z_{0,1}, Z_{0,2}, Z_{0,3}$. We have one measure of skills in period 1, $Z_{1,m}$. The measures satisfy Assumption 1.

We normalize initial period skills as $E(\ln\theta_0) = 0$ and initial investments $E(\ln I_0) = 0$. We normalize the factor loading for the first measure as $\lambda_{0,1} = 1$. Following the analysis above, we then identify the remaining measurement factor loadings $\lambda_{0,2}, \lambda_{0,3}$ and measurement intercepts $\mu_{0,1}, \mu_{0,2}, \mu_{0,3}$ for the initial period measures. We then identify the joint distribution of the latent skills and investments, $G_0(\theta_0, I_0)$. Applying Lemma 1 identifies $E(Z_{1,m}|\ln\theta_0, \ln I_0)$ for values of $\ln\theta_0, \ln I_0$ from the measures $Z_{1,m}$ and $Z_{0,m}$ and the identified measurement parameters $\mu_{0,m}$ and $\lambda_{0,m}$.

Next we apply Theorem 1 to identify the production function parameter γ_0. The key to our analysis is that we identify the production function primitive without making any assumptions about the values of measurement parameters $\mu_{1,m}$ or $\lambda_{1,m}$. We compute the following transformations of conditional expectations (algebra is given in the Appendix):

$$\frac{E(Z_{1,m}|\ln\theta_0 = a, \ln I_0 = 0) - E(Z_{1,m}|\ln\theta_0 = 0, \ln I_0 = 0)}{E(Z_{1,m}|\ln\theta_0 = 1, \ln I_0 = 1) - E(Z_{1,m}|\ln\theta_0 = 0, \ln I_0 = 0)} = \frac{\gamma_0 a}{1}$$

Letting Δ be the left-hand side of the expression and solving for the production function parameter, we have

$$\gamma_0 = \frac{\Delta}{a}$$

This expression shows that the unknown parameter γ_0 of the production technology is identified from the transformation of the conditional expectations. Identification of γ_0 requires only a single measure of latent skills in period 1 and is invariant to the measurement parameters, $\mu_{1,m}$ and $\lambda_{1,m}$.

With the production technology $f_0(\theta_0, I_0)$ identified, we can now identify the measurement parameters for $Z_{1,m}$. $\lambda_{1,m}$ is identified from

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} Cov(\ln\theta_1, \ln\theta_0)}$$

where we can use any of the three first period measures, $m = 1, 2, 3$ to form the right-hand side. Substituting for the production technology, we have

$$\begin{aligned}\lambda_{1,m} &= \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} Cov(\gamma_0 \ln\theta_0 + (1-\gamma_0)\ln I_0, \ln\theta_0)} \\ &= \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m}(\gamma_0 V(\ln\theta_0) + (1-\gamma_0) Cov(\ln\theta_0, \ln I_0))}.\end{aligned}$$

Given the identification of γ_0, and that we have already identified the initial joint distribution of θ_0, I_0 (and can compute $V(\ln\theta_0)$ and $Cov(\ln\theta_0, \ln I_0)$), we can compute the right-hand side.

$\mu_{1,m}$ is then identified from

$$\mu_{1,m} = E(Z_{1,m}) - \lambda_{1,m} E(\ln\theta_1),$$

where, for this particular production technology, we have $E(\ln\theta_1) = \gamma_0 E(\ln\theta_0) + (1-\gamma_0)E(\ln I_0) = 0$, given the normalization for the initial period ($E(\ln\theta_0) = 0$ and ($E(\ln I_0) = 0$). As described in more detail below, the mean of log latent skills will in general not be 0 in periods after the initial period. For alternative production functions, $E(\ln\theta_1)$ can be computed from the identified production technology.

Example 2 *General CES Technology*

In our second example, we maintain the same setup as Example 1 but consider the general CES function (8):

$$\theta_1 = (\gamma_0 \theta_0^{\phi_0} + (1-\gamma_0) I_0^{\phi_0})^{1/\phi_0}.$$

with parameters defined as in (8). For this technology, there are two unknown production function parameters we wish to identify, γ_0 and ϕ_0. We have the same measures as in Example 1 and identify the initial condition as before.

As in Example 1, we compute the following:

$$\frac{E(Z_{1,m}|\ln\theta_0=\ln a_1,\ln I_0=\ln\ell_1)-E(Z_{1,m}|\ln\theta_0=\ln a_2,\ln I_0=\ln\ell_2)}{E(Z_{1,m}|\ln\theta_0=\ln a_3,\ln I_0=\ln\ell_3)-E(Z_{1,m}|\ln\theta_0=\ln a_4,\ln I_0=\ln\ell_4)}=\frac{\ln f_0(a_1,\ell_1)-\ln f_0(a_2,\ell_2)}{\ln f_0(a_3,\ell_3)-\ln f_0(a_4,\ell_4)}$$

Now define Δ_1 to be the left-hand side of the above equation and take values $a_1 \neq 0$, $a_3 \neq 0$, where $a_1 \neq a_3$, $a_2 = a_4 = \ell_2 = \ell_4 = 1$, $\ell_1 = 0$ and $a_3 = \ell_3 = e^1$. We have (see Appendix for omitted algebra):

$$\Delta_1 = \frac{\ln f_0(a_1,0)-\ln f_0(1,1)}{\ln f_0(e^1,e^1)-\ln f_0(1,1)},$$

$$\Delta_1 = \frac{\ln(\gamma_0 a_1)}{1},$$

Solving for γ_0 , we have

$$\gamma_0 = \frac{e^{\Delta_1}}{a_1}$$

This expression identifies γ_0. With γ_0 identified, we form a second ratio:

$$\Delta_2 = \frac{\ln f_0(a_1,1)-\ln f_0(1,1)}{\ln f_0(a_3,0)-\ln f_0(1,1)},$$

$$= \frac{\ln(\gamma_0 a_1^{\phi_0}+1-\gamma_0)}{\ln(\gamma_0 a_3)},$$

Solving for ϕ_0 (see Appendix for omitted algebra), we have

$$\phi_0 = \frac{\ln\left(\frac{(\gamma_0 a_3)^{\Delta_2}-1+\gamma_0}{\gamma_0}\right)}{\ln(a_1)}$$

This analysis shows that a single measure of period 1 skills identifies the unknown production function parameters γ_0, ϕ_0 without imposing any restrictions on the values of the period 1 measurement parameters. We can follow the same analysis as in Example 1 to identify the measurement parameters.

4.5 Comparison to Cunha et al. (2010)

Our results show identification of a production function with known location and scale without imposing any particular values for the measurement parameters after the initial period. Cunha et al. (2010) provide identification results in which they not only normalize initial period latent skills, as we do here, but also "re-normalize" latent skills each period. In our notation, their re-normalization restriction is $E(\ln \theta_0) = E(\ln \theta_1) = \cdots = E(\ln \theta_T) = 0$ and $\lambda_{0,1} = \lambda_{1,1} = \cdots = \lambda_{T,1} = 1$ for the normalized measure $m = 1$.

Some normalization is necessary (and we impose a normalization on the initial period), but, as we prove here, the additional restrictions on later periods are not necessary for identification of a known location and scale technology. The function Cunha et al. (2010) estimate is a known location and scale CES technology of the form given by (8). Because this function is already restricted (as compared to the more general functions with non-constant returns to scale and TFP dynamics), the additional normalizations are unnecessary and over-identifying. Importantly these re-normalization restrictions are not cost free as these additional normalizations can bias the technology estimates toward the Cobb-Douglas technology and away from more general patterns of substitution (see Agostinelli and Wiswall, 2016).

4.6 Errors-in-Variables Formulation

The KLS class of technologies can also be understood as a restriction in a traditional error-in-variables model (Chamberlain (1977a)). In this literature, identification is often achieved by proportionality restriction (linear regression parameters are assumed proportional to each other) within the context of a "reduced form" linear regression model. 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.

Consider the Cobb-Douglas case (12). Using the normalizations on the initial period, we proceed as before and form measures for the initial period:

$$\widetilde{Z}_{0,m} = \frac{Z_{0,m} - \mu_{0,m}}{\lambda_{0,m}} = \ln \theta_0 + \epsilon_{0,m}.$$

We also have a single measure of period 1 skills θ_1 given by

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m} \ln \theta_1 + \epsilon_{1,m}$$

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

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}[\gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0] + \epsilon_{1,m}$$

Substituting one of the measures for $\ln \theta_0$, say $\widetilde{Z}_{0,m}$, we have

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}[\gamma_0(\widetilde{Z}_{0,m} - \widetilde{\epsilon}_{0,m}) + (1 - \gamma_0) \ln I_0] + \epsilon_{1,m}$$

with $\widetilde{\epsilon}_{0,m} = \epsilon_{0,m}/\lambda_{0,m}$.

Re-arranging, we have

$$\begin{aligned} Z_{1,m} &= \mu_{1,m} + \lambda_{1,m}\gamma_0\widetilde{Z}_{0,m} + \lambda_{1,m}(1 - \gamma_0) \ln I_0 + (\epsilon_{1,m} - \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}) \\ &= \beta_0 + \beta_1\widetilde{Z}_{0,m} + \beta_2 \ln I_0 + \pi_{1,m} \end{aligned} \quad (13)$$

where $\beta_0 = \mu_{1,m}$, $\beta_1 = \lambda_{1,m}\gamma_0$, $\beta_2 = \lambda_{1,m}(1 - \gamma_0)$, and $\pi_{1,m} = \epsilon_{1,m} - \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}$. The "reduced form" equation (13) now has the standard errors-in-variables form: (13) is a linear regression of a measure of period 1 skills $Z_{1,m}$ on a measure for period 0 skills $\widetilde{Z}_{0,m}$. The β_1 and β_2 coefficients are combinations of the measurement factor loading $\lambda_{1,m}$ and the production function parameter γ_0.

Identification takes two steps. First, the standard error-in-variables problem is that the OLS regression estimands for β_1 and β_2 do not identify β_1 and β_2. We can solve this problem using any number of standard techniques. In this setting with multiple measures available satisfying independence assumptions, a second measure for period 0 skills, $\widetilde{Z}_{0,m'}$, can be used as an instrument for $\widetilde{Z}_{0,m}$, and we identify β_1 and β_2. Second, with β_1 and β_2 identified, we can then solve for the underlying primitive parameters γ_0 and $\lambda_{1,m}$:

$$\gamma_0 = \frac{\beta_1}{\beta_1 + \beta_2}\ ,\ \lambda_{1,m} = \beta_1 + \beta_2 \text{ and } \mu_{1,m} = \beta_0$$

The key to the identification here is that this commonly used production function (12) is already restricted (the factor shares sum to 1) and hence we can identify the production function parameters separately from the measurement parameters. Without this restriction on the production function, a function $\ln \theta_1 = \gamma_{0,\theta} \ln \theta_0 + \gamma_{0,I} \ln I_0$, where $\gamma_{0,\theta}$ and $\gamma_{0,1}$ are free parameters and do not sum to 1, point identification is not possible as there would be three unknown parameters $\gamma_{0,\theta}, \gamma_{0,I}$, and $\lambda_{1,m}$ and only two regression coefficients β_1, β_2.

4.7 Robustness to Alternative Types of Measures

One of the characteristics of the data used to study child development is the rich variety of skill measures. Here we 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 the Appendix, we briefly discuss 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 sometimes encountered empirically: (i) *age-standardized* measures where the raw measures are transformed ex post (in the sample) to have mean 0 and standard deviation 1; (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 (i.e. other children); (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 the age-standardized and relative measures, we find that our identification results continue to hold because these alternative measures can be expressed as alternative linear functions of the latent skills with particular measurement intercepts and factor loadings. Our identification results are invariant to these measurement parameters as the measurement parameters would be "transformed away," as described above (7). More generally, our identification results are robust to any linear increasing transformation of the original raw measures. On the other hand, without additional assumptions, the latter two classes of measures would appear to not allow non-parametric identification, at least globally, as these measures do not provide a one-to-one mapping between latent variables and measures (in expectation) as with the linear continuous measurement system we consider here.

5 Identification of General Technologies

The preceding analysis demonstrated that production functions with known location and scale (KLS, Definition 1) are non-parametrically identified using measures of latent skills that satisfy Assumption 1. This class of production technologies include the CES technologies analyzed in much of the previous work (see Cunha and Heckman, 2007; Cunha et al., 2010). These types of production functions are restricted, and these restrictions can affect our inferences about the child development process and the effects of policy interventions, as we demonstrate empirically below.[14] We next consider classes of technologies which are more general and no longer have a known location and scale, and we analyze identification of these more general technologies under additional assumptions about the measurement error process. We conclude this section with a discussion of what empirical measures may justify these additional assumptions.

5.1 Identifying Production Technologies with Dynamics in TFP

Consider a general class of technologies which exhibit Hicks-neutral TFP growth:

$$\theta_{t+1} = A_t \tilde{f}_t(\theta_t, I_t) \tag{14}$$

where $A_t > 0$ is the TFP term and the $\tilde{f}_t(\theta_t, I_t)$ sub-function is a known location and scale (KLS) production technology. An example of this class of functions is the CES production technologies augmented with TFP dynamics:

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

We first establish that our identification result for KLS production technologies fails in this case because we cannot separately identify the TFP parameter A_t from the measurement parameters. To see this, write the production technology in logs:

$$\ln \theta_{t+1} = \ln A_t + \ln \tilde{f}_t(\theta_t, I_t)$$

Note that A_t is the scale of the production technology in levels, but $\ln A_t$ is the location of the production function in logs.

Next consider the following difference in the conditional expectations for a latent log skill measure m, $Z_{t+1,m}$:

$$\begin{aligned} &E(Z_{t+1,m} | \ln \theta_t = a_1, \ln I_t = \ell_1) - E(Z_{t+1,m} | \ln \theta_t = a_2, \ln I_t = \ell_2) \\ = &\mu_{t+1,m} + \lambda_{t+1,m}(\ln A_t + \ln \tilde{f}_t(e^{a_1}, e^{\ell_1})) - [\mu_{t+1,m} + \lambda_{t+1,m}(\ln A_t + \ln \tilde{f}_t(e^{a_2}, e^{\ell_2}))] \\ = &\lambda_{t+1,m}(\ln \tilde{f}_t(e^{a_1}, e^{\ell_1}) - \ln \tilde{f}_t(e^{a_2}, e^{\ell_2})) \end{aligned}$$

[14]See the Appendix for more discussion. For example, a CES technology with constant returns to scale implies that the elasticity of skill formation with respect to investment must be between 0 and 1, regardless of the data. This restriction can then bias downward the effects on skill formation of investment and interventions which increase investment.

From this expression, it is clear that the TFP location $\ln A_t$ cannot be identified. Without further restrictions, the location of the production function (in logs) cannot be separately identified from the location of the measurement equations (which measure skills in logs) given by $\mu_{t,m}$ intercept.

Given the failure of identification for this more general technology, it is natural to ask what additional assumptions would be sufficient for identification. We show that if we have some auxiliary information on the relationship between measurement intercepts over time, then we can identify the A_t TFP terms in production functions of the form (14. We consider identification under the following assumption:

Assumption 2 *For some measures $Z_{t+1,m} = \mu_{t+1,m} + \lambda_{t+1,m} \ln \theta_{t+1} + \epsilon_{t+1,m}$ and $Z_{t,m'} = \mu_{t,m'} + \lambda_{t,m'} \ln \theta_t + \epsilon_{t,m'}$, we have $\mu_{t+1,m} = g(\mu_{t,m'})$, where $g(\cdot)$ is a known relationship.*

Whether Assumption 2 holds depends on the particular measures the researcher has available. We discuss the applicability of this assumption to our particular data and measures in our empirical application. This assumption could be justified if the measure in period $t+1$ and period t are *age-invariant* measures, as discussed below, where for example the measure is the same test given to children of different ages. In this case, it is plausible that the measures have the same location so that $\mu_{t+1,m} = \mu_{t,m}$. This assumption of age-invariant intercepts is of course sufficient but not necessary. And, to be clear, Assumption 2 does not require the researcher to assume any particular values for the measurement intercepts, but simply that they are related to each other in a known way.

We next present identification results which show that with Assumption 2, and the other assumptions previously used to prove Theorem 1, we can now identify skill development technologies of the form given in (14) which do not have a known location and scale:

Theorem 2 *Consider a production technology of the form $f_t(\theta_t, I_t) = A_t \tilde{f}_t(\theta_t, I_t)$ where $\tilde{f}_t(\theta_t, I_t)$ has known scale and location (Definition 1) and $A_t \in \mathbb{R}_{++}$. Under Assumption 1, Assumption 2, the full support assumption on some measure $Z_{t+1,m}$ and the conditions for Theorem 1, the technology $\tilde{f}_t(\theta_t, I_t)$ and A_t are separately identified.*

Proof.

The identification of $\tilde{f}_t(\theta_t, I_t)$ follows directly from Theorem 1, as the unknown A_t term is "differenced" away allowing identification of the $\tilde{f}_t(\theta_t, I_t)$ KLS sub-function. We identify the factor loading for the measure $\lambda_{t+1,m}$ as well because identification of this parameter does not depend on the A_t value. We then identify A_t from the mean of the measure of skills in period $t+1$, $E(Z_{t+1,m})$, and re-arranging for $\ln A_t$:

$$\ln A_t = \frac{E(Z_{t+1,m}) - (\mu_{t+1,m} + \lambda_{t+1,m} E(\ln \tilde{f}_t(\theta_t, I_t)))}{\lambda_{t+1,m}}$$

From Assumption 2, with $\mu_{t,m}$ known, we also identify the measurement intercept for $t+1$ from $\mu_{t+1,m} = g(\mu_{t,m})$.

■

From the proof we have some intuition for our result. As is common in the literature estimating TFP in a variety of contexts, TFP here is also identified by the *residual* growth in mean measured skills from period 0 to period 1 (scaled by $\lambda_{t+1,m}$ factor loading), netting out the growth due to period t inputs θ_t, I_t. Identification of the full sequence of production technologies then proceeds as above in a sequential fashion, and we identify the production function parameters, including the sequence of A_t TFP terms, for all periods. In the estimation sections below, we use this identification result constructively to develop an estimator for the TFP sequence.

5.2 Example

Next consider an example:

Example 3 *Log-Linear (Cobb-Douglas) Technology with TFP*

Return to the two period Cobb-Douglas example considered above (Example 1) but now add a scaling factor $A_0 > 0$:

$$\ln \theta_1 = \ln A_0 + (\gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0)$$

Assume the single period 1 measure $Z_{1,m}$ satisfies Assumption 2 and $\mu_{1,m} = g(\mu_{0,m})$ for some $m = 1, 2, 3$. We proceed as before to identify γ_0 from

$$\frac{E(Z_{1,m} | \ln \theta_0 = a, \ln I_0 = 0) - E(Z_{1,m} | \ln \theta_0 = 0, \ln I_0 = 0)}{E(Z_{1,m} | \ln \theta_0 = 1, \ln I_0 = 1) - E(Z_{1,m} | \ln \theta_0 = 0, \ln I_0 = 0)} = \frac{\ln A_0 + \gamma_0 a - \ln A_0}{\ln A_0 + 1 - \ln A_0}$$

where the $\ln A_0$ TFP terms drop out of the expression. As in Example 1, we can then solve for the γ_0 production function parameter.

The TFP term $\ln A_0$ is identified from

$$\ln A_0 = \frac{E(Z_{1,m}) - (\mu_{1,m} + \lambda_{1,m} E(\ln \tilde{f}_0(\theta_0, I_0)))}{\lambda_{1,m}}$$

$$= \frac{E(Z_{1,m}) - g(\mu_{0,m})}{\lambda_{1,m}}$$

because $E(\ln \tilde{f}_0(\theta_0, I_0)) = 0$ for this log-linear production function. $\ln A_0$ is identified from the growth in mean measured skills because $\mu_{0,m} = E(Z_{0,m})$ given the normalization of the initial conditions. Substituting, the TFP term is then

$$\ln A_0 = \frac{E(Z_{1,m}) - g(E(Z_{0,m}))}{\lambda_{1,m}}.$$

TFP is identified from the growth in mean skills between periods 0 and 1, scaled by the identified measurement factor loading for the period 1 measures, $\lambda_{1,m}$.

5.3 Identifying Production Technologies with Unknown Scale

We next consider a parallel problem to that of identifying the location (in logs) of the production technology considered above: identifying a production technology with an unknown scale. Consider the following production technology:

$$\theta_{t+1} = \widetilde{f}_t(\theta_t, I_t)^{\psi_t}, \tag{15}$$

where $\psi_t \in R^+$ is an unknown scaling parameter and $\widetilde{f}_t$ is a sub-function with known location and scale. Given the unknown scaling parameter, the technology described in (15) is not a known location and scale technology (Definition 1). An example of this type of production function is the following CES function with unknown scale:

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1 - \gamma_t) I_t^{\phi_t})^{\psi_t/\phi_t} \tag{16}$$

As in the TFP case above, we cannot separately identify the scale parameter ψ_t from the measurement factor loading $\lambda_{t+1,m}$. We consider an auxiliary restriction on the factor loadings which would allow identification:

Assumption 3 *For some measures* $Z_{t+1,m} = \mu_{t+1,m} + \lambda_{t+1,m} \ln \theta_{t+1} + \epsilon_{t+1,m}$ *and* $Z_{t,m'} = \mu_{t,m'} + \lambda_{t,m'} \ln \theta_t + \epsilon_{t,m'}$*, we have* $\lambda_{t+1,m} = q(\lambda_{t,m'})$*, where* $q(\cdot)$ *is a known function.*

We show that with Assumption 3, together with the other assumptions previously used to prove Theorem 1, we can now identify skill development technologies of the form given in (15), which do not have a known scale:

Theorem 3 *Consider a production technology of the form* $f_t(\theta_t, I_t) = \tilde{f}_t(\theta_t, I_t)^{\psi_t}$*, where* $\tilde{f}_t(\theta_t, I_t)$ *has known scale and location (Definition 1) and* $\psi_t \in \mathbb{R}_{++}$*. Under Assumption 1, Assumption 3, the full support assumption on some measure* $Z_{t+1,m}$ *and the conditions for Theorem 1, the technology* $\tilde{f}_t(\theta_t, I_t)$ *and* ψ_t *are separately identified.*

Proof.

Identification of $\tilde{f}_t(\theta_t, I_t)$ follows directly from Theorem 1, as the unknown ψ_t term drops out allowing identification of the $\tilde{f}_t(\theta_t, I_t)$ KLS sub-function. To identify ψ_t, take the covariance between a measure of latent skills at age $t+1$ and at age t:

$$\begin{aligned} Cov(Z_{t+1,m}, Z_{t,m}) &= \lambda_{t+1,m}\lambda_{t,m}Cov(\ln\theta_{t+1}, \ln\theta_t) \\ &= q(\lambda_{t,m})\lambda_{t,m}\psi_t Cov(\ln \tilde{f}_t(\theta_t, I_t), \ln\theta_t) \end{aligned}$$

Given $\lambda_{t,m}$ is known and $Cov(\ln \tilde{f}_t(\theta_t, I_t), \ln\theta_t)$ can be computed from the identified sub-function $\tilde{f}_t(\theta_t, I_t)$, then we can re-arrange this expression to solve for ψ_t.

■

This proof mirrors the identification result for TFP. If we assume that the factor loading in the measurement equation (which provides the scale of the measure) has some known relationship with already identified factor loadings, then we can identify the scale of the production technology.

5.4 Age-Invariant Measures

We conclude this section with a discussion of measures which would satisfy these auxiliary assumptions. An extensive literature, principally in the field of psychometrics, is concerned with designing skill measures which can be "equated" across children of different ages so that the development of children can be tracked using a coherent single measure. These measures consist of tests which are designed to be applicable for children of various ages, and include a range of test items (questions) which show meaningful variation for both younger and older children. Tests such as the Peabody Individual Achievement Test (PIAT)and the Woodcock-Johnson tests are designed so that they include a range of questions of various difficulty levels. The simple raw scores on these tests, reflecting the total number of questions answered correctly, can then be interpreted as an *age-invariant* measure of skills.[15]

[15]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

We formalize this notion of age-invariant measures in the following definition. A pair of measures are age-invariant if their measurement parameters are constant across child ages:

Definition 2 *A pair of measures $Z_{t,m}$ and $Z_{t+1,m}$ 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: $\theta_t = \theta_{t+1} = p$.[16] 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 for age-invariant measures both Assumption 2 and Assumption 3 hold, allowing identification of the technology with unknown TFP and unknown return to scale (see Theorem 3 (i)-(ii)).[17]

Finally note that whether a given pair of measures is age-invariant depends on the measures and must be evaluated on a case-by-case basis. Using pairs of unrelated measures, such as birth weight to measure cognitive skills at birth and SAT scores to measures skills at age 18, would not seem to constitute a set of age-invariant measures as there is no reason to believe these measures would have a common location and scale.

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).

[16] 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 (See the Appendix). Our concept of age-invariant measures concerns the underlying primitive and unobserved 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.

[17] 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 $\lambda_t = \lambda_{t+1}$.

6 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 sections, and briefly describe the data. Additional details about the data and sample are left for the Appendix.

6.1 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 tradeoff 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.

6.1.1 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 specification:

$$\ln \theta_{t+1} = \ln A_t + \gamma_{1,t} \ln \theta_t + \gamma_{2,t} \ln I_t + \gamma_{3,t} \ln I_t \cdot \ln \theta_t + \eta_{\theta,t}, \tag{17}$$

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 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 can be expanded with the inclusion of additional terms to a close provide an approximation of any unknown production technology. The log-linear form of the function also facilitates convenient and fast closed form estimators, as detailed below. Our general translog function also allows non-constant returns to scale. 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) affect the productivity of current investments. Moreover, $\gamma_{3,t} \neq 0$ implies that the elasticity of next period skills with respect to investment is a function of the child's stock of skills.

6.1.2 Parental Investment

We specify a parametric policy function for parental investment. Investment is endogenously determined by the current stock of the child's skills, mother's skills, and family income:

$$\ln I_t = \alpha_{1,t} \ln \theta_t + \alpha_{2,t} \ln \theta_{MC} + \alpha_{3,t} \ln \theta_{MN} + \alpha_{4,t} \ln Y_t + \eta_{I,t} \tag{18}$$

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 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 a kind of "reduced form" specification representing a policy function for parental investment 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, this approach 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 work derives parental investment from explicit models of the household, including explicit representations of household preferences, decision making, beliefs, and constraints (see for example Del Boca et al., 2014a,b; Cunha, 2013; Cunha et al., 2013; Bernal, 2008). 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, the interpretation of the parameters is in some sense speculative. $\alpha_{1,t}$ can be interpreted as reflecting whether parents "reinforce" existing skill stocks ($\alpha_{1,t} > 0$) or "compensate" for low skill stocks ($\alpha_{1,t} < 0$). $\alpha_{2,t}$ 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, $\alpha_{4,t}$ reflects the influences that household resources have on the extent of parental investments, and reflects the combined effects of constraints the household faces (such as credit market constraints) and preferences the household has to invest 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:

$$\ln Y_{t+1} = \mu_Y + \delta_Y \cdot t + \rho_Y \ln Y_t + \eta_{Y,t} \tag{19}$$

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.

6.1.3 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). 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 the family log income when children are 5-6 years old. The means of the remaining variables are set to zero by Normalization 1. Σ_Ω is the variance-covariance matrix for the initial conditions.

6.1.4 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):

$$Q = \mu_Q + \alpha_Q \ln \theta_T + \eta_Q, \tag{20}$$

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.

6.1.5 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: $\omega \in \{\theta, \theta_{MC}, \theta_{MN}, I\}$. There are in general multiple measures for each latent variable. 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 $\omega \in \{\theta, \theta_{MC}, \theta_{MN}, I\}$.

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 $(\eta_{I,t}, \eta_{\theta,t}, \eta_Q)$. This assumptions is strong, and weaker assumptions of mean-independence are sufficient for identification of the parametric model. While we assume strong independence assumption, 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. Our sequential estimator, described below, is therefore robust to mis-specification of the marginal distributions of measurement errors.

6.2 Estimation Algorithm

Our estimation algorithm is formed from the identification results presented above, and in particular relies on the error-in-variable formulation from Section 4.6. Before describing the steps of the algorithm, consider several estimation options. One approach, a kind of "brute force" approach, is to simulate the full sequence of latent variables and measures from candidate primitive parameters and explicit assumptions about the distribution of measurement errors (e.g. assume they are Normally distributed) and 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. This approach may also involve a tremendous amount of computationally costly simulation given the non-linear nature of the model.

A second estimation approach is to use the measures directly to simulate the distribution of latent 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 is the approach of Cunha et al. (2010) and Attanasio et al. (2015a,b) in which both assume the latent variables are distributed according to a mixture of 2 Normal distributions. This approach too makes specific parametric assumptions which are not required.

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.

We present two versions of the estimation algorithm. The first version works with a unrestricted version of the technology:

Model 1 (General): $\ln A_t$ free and $\sum_{j=1}^{3} \gamma_{j,t}$ free. At least one measure is age-invariant.

The availability of this age-invariant measure allows us to identify the more general technology.

The second version of the model restricts the production technology (17) to have a known location and scale.

Model 2 (Restricted): $\ln A_t = 0$ for all t (no TFP dynamics) and $\sum_{j=1}^{3} \gamma_{j,t} = 1$ for all t (constant returns to scale).

6.2.1 Estimation of Model 2 (Restricted)

We begin with the estimator for the second version of the model, using the restricted technology. The estimator for the more general technology (Model 2) is below.

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 means as outlined above (3) and (4). 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

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

[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 (18).

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 (19) using a regression of income on lagged income and a time 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 (18):

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

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

Re-arranging, we have

$$\begin{aligned} Z_{I,0,m} \;&=\; \mu_{I,0,m} + \lambda_{I,0,m}\alpha_{1,0}\widetilde{Z}_{\theta,0,m} + \lambda_{I,0,m}\alpha_{2,0}\widetilde{Z}_{MC,m} + \lambda_{I,0,m}\alpha_{3,0}\widetilde{Z}_{MN,m} + \lambda_{I,0,m}\alpha_{4,0} \ln Y_0 \\ &+\; \epsilon_{I,0,m} + \lambda_{I,0,m}(\eta_{I,0} - \widetilde{\epsilon}_{\theta,0,m} - \widetilde{\epsilon}_{MC,m} - \widetilde{\epsilon}_{MN,m}) \\ &=\; \beta_{0,0,m} + \beta_{1,0,m}\widetilde{Z}_{\theta,0,m} + \beta_{2,0,m}\widetilde{Z}_{MC,m} + \beta_{3,0,m}\widetilde{Z}_{MN,m} + \beta_{4,0,m} \ln Y_0 + \pi_{I,0,m} \qquad (21) \end{aligned}$$

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

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

Estimation of (21) 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 $\pi_{I,0,m}$). 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 Assumption 1, these instruments are valid because each of these alternative measures is uncorrelated with all of the components of $\pi_{I,0,m}$. 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

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

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

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

and

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

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

$$\widetilde{Z}_{I,0,m} = \frac{Z_{I,0,m} - \mu_{I,0,m}}{\lambda_{I,0,m}}$$

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 (17), 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:

$$Z_{\theta,1,m} = \delta_{0,0,m} + \delta_{1,0,m}\widetilde{Z}_{\theta,0,m} + \delta_{2,0,m}\widetilde{Z}_{I,0,m} + \delta_{3,0,m}\widetilde{Z}_{\theta,0,m} \cdot \widetilde{Z}_{I,0,m} + \pi_{\theta,0,m} \tag{22}$$

where $\delta_{0,0,m} = \mu_{\theta,0,m}$, $\delta_{j,0,m} = \lambda_{\theta,1,m}\gamma_{j,0}$ for $j = 1, 2, 3$ and

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

As with the investment function, estimation of 22 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 Assumption 1 these instruments are uncorrelated the residual error term $\pi_{\theta,0,m}$.[19] With consistent estimates of $\delta_{j,0,m}$s 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

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

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

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

We then form the residual measure for latent skill as

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

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

The remaining parameters to be estimated for this period are the variances of the investment and production function shocks, $\sigma^2_{I,0}$ and $\sigma^2_{\theta,0}$. To estimate $\sigma_{I,0,}$ we use the covariance between the residual from (21), $\pi_{I,0,m}$ and an alternative residual measure of investment $\widetilde{Z}_{I,0,m'} = \ln I_0 + \epsilon_{I,0,m'}$:

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

To compute the residual measure $\widetilde{Z}_{I,0,m}$ we need to compute the measurement parameters for this measure. We do this by repeating the estimation in Steps 2 and 3 replacing the left-hand side variable in (21) 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 $t = 1$:

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

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 of the investment and production functions for all child ages.

6.2.2 Estimation of Model 1 (Unrestricted)

The preceding algorithm restricted the production technology to have no TFP dynamics and constant returns to scale (Model 2). Following Theorem 2 and Theorem 3, identification of the more general model can be accomplished with restrictions on the measurement parameters. We assume we have available at least one child skill measure which is age-invariant (Definition 2). 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 (22) and estimation of the $\delta_{j,0,m}$ parameters remains the same. To allow for non-constant returns to scale we do not restrict the structural $\gamma_{j,0}$ parameters to sum to 1. The structural parameters are computed as

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

With the inclusion of the TFP term $\ln A_0$, the $\delta_{0,0,m}$ intercept from the reduced form equation (22) is now

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

Given the age-invariance assumption, we can consistently estimate $\mu_{\theta,1,m}$ and $\lambda_{\theta,1,m}$ and compute $\ln A_0$.

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

6.2.3 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 (20). We focus on both final years of education at age 23 and log earnings at age 30. 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 (20), we have:

$$Q = \mu_Q + \alpha_Q \widetilde{Z}_{\theta,4,m} + (\eta_Q - \alpha_Q \widetilde{\epsilon}_{\theta,4,m}) \tag{23}$$

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

6.3 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.

7 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 the general model in which we allow for Total Factor Productivity (TFP) dynamics and non-constant returns to scale (Model 1). Because this general technology no longer has a known location and scale, we use an age-invariance restriction for identification. Given the structure of the 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). Note that we do not assume any particular values for these measurement parameters, only the age invariance of them, and treat the measurement parameters as free parameters to be estimated.

We also consider results using alternative restricted models, and estimates which do not correct for measurement error and treat the measures as error free measures. We briefly discuss the policy predictions of these models below, but, for brevity, we report estimates of these several alternative models in the Appendix.

Another key issue involves interpreting the magnitude of the parameter estimates.

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 counter-factual experiments using the estimated model. These exercises provide necessary metrics to interpret the estimates with respect to adult outcomes, years of completed schooling at age 23 and earnings at age 29.

7.1 Parameter Estimates

7.1.1 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-Esteem1 (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.

7.1.2 Investment Function

Table 3 reports the estimates of the investment function specified in Section 6.1.2. 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 are 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.

7.1.3 Production Function

Table 4 reports the parameter estimates for the technology of skill formation, as described in Section 6.1.1. We present measurement error corrected estimates of the two versions of the model: our preferred unrestricted Model 1 and the restricted Model 2. We turn first to the unrestricted Model 1 estimates.

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 statistically significant from 0 negative coefficient on the interaction term $(\ln \theta_t \ln I_t)$ indicating that we reject the Cobb-Douglas special case.

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. These heterogeneous investment elasticities suggest that targeting interventions to improve children's skills would have the largest effect on skill disadvantage children. This estimate stands in contrast to the estimates reported in Cunha et al. (2010). They estimate a CES technology which implies that the marginal productivity of investment is *higher* for high skill children given that current investments and the current stock of skills are complements. Note also that unlike the constant returns to scale CES case, our unrestricted model allows investment elasticities to be larger than 1, and we estimate, at least for some children, an elastic response of skill formation to investment.

The high TFP estimate for age 5-6 and the increasing returns to scale (indicated by the sum of the coefficients being greater than 1) 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 the unrestricted Model 1 to the restricted 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 the estimated return to scale of 2.66 in the early period indicating increasing returns to scale. The estimated return to scale declines with the child's age to a value of 1.3 at older ages, revealing that even for older children we can reject constant returns to scale. In addition, the estimate of high positive TFP term also indicates that we clearly reject the assumption of a 0 log TFP in Model 2. As discussed below, these differences in production function estimates imply very different investment and policy effects, with the restricted Model 2 estimates implying a much smaller effect of an income transfer on children's skill development than in our preferred unrestricted model.

7.1.4 Adult Outcomes

Table 5 presents our estimates of the completed schooling outcome equation and log earnings equation. We estimate that a percentage change in children skills at age 13-14 leads to an increase of 0.086 years of school. We also find that a 1 percentage change in children skills leads to a 0.021 percentage change in earnings at age 29. Below, we use these estimates to "anchor" our policy estimates to a meaningful adult outcome metric.

7.2 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. In addition to growth in mean skills, we estimate that the latent distribution of cognitive skills becomes more dispersed as children age. Inequality rises substantially as there are different rates of skills 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.

7.3 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. While we do not have a fully developed model of household resource allocation to provide a more realistic setting to evaluate these policy, 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 measurement error and the specification of general technologies.

7.3.1 Short and Long-Term Effects

Before we analyze the results for our particular parameter estimates, 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 $t+1$:

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

$\Delta_{t+1,t}(\Omega_t)$ 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

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

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 kind of "multiplier effect" on later skill development).

7.3.2 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 percent change in the level of latent children's skills at age 13-14 by the different timing (age) of income transfer:

$$100 \cdot E(\frac{\theta_T^{'}(a) - \theta_T}{\theta_T})$$

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).

7.3.3 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.

7.3.4 Comparison with Dahl and Lochner (2012)

Our estimated effects of a family income transfer on children outcomes are similar to some previous finding in the literature using different sources of identifying variation. Using changes in the Earned Income Tax Credit (EITC) to instrument for family income, Dahl and Lochner (2012) find that a \$1,000 dollars increase in family income implies an increase in PIAT score of about 4.5% of a standard deviation.[20] To directly compare our estimates to their reported effects, we compute the equivalence between their PIAT score outcome and years of schooling. We calculate that the Dahl and Lochner (2012) estimates imply an average increase of about 0.54 months of schooling for a $1,000 transfer at age 11-12.[21] This results is quite similar to our main results (see Figure 8).

7.3.5 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,

[20]This result is based on the results reported in the correction dated March 2016 to the previous results (Table 4). In comparing our results to their results, it should be noted that the policy considered is different. Dahl and Lochner (2012) consider a change in the EITC, which affects after-tax wage rates, parental labor supply, and hence parental time allocation, and we consider here a pure income transfer (where we do not distinguish between income from labor and other sources).

[21]As an outcome, Dahl and Lochner (2012) use a combined PIAT test score (the average of the three separately age-standardized tests in Math, Reading Recognition, and Reading Comprehension). We rescale the PIAT scores in terms of schooling in the same way as we estimate the factor loadings for different skill measures. Define S to be years of schooling at age 23, Z_T to be the PIAT test score at age 13-14 (period T = 4), and Z_{T-1} the PIAT test score at age 11-12 (period 3). We write

$$S = \mu_S + \alpha_S \ln \theta_T + \eta_S$$

$$Z_T = \mu_T + \lambda_T \ln \theta_T + \epsilon_T$$

$$Z_{T-1} = \mu_{T-1} + \lambda_{T-1} \ln \theta_{T-1} + \epsilon_{T-1}$$

Under the assumption that error terms are uncorrelated, the following ratio of covariances provides the scaling of adult schooling with respect to the PIAT test score: $\frac{\alpha_S}{\lambda_T} = \frac{Cov(S, Z_{T-1})}{Cov(Z_{T-1}, Z_T)}$.

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). 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 but 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 7 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 suggest 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.

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

Our results presented thus far have been focused on our preferred model estimates: estimates of the general unrestricted technology (Model 1) with measurement error correction. 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 repeats the estimates from the preferred model: using the unrestricted Model 1 and correcting for measurement error, we estimate that $1,000 income transfer at age 5-6 would increase average schooling by about 1.8 additional month. In comparison, using the restricted Model 2 (assuming constant 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 restricting the model and ignoring possible TFP dynamics and non-constant 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. Using these uncorrected estimates, we estimate policy effects less than half the size of the preferred measurement error corrected estimates of the most general model, Model 1. These substantially lower estimates of the effect of an income transfer are consistent with the standard attenuation bias in standard linear models, where classical measurement error biases coefficient estimates toward 0. Our models are dynamic, non-linear, and consist of inter-related multiple equations, so there is no clear theoretical prediction about the sign of the measurement error bias. But we estimate in this case that ignoring measurement error would substantially bias downward the estimates of the ATE of the income transfer policy.

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 model to have constant returns to scale and no 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.

7.5 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).[22] The benefit of this policy is the comparison between the present value of worker's earnings 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.

8 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 type 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 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 tradeoffs, 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 develop-

[22]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 %.

ment allowing for non-constant returns to scale, dynamics in TFP, and for parental investment to endogenously depend on unobserved children's skills.

Our empirical results show a pattern of rapid skill development from age 5 to 14. We find that as children age, not only does their mean skill level increase, but the level of skill inequality also increases. 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) 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 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.

Online Appendix

References

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.

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, 2013a.

Timothy N. Bond and Kevin Lang. The black-white education-scaled test-score gap in grades k-7. *NBER Working Paper*, (19243), 2013b.

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.

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

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.

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.

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. Standard errors in parenthesis are computed using a cluster bootstrap.

Table 3: Estimates for Investment (Model 1)

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 (see Section 6.2.1). 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 cluster bootstrap.

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

	Model 1 (Free Return to Scale Technology and TFP Dynamics)				Model 2 (Restricted Return to Scale Technology and No TFP Dynamics)			
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 skills formation for both Model 1 and Model 2 (see Sections 6.2.1 and 6.2.2) . 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 $t + 1$, 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 cluster bootstrap.

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	0.151	0.021
at age 13-14	(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 (see Section 6.2.1) 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 cluster bootstrap.

Table 6: Estimated Policy Effects under Different Modeling Assumptions

Panel A: ATE by Age of Income Transfer

Measurement Error Corrected

	Age of Income Transfer ($ 1000)			
Model	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

	Age of Income Transfer ($ 1000)			
Model	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 vs Model 2, see Sections 6.2.1 and 6.2.2) and different estimators (controlling for measurement error or not). The 90% confidence interval in square brackets are computed using a cluster bootstrap. Panel B shows the ATE respect to family income for the different model specifications and different

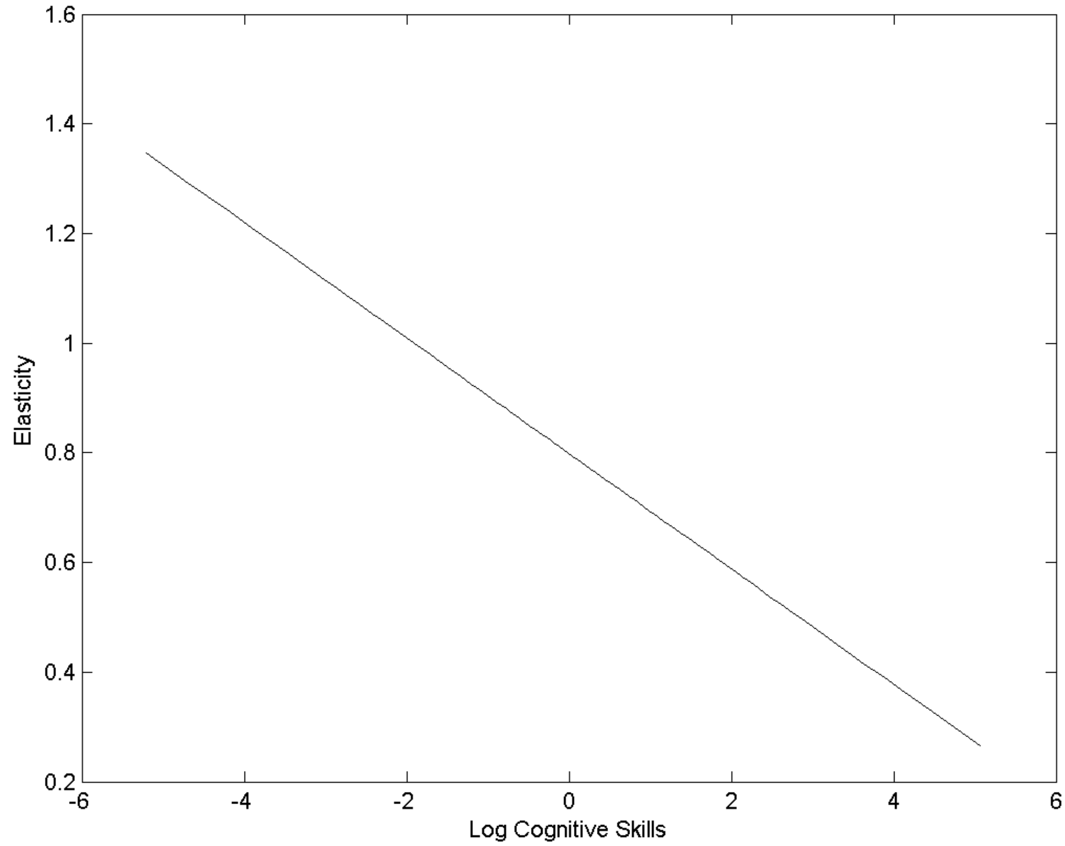
Table 7: Average Effect of an Income Transfer by Age of Transfer (Outcome: PV of Earnings)

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

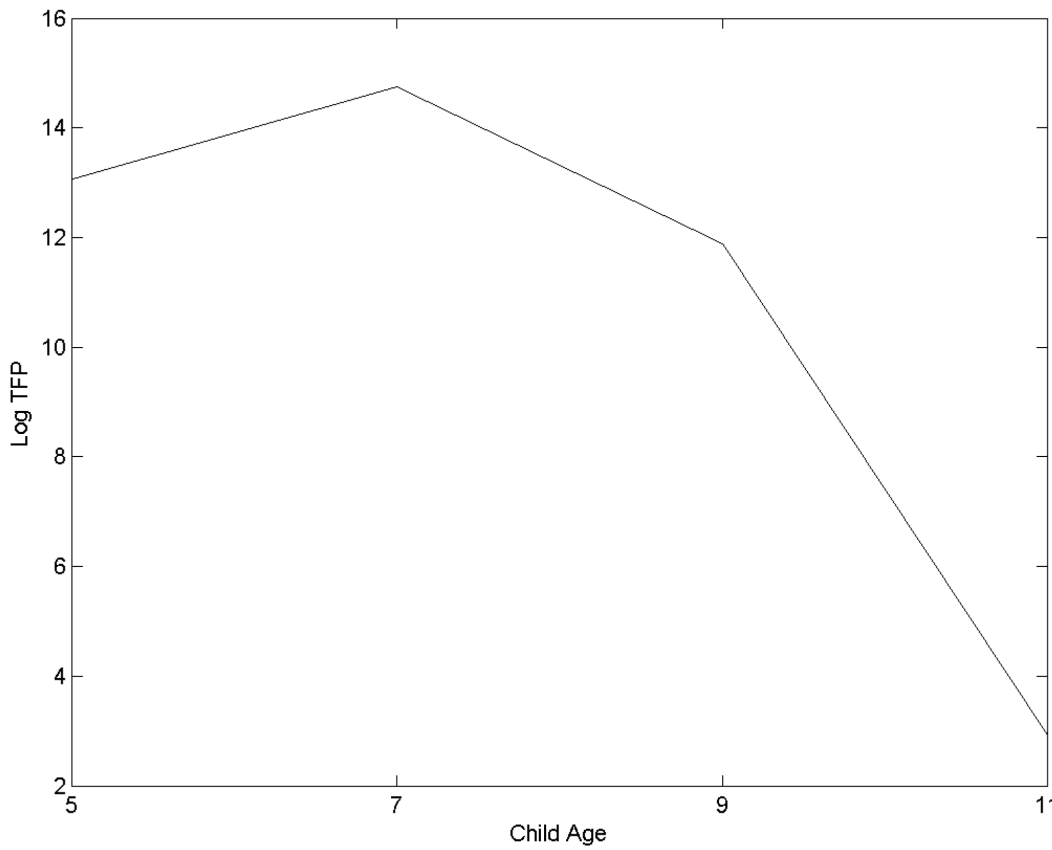
Notes: This table shows the benefit-cost analysis for a 1000 dollars transfer to family of a future median earner workers with 12 years of completed education. The benefit on the PV of earnings is the difference between the present value of earnings with and without that transfer when worker was age 5-6. The effect of family income transfer on earning 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)



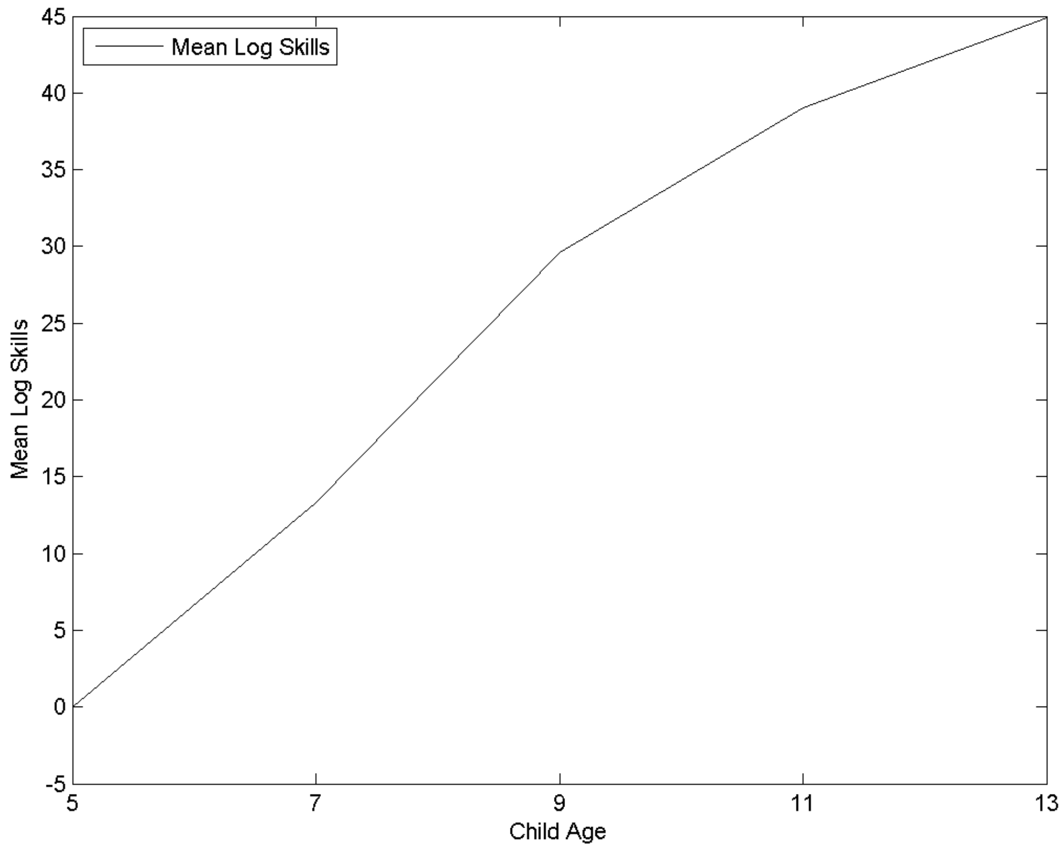
Notes: This figure shows the measurement error corrected estimates of the elasticity of children's skills at age 7-8 (θ_1) with respect to parental investments at age 5-6 (I_0) for Model 1: $\frac{\partial \ln \theta_1}{\partial \ln I_0} = \gamma_{2,0} + \gamma_{3,0} \ln \theta_0$.

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



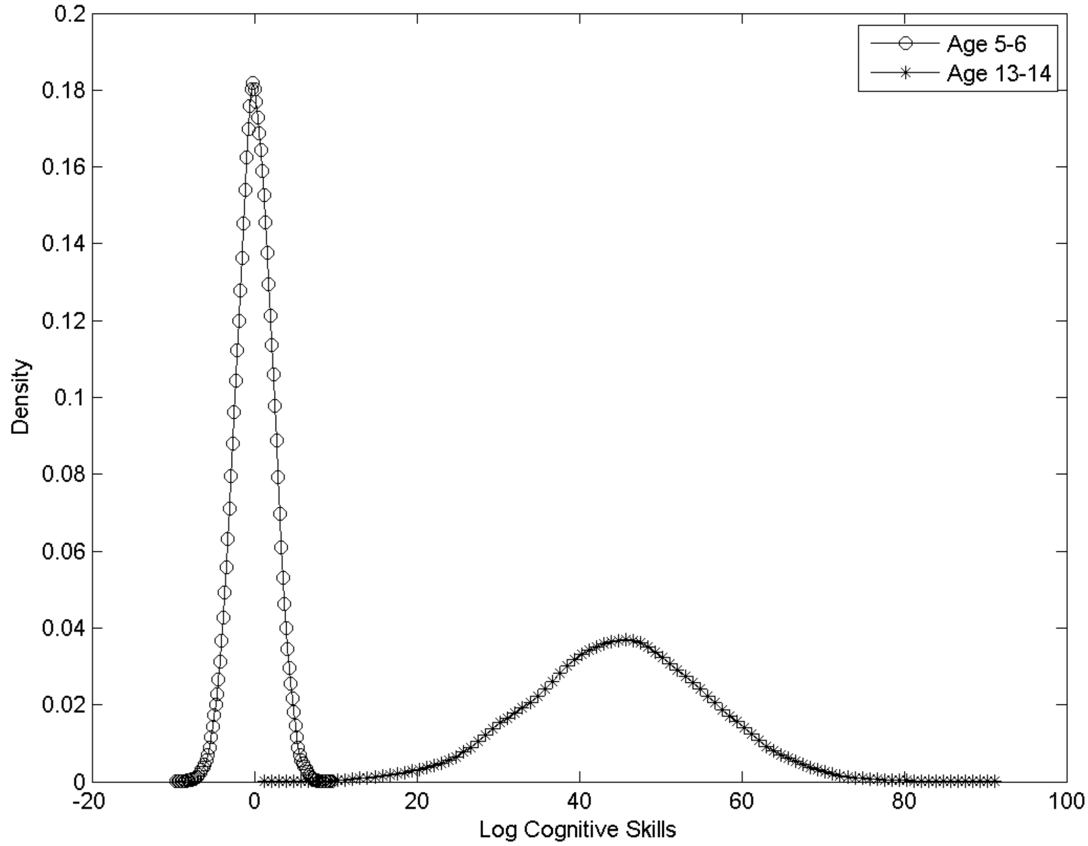
Notes: This figure shows the estimated log TFP (correcting for measurement error) for Model 1 (see Section 6.2.1). The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on.

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



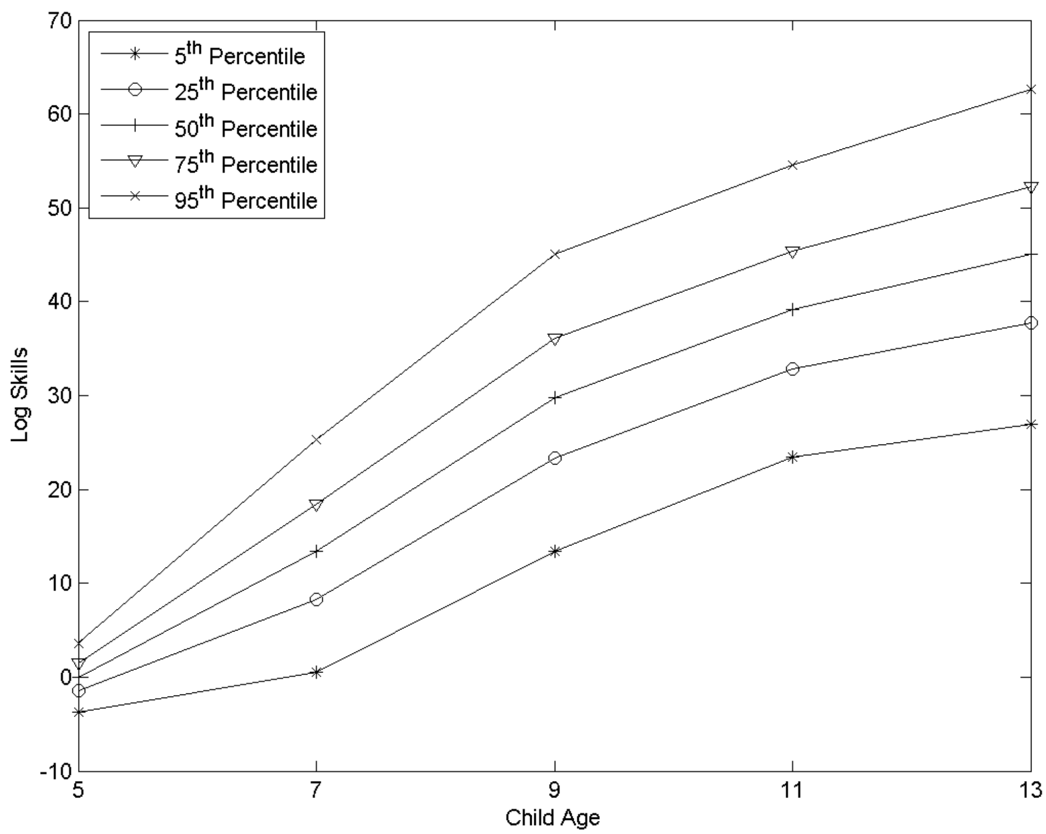
Notes: This figure provides the mean log latent skills ($E(\ln\theta_t)$) predicted by the estimated Model 1 (see Section 6.2.1), controlling for measurement error) . The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on. Log latent skills at age 5-6 are normalized to be mean 0.

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



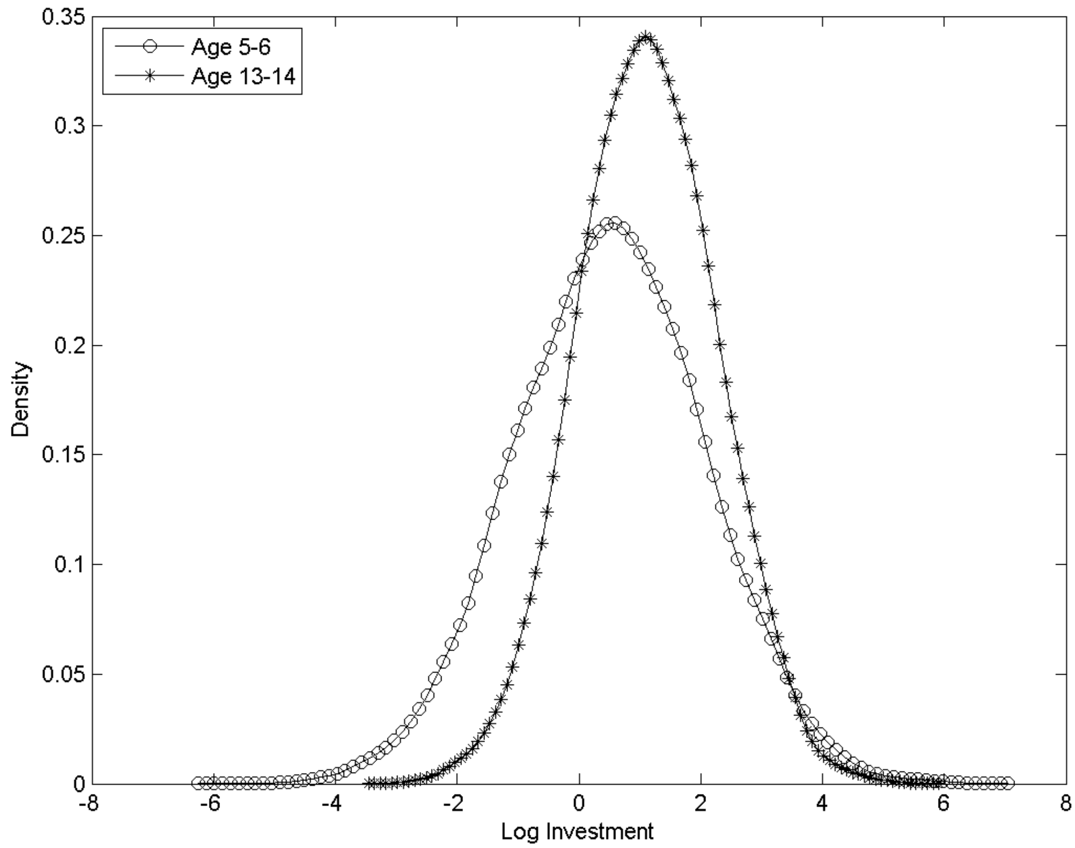
Notes: This figure shows the distribution of log latent skills at age 5-6 and at age 13-14 simulated from the estimated Model 1 (see Section 6.2.1), controlling for measurement error. Log latent skills at age 5-6 are normalized to be mean 0.

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



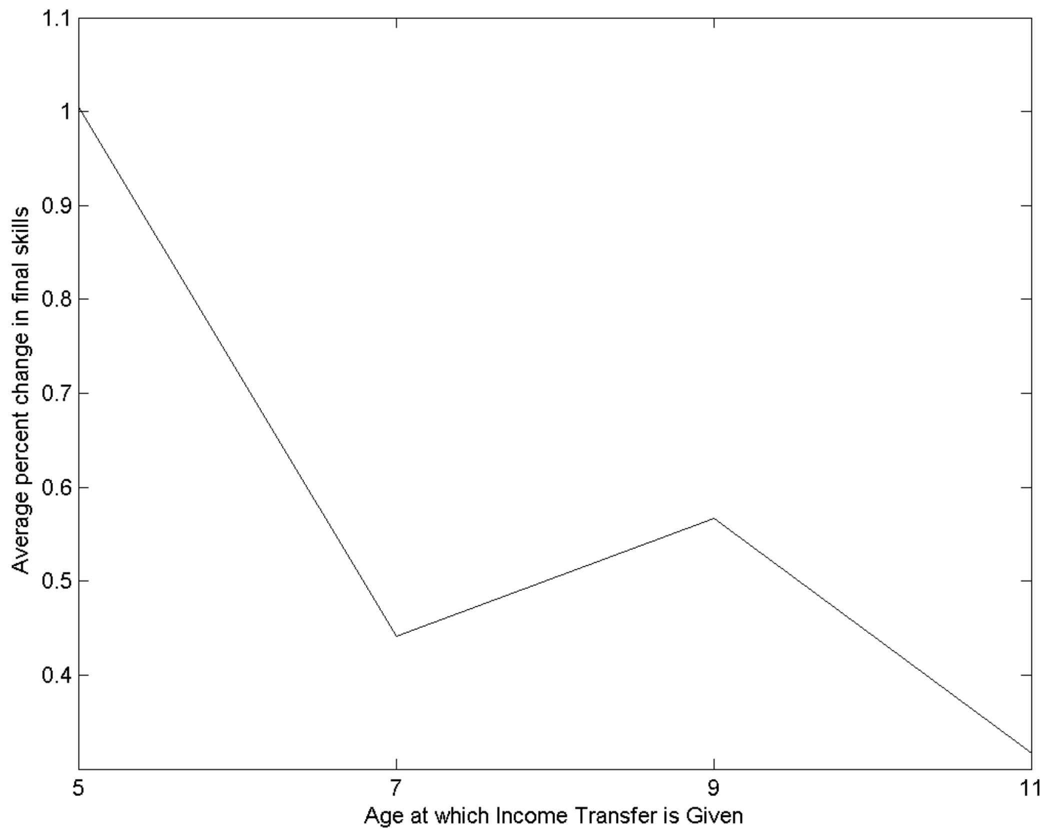
Notes: This figure shows the dynamics in the distribution of the log latent skill distribution for the estimated Model 1 (see Section 6.2.1), controlling for measurement error. Log latent skills at age 5-6 are normalized to be mean 0.

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



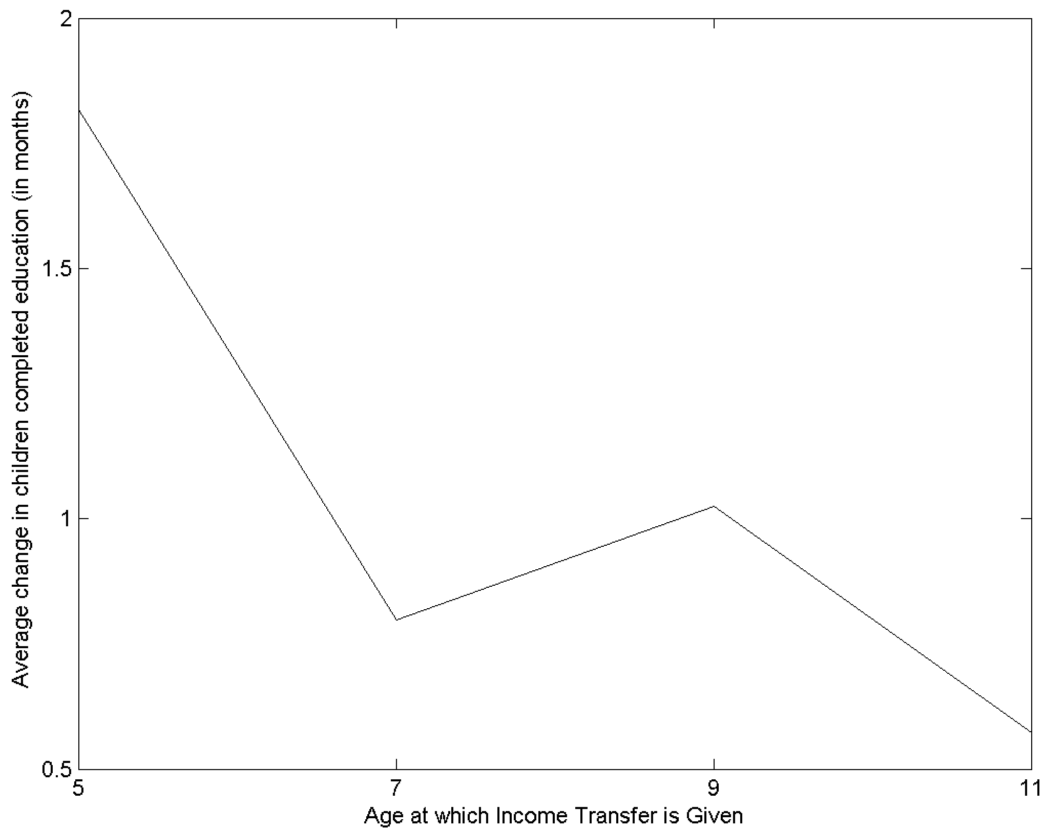
Notes: This figure shows the distribution of log latent investments at age 5-6 and at age 13-14 simulated from the estimated Model 1 (see Section 6.2.1), controlling for measurement error.

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



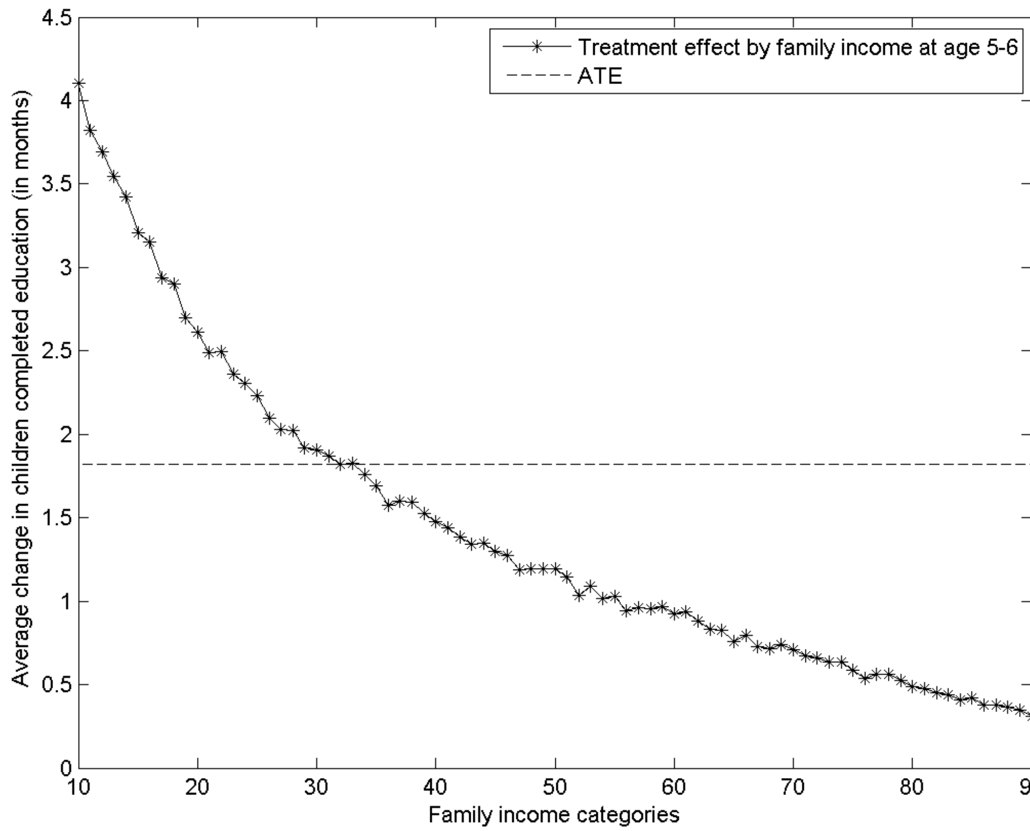
Notes: This figure shows the average percent change in the level of latent children's skills at age 13-14 by the different timing (age) of income transfer for the estimated Model 1 (see Section 6.2.1), controlling for measurement error. The transfer is \$1,000 in family income at some age t. We report $100 \cdot E(\frac{\theta'_T(a)-\theta_T}{\theta_T})$, 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)



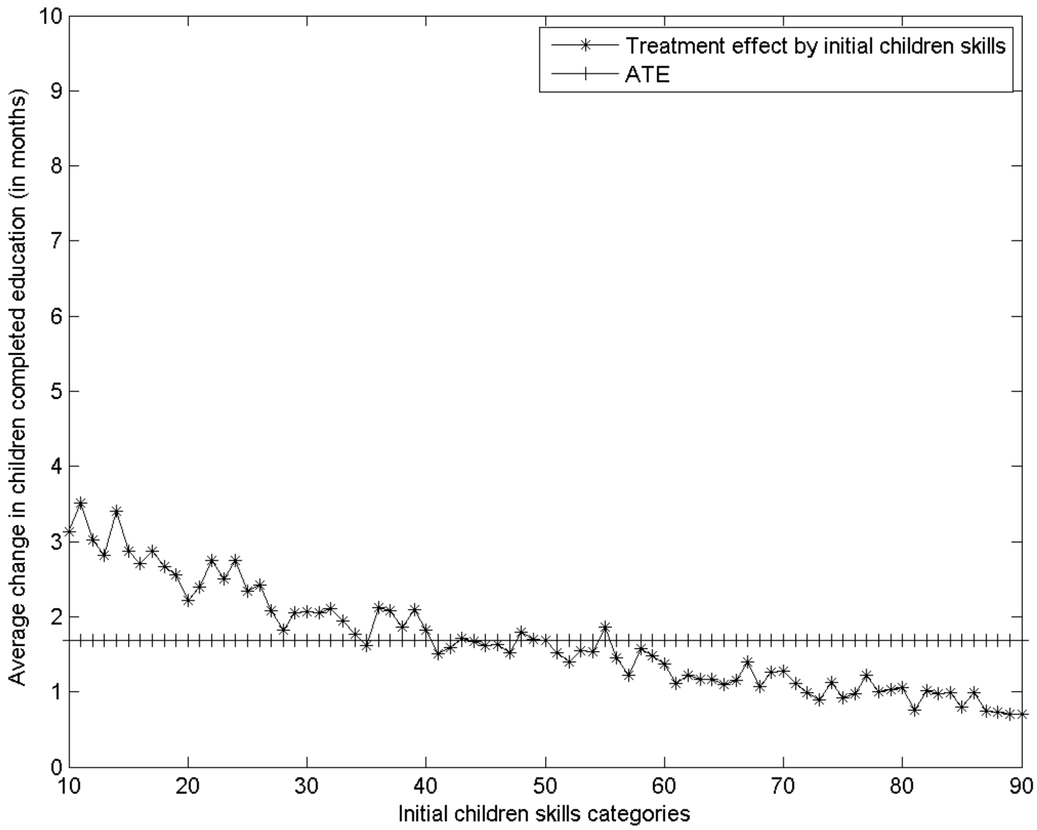
Notes: This figure shows the average change in the number of months of completed schooling at age 23 by different timing (age) of income transfer for the estimated Model 1 (see Section 6.2.1), controlling for measurement error. We report $E\left[S'(a) - S\right]$, where $S'(a)$ is the number of months 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 (see Section 6.2.1), 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 (see Section 6.2.1), 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.