NBER WORKING PAPER SERIES

IDENTIFICATION OF DYNAMIC LATENT FACTOR MODELS:
THE IMPLICATIONS OF RE-NORMALIZATION IN A MODEL OF CHILD DEVELOPMENT

Francesco Agostinelli
Matthew Wiswall

Working Paper 22441
http://www.nber.org/papers/w22441

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

We thank Orazio Attanasio, Ryan Bush, Flavio Cunha, Lance Lochner, Costas Meghir, 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development
Francesco Agostinelli and Matthew Wiswall
NBER Working Paper No. 22441
July 2016
JEL No. C38,J13

ABSTRACT

A recent and growing area of research applies latent factor models to study the development of children's skills. Some normalization is required in these models because the latent variables have no natural units and no known location or scale. We show that the standard practice of “re-normalizing” the latent variables each period is over-identifying and restrictive when used simultaneously with common skill production technologies that already have a known location and scale (KLS). The KLS class of functions include the Constant Elasticity of Substitution (CES) production technologies several papers use in their estimation. We show that these KLS production functions are already restricted in the sense that their location and scale is known (does not need to be identified and estimated) and therefore further restrictions on location and scale by re-normalizing the model each period is unnecessary and over-identifying. The most common type of re-normalization restriction imposes that latent skills are mean log-stationary, which restricts the class of CES technologies to be of the log-linear (Cobb-Douglas) sub-class, and does not allow for more general forms of complementarities. Even when a mean log-stationary model is correctly assumed, re-normalization can further bias the estimates of the skill production function. We support our analytic results through a series of Monte Carlo exercises. We show that in typical cases, estimators based on “re-normalizations” are biased, and simple alternative estimators, which do not impose these restrictions, can recover the underlying primitive parameters of the production technology.

Francesco Agostinelli
Arizona State University
W.P. Carey School of Business
P.O. Box 879801
Tempe, AZ 85287
fagostin@asu.edu

Matthew Wiswall
Arizona State University
W.P. Carey School of Business
P.O. Box 879801
Tempe, AZ 85287
and University of Wisconsin-Madison
and also NBER
matt.wiswall@gmail.com

1 Introduction

A recent and growing area of research applies latent factor models to study the development of skills in children. The goal of this research is to characterize the optimal timing and form of interventions to improve children's skills (Cunha and Heckman (2007); Cunha et al. (2010); Cunha and Heckman (2008); Attanasio et al. (2015a,b); Pavan (2015)).[1] In these models, the stock of children's skills develop dynamically through childhood according to a specified skill production technology, and children's skills, and in some work investments in skills as well, are assumed to be measured with error.

The identification of these models applies the techniques developed for cross-sectional latent factor models to the dynamic models describing the development of children's skills.[2] Some normalization is required in these models because the latent variables have no natural units and no known location or scale. Much of the recent literature "re-normalizes" the model every period, essentially treating each period as a separate cross-section and the stock of children's skills in each period as a separate latent factor. In several papers, re-normalization takes the form of setting the latent (log) stock of skills to be mean 0 in every period, and the factor loading for one measure each period is assumed to be 1. This re-normalization approach is in contrast to normalizing the latent variables to one particular period (say the initial earliest period of childhood) and therefore locating and scaling all latent variables to this initial period. Because the re-normalization approach treats the inputs and outputs of the production process–the latent factors in all periods–as separate latent variables, re-normalization has implications for the dynamic production relationships implied by the technology of child development.

In this paper we show that re-normalization of dynamic latent factor model is over-identifying and restrictive when used simultaneously with production technologies which already have a known location and scale (KLS). While in principle the re-normalization approach is without loss of generality with general enough production technologies, the parametric technologies estimated in practice are often already sufficiently restricted to allow identification without re-normalization. KLS functions include the Constant Elasticity of Substitution (CES) production technologies several papers use in their estimation, as in the influential paper Cunha et al. (2010).

[1]Other recent work shows the importance of measurement issues to understanding the level and growth of inequality in children's skills, such as the black-white test score gap (see Bond and Lang, 2013a,b).

[2]For the early literature on factor models in economics see Anderson and Rubin (1956); Jöreskog and Goldberger (1975); Goldberger (1972); Chamberlain and Griliches (1975); Chamberlain (1977a,b). For a more recent reference, see Carneiro et al. (2003).

We show that these KLS production functions are already restricted in the sense that their location and/or scale is fixed and known (does not need to be estimated) and therefore further restrictions on the location and scale by re-normalizing the model each period is unnecessary and over-identifying. In these cases, re-normalization is not a set of normalizations imposed without loss of generality, but a set of testable empirical restrictions on the underlying model of child development.

In addition to being unnecessary and over-identifying, we also show that in standard cases using KLS production technologies, re-normalizing the model each period can actually bias the estimation of the production function. Rather than allowing fully unrestricted non-parametric identification of the production technology, re-normalization imposes a mean log-stationary restriction on the dynamics of the latent skill stock, and this restriction is generally consistent only with particular types of production technologies, such as the log-linear (Cobb-Douglas) production technology, and not with more general technologies, such as the general CES technology which some of the previous work actually estimate. Even in the case where the mean-log-stationary restriction is correctly imposed, re-normalization imposes further restrictions on the measurement factor loadings and the scale of the latent variables, which can also bias the estimates of the production function parameters.

The remainder of the paper is organized as follows. We first present a stylized child development model and review the existing identification analysis based on re-normalization. The second part of the paper presents our analysis in a series of remarks. We conclude with a Monte Carlo simulation to demonstrate the validity of our analysis and quantify the extent of the bias. In our Monte Carlo analysis, we show that simple alternative estimators, which do not re-normalize the model and instead impose a normalization on the initial period only, can recover the underlying parameters.

2 Model and Measurement

In this section we present a stylized model of skill formation and measurement, a simplified version of the influential model developed in Cunha et al. (2010). Several subsequent papers follow this model. We conclude this section by briefly summarizing the identification analysis under re-normalization restrictions.

2.1 Skill Formation Technology

Child development takes place over a discrete and finite period, $t = 0, 1, \ldots, T$, where $t = 0$ is the initial period (say birth) and $t = T$ is the final period of childhood (say age 18). There is a population of children and each child in the population is indexed i. For each period, each child is characterized by a stock of skills $\theta_{i,t}$, with $\theta_{i,t} > 0$ for all i and t, and a flow level of investments $I_{i,t}$, with $I_{i,t} > 0$ 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:

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

(1) can be viewed as a dynamic state space model with $\theta_{i,t}$ 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. Given some initial distribution of skills in the population, G_0, and the sequence of investments in children $I_{i,0}, I_{i,1}, \ldots, I_{i,T-1}$ for all i, the technology (1) defines the dynamic process of skill development producing the stock of skills from the birth to adulthood, $\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,T}$ for all i, and the population distribution of skill stocks in all subsequent periods $G_1, \ldots, G_T$.

2.2 Measurement

The stock of skills $\theta_{i,t}$ is not observed in data directly. Instead we have a system of skill measures given by

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

Our measurement system presumes panel data with 3 dimensions: children i, time t, and measure m. We have M_t measures for latent skill in each period t, indexed $m = 1, \ldots, M_t$. $Z_{i,t,1}, \ldots, Z_{i,t,M_t}$ are the measures, $\mu_{t,1}, \ldots, \mu_{t,M_t}$ are the measurement intercepts, and $\lambda_{t,1}, \ldots, \lambda_{t,M_t}$ are the measurement "factor loadings," with $\lambda_{t,m} > 0$ for all t, m. The measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$, vary over time and across measures but are homogenous across children. Finally, $\epsilon_{i,t,1}, \ldots, \epsilon_{i,t,M_t}$ are the measurement errors, with $E(\epsilon_{i,t,m}) = 0$ for all t, m. Given the intercept $\mu_{t,m}$, the assumption of mean zero $\epsilon_{i,t,m}$ errors is without loss of generality. To focus on the already complex identification issues involved with measurement error in skills, we assume investments $I_{i,t}$ are observed without error.[3]

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

[3]To leave aside any issues with normalizations regarding investments, we set log investment to be mean zero in all periods, $E(\ln I_{i,t}) = 0$ for all t, m. 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 $\ln I_{i,t}$ is zero in each period.

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 $\kappa_t \equiv E(X_t)$.

Various assumptions about the dependence of the measurement errors with each other and dependence of the measurement errors and the latent stock of skills and investment play a limited role in our analysis here, but are otherwise important general considerations in these types of measurement models. A sufficient assumption we maintain in this paper is that measurement errors are independent contemporaneously across measures ($\epsilon_{t,m} \perp \epsilon_{t,m'}$ for all $m \neq m'$ and all t), independent over time ($\epsilon_{t,m} \perp \epsilon_{t',m'}$ for all $t \neq t'$ and all m and m'), and independent of the latent stock of skills and investments in any period ($\epsilon_{t,m} \perp \theta_{t'}$ for all t, t' and all m and $\epsilon_{t,m} \perp I_{t'}$ for all t, t' and all m). These assumptions are sufficient, but not strictly necessary, as weaker assumptions allowing for some forms of dependence among measures and among measures and latent variables can be used for identification.

Because latent skills are unobserved and have no natural units, some normalization is clearly necessary. We normalize the initial period as $E(\ln \theta_0) = 0$ and $\lambda_{0,1} = 1$. Under this normalization and with at least 3 measures in the first period, we identify the remaining factor loadings for the period $t = 0$ measures $\lambda_{0,2}, \lambda_{0,3}, \ldots, \lambda_{0,M_0}$ and the $\mu_{0,1}, \mu_{0,2}, \ldots, \mu_{0,M_0}$ measurement intercepts. Following standard arguments, we then identify the distribution of latent skills in the initial period, up to this normalization. For the remainder of the paper, we maintain this initial period normalizations and discuss normalizations and restrictions for periods only after the initial period (for periods $t > 0$).

3 Identification of Dynamic Production Technologies

3.1 Location and Scale of Production Technologies

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 and have unknown scale and location. Central to our analysis is whether the production technology already 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. If the production function already has a known location and scale, then simultaneously imposing normalizations on the latent variables or restrictions on the measurement process to further restrict the scale and location is over-identifying.

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 we identify the location of the production function at either known input vector, and 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 also known. Other points in the production possibilities set may be unknown, i.e. depend on free parameters which then need 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 studies such as (e.g. Cunha et al., 2010). The standard CES technology is

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1 - \gamma_t) I_t^{\phi_t})^{1/\phi_t}. \tag{3}$$

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 (3) satisfies Definition 1 because for any inputs $I_t = \theta_t = a > 0$, $\theta_{t+1} = a$. That is, for inputs which are known to be equal at any value a, we also know the output is a 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 sufficient property to satisfy Definition 1, as shown below. While the scale and location of the production function (3) are known, other points in the production possibilities set are determined by the free production function parameters γ_t and ϕ_t, and these parameters need to be estimated.

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 between inputs to be constant:

$$\ln \theta_{t+1} = \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} (\ln \theta_t)(\ln I_t) \tag{4}$$

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 have a known location is the following CES function with a free Total Factor Productivity (TFP) term:

$$\theta_{t+1} = A_t(\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{1/\phi_t} \tag{5}$$

with $A_t > 0$ representing TFP.[4] In this case the location (in logs) depends on the unknown/free TFP term A_t.

Another class of technologies which does not satisfy Definition 1 are CES technologies without a known returns to scale:

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{\psi_t/\phi_t}. \tag{6}$$

where $\psi_t > 0$ is an unknown 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, $f_t(a, a) = a^{\psi_t}$. For this function, while we know the point $f_t(1, 1) = 1$, we do not know a second point in the production possibilities set. This function then has a known location but an unknown scale.

3.2 Under-Identification

We analyze the under-identification of the general model by examining a general form for the production technology given by

$$\ln \theta_{t+1} = \alpha_t + \beta_t \ln h_t(\theta_t, I_t) \tag{7}$$

where $h_t(\theta_t, I_t)$ is a known location and scale (KLS) function, such as the CES function (3). α_t and $\beta_t > 0$ are free (unknown) location and scaling parameters, respectively, for the technology.

Substituting the latent technology equation (7) into the measurement equation (2), we have the following:

$$Z_{t,m} = \mu_{t,m} + \lambda_{t,m}[\alpha_t + \beta_t \ln h_t(\theta_t, I_t)] + \epsilon_{t,m}$$

Re-arranging,

$$\begin{aligned} Z_{t,m} &= (\mu_{t,m} + \lambda_{t,m}\alpha_t) + (\lambda_{t,m}\beta_t) \ln h_t(\theta_t, I_t) + \epsilon_{t,m} \\ &= a_t + b_t \ln h_t(\theta_t, I_t) + \epsilon_{t,m} \end{aligned} \tag{8}$$

[4] Equivalently, one could write this function as $\theta_{t+1} = (\gamma_{1t}\theta_t^{\phi_t} + \gamma_{2t} I_t^{\phi_t})^{1/\phi_t}$, where $\gamma_{1t} + \gamma_{2t}$ does not equal a known constant. The previous case (3) is a special case of this function with $\gamma_{1t} + \gamma_{2t} = 1$.

From this expression, it is clear that we cannot separately identify the measurement parameters $\mu_{t,m}$, $\lambda_{t,m}$ from the production function parameters α_t, β_t.

3.3 Observationally Equivalent Models

We conclude our discussion of the under-identification issue, by examining two alternative models and show that they are observationally equivalent. We define a "model" as a combination of assumptions about (i) the production technology f_t, (ii) the latent variables θ_t, and (iii) the measurement parameters $\mu_{t,m}, \lambda_{t,m}$. For each model, we assume the initial ($t = 0$) distribution of latent skills have been identified, as outlined above. Recall that measures take the form:

$$Z_{t,m} = \mu_{t,m} + \lambda_{t,m} \ln \theta_t + \epsilon_{t,m}.$$

The first model assumes the production technology is KLS, but leaves the latent variables and measurement parameters free:

Model 1 *(i)* $\ln(\theta_{t+1}) = h_t(\theta_t, I_t)$ *(*$\alpha_t = 0$, $\beta_t = 1$*)*
(ii) $E(\ln(\theta_t))$ *free for all* $t > 0$
(iii) $\lambda_{t,m}$ *free for all* $t > 0$

where $h_t(\theta_t, I_t)$ is a known location and scale technology (Definition 1). The second model leaves the production function free (α_t, β_t free production function parameters), but restricts the latent variables and measurement parameters:

Model 2 *(i)* $\ln(\theta_{t+1}) = \alpha_t + \beta_t h(\theta_0, I_0)$
(ii) $E(\ln(\theta_t)) = 0$ *for all* $t > 0$
(iii) $\lambda_{t,m} = 1$ *for all* $t > 0$

Model 1 and Model 2 are observationally equivalent models. The two models present observationally equivalent normalizations for the under-identified general formulation (8). Model 1 normalizes the technology (sets $\alpha_t = 0$ and $\beta_t = 1$) but leaves the latent variables and measurement parameters free. Model 2 normalizes the latent variables and measurement parameters ($E(\ln \theta_t) = 0$ for all t and $\lambda_{t,m} = 1$) but leaves the production technology free. Note that in Model 2, the normalization on the latent variable $E(\ln \theta_t)$ implicitly fixes the measurement intercepts to be the unconditional mean of the measures: $\mu_{t,m} = E(Z_{t,m})$.[5]

[5] There are of course other possible normalizations. One possibility is to replace Model 2 (ii) with $\mu_t = 0$ for all t.

4 Re-Normalization

Our main argument is that while some normalization in this class of latent factor models is clearly necessary, in practice prior papers "over-normalize" the model: they simultaneously estimate production technologies that already have a known location and scale *and* "re-normalize" the latent variables. In this section, we review the previous approaches, showing that these approaches impose over-identifying restrictions, and showing that these over-identifying restrictions can bias the estimation.

4.1 Review of Identification with Re-Normalization

We define *re-normalization* as:

Definition 2 *Re-Normalization:*
(i) $E(\ln \theta_t) = 0$ *for all* $t > 0$
(ii) $\lambda_{t,1} = 1$ *for all* $t > 0$

where we have labeled the arbitrarily chosen normalized measure in each period to be measure $m = 1$. We continue to refer to re-normalization as a "normalization," but in fact we argue below that in common cases where the technology is already a known location and scale technology, re-normalization is a set assumption with empirical content and testable restrictions.

Under re-normalization, the latent skill stock in each period is treated as a separate latent factor and the measurement system is "re-normalized" every period. Specifically, re-normalization (i) imposes that latent skills are mean log stationary:

$$E(\ln \theta_0) = E(\ln \theta_1) = \cdots = E(\ln \theta_T).$$

In addition, (ii) restricts latent skills to "load onto" one arbitrarily chosen measure each period in the same way:

$$\frac{\partial E(Z_{0,1} | \ln \theta_0)}{\partial \ln \theta_0} = \frac{\partial E(Z_{1,1} | \ln \theta_1)}{\partial \ln \theta_1} = \cdots = \frac{\partial E(Z_{T,1} | \ln \theta_T)}{\partial \ln \theta_T}.$$

Below, we discuss each of these restrictions and argue that each can bias the estimation of the production function parameters.

We first briefly review the identification analysis under re-normalization, following the analysis in Cunha et al. (2010). It is useful to think of their identification procedure as a three step procedure. In the first step, they recover the measurement error parameters, $\mu_{t,m}$ and $\lambda_{t,m}$, for all periods. With $M_t \geq 3$ measures for each period,[6] re-normalization (ii) ($\lambda_{t,1} = 1$ for all t) allows one to use covariances among measures to identify the $\lambda_{t,m}$ parameters for all t and all measures m:

$$\lambda_{t,m} = \frac{Cov(Z_{t,m}, Z_{t,m'})}{Cov(Z_{t,m'}, Z_{t,1})} \text{ for all } m \neq m' \neq 1 \text{ and all } t$$

The measurement intercepts are recovered using the mean log-stationarity assumption, re-normalization (i) ($E(\ln \theta_t) = 0$ for all t):

$$\mu_{t,m} = E(Z_{t,m}) \text{ for all } m \text{ and } t.$$

In the second step, the cross-sectional distribution of latent skills and investments for all periods are identified using the identified measurement parameters. For each period, residual measures can be constructed, and each residual measure identifies the latent skill up to the measurement error:

$$\begin{aligned} \widetilde{Z}_{t,m} &= (Z_{t,m} - \mu_{t,m})/\lambda_{t,m} \\ &= \ln \theta_t + \epsilon_t/\lambda_{t,m} \text{ for all } m, t \end{aligned}$$

The joint distribution of latent skills and investments for all periods $(\theta_0, I_0, \theta_1, I_1, \ldots, \theta_T, I_T)$, is then identified using these residual measures.

Finally, in the third step, Cunha et al. (2010) identify the production technology linking latent skill stocks in period $t+1$ (θ_{t+1}) to period t latent skill stocks and investment (θ_t, I_t). The claim is that non-parametric identification of the production technology follows because the joint distribution of all of the latent input and output variables have been identified in the previous steps. As we detail below, the re-normalization approach has already implicitly restricted the technology in the previous steps. Therefore the distribution of latent variables identified in the previous steps must satisfy the imposed re-normalization restrictions. These identification results therefore do not imply that *any* arbitrary production function can be identified.

4.2 Over-Identification with Re-Normalization

We now turn to our first result. If the $f_t(\theta_t, I_t)$ production functions characterizing the skill development technology already have a known location and scale, then further restrictions to fix the location and scale are unnecessary. We classify these KLS models with re-normalization imposed as Model 3:

[6] While $M_t \geq 3$ measures is sufficient for identification, it is not strictly necessary. Under some conditions, as few as two measures can be sufficient.

Model 3 *KLS with Re-Normalization*
(i) $\ln(\theta_{t+1}) = h_t(\theta_t, I_t)$
(ii) $E(\ln(\theta_t)) = 0$ *for all* $t > 0$
(iii) $\lambda_{t,m} = 1$ *for all* $t > 0$

This model "over-normalizes" the under-identified model (8), simultaneously restricting the technology *and* the latent variables and measurement parameters. Model 3 is not observationally equivalent to the Models 1 and 2, and Model 3 can be empirically tested against those more general models. We summarize this conclusion in the following remark:

Remark 1 *Re-normalization is over-identifying when the production technology has a known location or scale.*

For example, Cunha et al. (2010) propose estimation of a KLS technology of skill formation (CES production function) under re-normalization.[7] In the models considered by Cunha et al. (2010), the re-normalization assumptions are not without loss of generality given the production technologies they consider already have a fixed location and scale.

4.3 Identification of KLS Functions without Re-Normalization

A corollary of our result that re-normalization is over-identifying with known location and scale functions is that point identification of these cases is possible without re-normalization. We next consider a simple example which shows identification of a KLS production technology without imposing the re-normalization restrictions.

Consider the Cobb-Douglas specification of the technology and a two period model $t = 0, 1$:

$$\ln \theta_1 = \gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0 \tag{9}$$

[7] In a recent private email correspondence, the authors of Cunha et al. (2010) have revealed that, although their identification and estimation description in the published paper indicates otherwise (see p. 905 in particular which indicates that the latent distribution of log skills is assumed to be mean 0 in all periods), their specific estimation algorithm did not actually impose the mean log-stationarity assumption, re-normalization (i). This explains why their estimated elasticity of substitution is not 1 (Cobb-Douglas) as would be necessarily imposed by the re-normalization assumption (see discussion below). Cunha et al. (2010) have revealed that they did maintain in the estimation the constant factor loading re-normalization restriction, re-normalization (ii), which we show below is over-identifying and potentially biasing.

with $\gamma_0 \in (0,1)$. The Cobb-Douglas function is a KLS function, as shown above. Our goal is to identify the primitive production function parameter γ_0 and the measurement parameters $\mu_{1,m}$ and $\lambda_{1,m}$.

Identification proceeds using empirical covariances of measures of skills and investments. The covariance between a measure of the stock of skills in period 1 $Z_{1,m}$ and observed log investment in the initial period $\ln I_0$ is given by

$$\begin{aligned} Cov(Z_{1,m}, \ln I_0) &= \lambda_{1,m} Cov(\ln \theta_1, \ln I_0) \\ &= \lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, \ln I_0) + (1-\gamma_0) Var(\ln I_0)] \end{aligned} \quad (10)$$

This covariance is a combination of the production function parameter γ_0, the measurement parameter (factor loading) for this measure $\lambda_{1,m}$, and moments of the joint distribution of initial skills and investments $Cov(\ln \theta_0, \ln I_0)$ and $Var(\ln I_0)$. Consider a second covariance using squared log investment but the *same* measure of period 1 skills $Z_{1,m}$:

$$\begin{aligned} Cov(Z_{1,m}, (\ln I_0)^2) &= \lambda_{1,m} Cov(\ln \theta_1, (\ln I_0)^2) \\ &= \lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, (\ln I_0)^2) + (1-\gamma_0) Cov(\ln I_0, (\ln I_0)^2)] \end{aligned} \quad (11)$$

The ratio of these two covariances is

$$\begin{aligned} \frac{Cov(Z_{1,m}, \ln I_0)}{Cov(Z_{1,m}, (\ln I_0)^2)} &= \frac{\lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, \ln I_0) + (1-\gamma_0) Var(\ln I_0)]}{\lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, (\ln I_0)^2) + (1-\gamma_0) Cov(\ln I_0, (\ln I_0)^2)]} \\ &= \frac{\gamma_0 Cov(\ln \theta_0, \ln I_0) + (1-\gamma_0) V(\ln I_0)}{\gamma_0 Cov(\ln \theta_0, (\ln I_0)^2) + (1-\gamma_0) Cov(\ln I_0, (\ln I_0)^2)} \end{aligned} \quad (12)$$

Taking the ratio of these covariances has eliminated the unknown measurement parameter $\lambda_{1,m}$. Our approach here is an example of the general approach we develop in a companion paper (Agostinelli and Wiswall (2016)). We treat the measurement parameters as "nuisance parameters" and use particular transformations of observed data moments to eliminate them, a method similar in spirit to that used to eliminate fixed effects in standard panel data analysis.

Given the initial period normalizations (described above), we identify the initial period moments, $Var(\ln \theta_0)$, $Var(\ln \theta_0)$, and $Cov(\ln \theta_0, \ln I_0)$. Solving (12) for the production function primitive γ_0 then shows that the production function parameter is identified without imposing *any* restrictions on the measurement parameters $\mu_{1,m}, \lambda_{1,m}$ or the latent distribution of θ_1. One remarkable aspect of this identification concept is that identification of the production function in this case requires only a single measure of period 1 skills, rather than the multiple measures typically thought to be required.

Once γ_0 has been identified, we can also then identify the parameters of the measurement equation. The measurement factor loading $\lambda_{1,m}$ is identified from

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} \cdot Cov(\ln \theta_1, \ln \theta_0)} \tag{13}$$

$$= \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} \cdot (\gamma_0 Var(\ln \theta_0) + (1-\gamma_0) Cov(\ln \theta_0, \ln I_0))}, \tag{14}$$

where $\lambda_{0,m}$ is identified up to the initial period normalization. We can also identify the measurement intercept for $Z_{1,m}$ from

$$\mu_{t,m} = E(Z_{t,m}) - \lambda_{t,m} E(\gamma_0 E(\ln \theta_0) + (1-\gamma_0) E(\ln I_0)).$$

which in this Cobb-Douglas case is simply $\mu_{t,m} = E(Z_{t,m})$ given the initial period normalization.

Agostinelli and Wiswall (2016) provide more general identification results and other examples, including for general CES technologies. The Monte Carlo exercises we conduct below show how to use these identification results to develop an estimator that works well in practice.

4.4 Errors-in-Variables Formulation

To further understand the over-identifying nature of the re-normalization restrictions, it is useful to re-formulate the problem as a traditional errors-in-variables linear regression model Chamberlain and Griliches (1975). For simplicity, we again consider the Cobb-Douglas case (9). We proceed as before using the initial period normalization and forming measures for the first period:

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

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

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

The measurement parameters $\mu_{1,m}$ and $\lambda_{1,m}$ are treated as free parameters, and we do not impose the re-normalization restriction.

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

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

Re-arranging, we have

$$\begin{aligned} Z_{1,m} &= \mu_{1,m} + \lambda_{1,m}\gamma_0\widetilde{Z}_{0,m} + \lambda_{1,m}(1-\gamma_0)\ln I_0 + (\epsilon_{1,m} + \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}) \\ &= \beta_0 + \beta_1\widetilde{Z}_{0,m} + \beta_2 \ln I_0 + \pi_{1,m} \end{aligned} \tag{15}$$

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 (15) now has the standard errors-in-variables form: (15) is a linear regression of a measure of period 1 skills $Z_{1,m}$ on a measure for period 0 skills $\tilde{Z}_{0,m}$ and observed investment. 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 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 ($\tilde{Z}_{0,m'}$) can be used as an instrument for $\tilde{Z}_{0,m}$. Using this instrumental variables approach we identify β_0, β_1, and β_2. Second, with β_1 and β_2 identified, we can then solve for the underlying primitive parameters γ_0, $\mu_{1,m}$, and $\lambda_{1,m}$:

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

The key to identification in this case is that this commonly used production function (9) has a known location and scale (i.e., the factor shares sum to 1 and there is no free intercept) and hence we can identify the production function parameters separately from the measurement parameters. The primitive restriction on the production technology we consider here is quite similar to the proportionality restriction (linear regression parameters are assumed proportional to each other) as considered by Chamberlain (1977a) in a traditional "reduced form" error-in-variables model. In contrast, in the more general case of a Cobb-Douglas function with unknown scale:

$$\ln \theta_1 = \gamma_{0,\theta} \ln \theta_0 + \gamma_{0,I} \ln I_0, \tag{16}$$

where $\gamma_{0,\theta}$ and $\gamma_{0,1}$ are free parameters and do not sum to 1, the structural parameters in equation 16 $\gamma_{0,\theta}$ and $\gamma_{0,I}$ are not point identified because there would be four total unknown parameters $\gamma_{0,\theta}, \gamma_{0,I}$, $\mu_{1,m}$ and $\lambda_{1,m}$ and only three regression coefficients β_0,β_1, β_2.

A similar issue arises if we omit the investment input altogether, and the technology takes the form of a simple panel AR(1) process for the latent stock of skills:

$$\ln \theta_1 = \gamma_0 \ln \theta_0$$

Substituting measures as above, we have

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}\gamma_0 \widetilde{Z}_{0,m} + (\epsilon_{1,m} + \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}) \tag{17}$$

In this case, an additional assumption is required to separately identify the factor loading $\lambda_{1,m}$ of the measurement equation from the primitive production function parameter γ_0.

4.5 Restrictions Implied by Re-Normalization: Mean Log-Stationarity

Our previous results show that re-normalization restrictions are over-identifying when the production technology already has a known location and/or scale. We next show that in these cases, re-normalizing the model can also impose important biases in the estimation of the primitive production function parameters.

First, we show that because re-normalization imposes mean log-stationarity it restricts the *dynamic* relationships in skill development:

Remark 2 *Re-normalization (i) restricts the permissible production technologies to those that respect mean log-stationarity.*

As an example, consider again the class of constant elasticity of substitution (CES) technologies defined in (3). The stock of period $t = 1$ skills are given by

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

where $\gamma_0 \in (0, 1)$ and $\phi_0 \to 0$ is the log-linear, Cobb-Douglas special case. Recall that the initial conditions are normalized such that $E(\ln\theta_0) = 0$ and $E(\ln I_0) = 0$ and we identify the joint distribution of initial latent skills and investment, which we define as $G_0(\theta_0, I_0)$.

Consider the dynamics in skill development. Given the CES technology, the mean of log skills in the next period, say period 1, is given by

$$E(\ln\theta_1) = \begin{cases} \gamma_0 E\left(\ln\theta_0\right) + (1-\gamma_0)E\left(\ln I_0\right) & \text{if} \quad \phi_0 = 0 \\ \dfrac{1}{\phi_0}\displaystyle\int \ln\left((\gamma_0\theta_0^{\phi_0} + (1-\gamma_0)I_0^{\phi_0})\right) dG_0(\theta_0, I_0) & \text{if} \quad \phi_0 \neq 0 \end{cases}$$

The log-linear production technology ($\phi_0 \to 0$, Cobb-Douglas) is always consistent with re-normalization for any distribution of skills and investments in the initial period, $G_0(\theta_0, I_0)$, with $E(\ln\theta_0) = 0$ and $E(\ln I_0) = 0$ (as imposed in the initial period normalization). For at least some non-log-linear technologies (with $\phi_0 \neq 0$), re-normalization (i) may not hold and mean log skills could grow or decline: $E(\ln\theta_0) \neq E(\ln\theta_1)$. Re-normalization (i) is therefore not without loss of generality in an environment where the dynamics in children's skills are generated by a non-log-linear production technology. These technologies can imply either growth or decline in mean log skills over time, and the dynamics depend on the curvature of the production function as well as on the joint distribution of initial period skills and investments.

4.6 Restrictions Implied by Re-Normalization: Constant Factor Loading

The second part of the re-normalization definition, re-normalization (ii), is that the factor loading for one measure each period is normalized to be 1, $\lambda_{t,1} = 1$ for all t, where $m = 1$ is the (arbitrarily chosen) normalizing measure. This restricts skills to "load onto" different measures in each period, where each of these normalizing measures can have different scales. We have previously shown that if the production technology has a known location and scale, then fixing the scale again through re-normalization is not necessary, and the factor loadings $\lambda_{t,m}$ for all $t > 0$ and all m can be identified without fixing *any* of the factor loadings for periods after the initial period. In addition to being unnecessary and over-identifying, we show that this restriction on the factor loadings could also bias the estimation of production function parameters.

Remark 3 *Even when the technology respects mean log-stationarity, re-normalization (ii) (constant factor loadings) can bias estimation of production function parameters.*

Consider again the following simple example of a Cobb-Douglas production technology:

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

where $\gamma_0 \in (0, 1)$. This production technology is a KLS function and yields a path of skills which are always mean log-stationary. This function is therefore consistent with the restrictions imposed by re-normalization (i): $E(\ln \theta_1) = E(\ln \theta_0)$ for any $\gamma_0 \in (0, 1)$.

The bias in the estimation depends on the particular estimator. Consider the intuitively appealing idea of using the covariance between a log skill measure in period 1 $Z_{1,m}$ and log investment in period 0 $\ln I_0$. This moment captures the input-output relationship and could be plausibly used to identify and estimate the production function parameter γ_0. To simplify the derivation, we further assume the initial conditions are such that $Cov(\ln \theta_0, \ln I_0) = 0$, $Var(\ln \theta_0) = 1$ and $Var(\ln I_0) = 1$. The covariance in the skill measure and observed investment is given by

$$Cov(Z_{1,m}, \ln I_0) = \lambda_{1,m} Cov(\ln \theta_1, \ln I_0)$$

Substituting the production technology and solving for γ_0, we have

$$\gamma_0 = 1 - \frac{Cov(Z_{1,m}, \ln I_0)}{\lambda_{1,m}}$$

This expression indicates that the primitive parameter of the production technology γ_0 is a function of the covariance observed in data *and* the factor loading $\lambda_{1,m}$ for the period 1 measure. The re-normalization constant factor loading assumption imposes that $\lambda_{1,m} = 1$, and therefore under re-normalization, the primitive parameter is identified as $\gamma_0 = 1 - Cov(Z_{1,m}, Z_{0,m})$.

To see the problem with this approach, consider a new measure indexed m', which is a "scaled" version of the original measure m: $Z_{1,m'} = \delta Z_{1,m}$. Using this measure in the estimation implies a new value for the production function parameter: $\gamma_0' = 1 - Cov(Z_{1,m'}, Z_{0,m}) \neq \gamma_0$. We conclude that under re-normalization, the estimated γ_0 value depends on the scale of the measure used: different measures yield different estimates of the technology parameters. This is of course not an attractive feature of the estimator.

As described above (Section 4.3), in the case with a KLS production technology, a simple alternative estimator allows estimation of γ_0 which is invariant to the scale of the period 1 measure and the factor loading. Consider following the derivation above (12) and constructing the following ratio using the measure $Z_{1,m'} = \delta Z_{1,m}$:

$$\begin{aligned}\frac{Cov(Z_{1,m'}, \ln I_0)}{Cov(Z_{1,m'}, (\ln I_0)^2)} &= \frac{\delta Cov(Z_{1,m}, \ln I_0)}{\delta Cov(Z_{1,m}, (\ln I_0)^2)} \\ &= \frac{\delta \lambda_{1,m} Cov(\ln \theta_1, \ln I_0)}{\delta \lambda_{1,m} Cov(\ln \theta_1, (\ln I_0)^2)} \\ &= \frac{Cov(\ln \theta_1, \ln I_0)}{Cov(\ln \theta_1, (\ln I_0)^2)}\end{aligned}$$

Taking the ratio of covariances has eliminated the factor loading $\lambda_{1,m}$ and the "scaling" factor δ. This equation can be solved to identify γ_0 and can form an estimator for γ_0 which is *invariant* to the scale of the measures. This expression makes clear that because the production technology already has a known scale, we can identify the technology without re-normalizing the scale of the latent variables and imposing restrictions on the measurement equations. Imposing re-normalization and $\lambda_{1,m} = 1$ and using this restriction to estimate γ_0 is not only unnecessary, but can also therefore bias the resulting estimator if the true factor loading is not exactly 1. We examine this type of bias in more detail in the Monte Carlo simulations we present below.

4.7 Age-Standardization

One common approach to dealing with various skill measures which have different scales and locations is to "age-standardize" the measures. We show that this approach does not resolve the issues we have raised above.

Our previous measures $Z_{t,m}$ can be considered "raw" measures, with the mean and variance of the raw measure unrestricted. Using the raw measures, researchers often form age-standardized measures $S_{t,m}$:

$$S_{t,m} = \frac{Z_{t,m} - E(Z_{t,m})}{Var(Z_{t,m})^{1/2}} \tag{18}$$

$S_{t,m}$ has mean 0 and standard deviation 1, by construction.[8]

Following the same measurement model as above, we can write the age-standardized measure as a linear measure of the underlying latent skills:

[8] For example, to fix ideas, $Z_{t,m}$ could be a test score measure of skills with 76 items. $Z_{t,m}$ provides the number of questions the child answered correctly on the test, ranging from 0 (no questions answered correctly) to 76 (all questions answered correctly). $Z_{t,m}$ has a mean $E(Z_{t,m}) = 32$, and a standard deviation of $V(Z_{t,m})^{1/2} = 11$. In this case, the standardized measure would be constructed as $S_{t,m} = (Z_{t,m} - 32)/11$, where $S_{t,m}$ has mean 0 and standard deviation 1 by construction.

$$S_{t,m} = \mu_{S,t,m} + \lambda_{S,t,m} \ln \theta_t + \epsilon_{S,t,m} \tag{19}$$

where the measurement parameters are can be written in terms of the original parameters for the raw measure:

$$\mu_{S,t,m} = -\lambda_{S,t,m} E(\ln \theta_t),$$
$$\lambda_{S,t,m} = \frac{\lambda_{t,m}}{V(Z_{t,m})^{1/2}}, \text{and}$$
$$\epsilon_{S,t,m} = \epsilon_{t,m}/V(Z_{t,m})^{1/2}.$$

A key question is whether using age-standardized measures $S_{t,m}$, rather than the raw measures $Z_{t,m}$, would then imply that the re-normalization assumption holds without any loss of generality. We find this is not the case, as we summarize in the following remark:

Remark 4 *Age-standardized measures do not necessarily imply re-normalization.*

Note first that while $E(S_{t,m}) = 0$ by construction, this does not imply that $E(\ln \theta_t) = 0$ for any period. It is important to distinguish the sample construction of measures, which can be constructed to be mean 0 in the sample, from the latent distribution of skills, which can evolve dynamically to have a non-zero mean. Second, the factor loadings on the standardized measure, $\lambda_{S,t,m}$, are in general not equal to 1, and $\lambda_{S,t,m}$ can vary over time as the variance of latent skills $V(\ln \theta_t)$, the factor loading on the original raw measure $\lambda_{t,m}$, and measurement error variance $V(\epsilon_{t,m})$ vary over time. Age-standardization techniques therefore do not resolve the issues raised above with re-normalization. We directly evaluate this approach in the Monte Carlo exercises and show that estimators using re-normalization with age-standardized measures are still biased, although the form of the bias using the age-standardized measures can be different from that imposing re-normalization on the raw measures directly.

4.8 Anchoring

Cunha et al. (2010) consider "anchoring" the latent skills to variables which might be particularly meaningful from an economic or policy perspective. The idea is that latent skills should be anchored to some adult outcome (e.g. adult earnings or adult schooling) over which we might construct some sense of individual welfare. Their approach uses an equation which relates some adult outcome Y, e.g. adult earnings, to final period latent skills of children at the "terminal" age T (e.g. age $T = 14$ in their framework):

$$Y = \mu_A + \alpha_A \ln \theta_T + \epsilon_Y \tag{20}$$

where $E(\epsilon_Y) = 0$ and $\ln \theta_T$ and ϵ_Y are independent.

Using these anchoring parameters, μ_A and α_A, which are specific to the adult outcome Y, we can then relate "unanchored" latent skills (what we have to this point denoted θ_t) to anchored skills (which we now denote as $\theta_{A,t}$).

$$\ln \theta_{A,t} = \mu_A + \alpha_A \ln \theta_t \tag{21}$$

Inverting the function, we can also write unanchored skills as a function of anchored skills:

$$\ln \theta_t = -\frac{\mu_A}{\alpha_A} + \frac{1}{\alpha_A} \ln \theta_{A,t} \tag{22}$$

And, we can re-formulate the production technology in terms of anchored skills:

$$\ln \theta_{t+1} = -\frac{\mu_A}{\alpha_A} + \frac{1}{\alpha_A} \ln f_t(e^{\mu_A + \alpha_A \ln \theta_t}, I_t), \tag{23}$$

where the level of unanchored skills in period t is given by $\theta_t = e^{\ln \theta_{A,t}} = e^{\mu_A + \alpha_A \ln \theta_t}$. In this setting, we follow Cunha et al. (2010) and assume re-normalization continues to hold for the unanchored skills, but there are no explicit conditions on the anchored skills.

Formulating the technology in terms of anchored skills changes the interpretation of the production function parameters, as they are now in terms of the anchored skills rather than unanchored skills. In particular, the μ_A and α_A parameters can change the curvature of the production technology, possibly violating the assumption of mean log-stationarity ($E(\ln \theta_1) = \cdots = E(\ln \theta_T) = 0$). We summarize this finding in the following remark:

Remark 5 *Re-normalization can restrict the anchoring and production function parameters.*

Consider again the log-linear, Cobb-Douglas function, but now with anchoring:

$$\ln \theta_1 = -\frac{\mu_A}{\alpha_A} + \frac{1}{\alpha_A} [\gamma_0(\mu_A + \alpha_A \ln \theta_0) + (1 - \gamma_0) \ln I_0]$$

$$= \frac{\mu_A}{\alpha_A}(\gamma_0 - 1) + \gamma_0 \ln \theta_0 + \frac{1-\gamma_0}{\alpha_A} \ln I_0$$

With $\gamma_0 \in (0,1)$, $E(\ln \theta_0) = 0$ and $E(\ln I_0) = 0$, the re-normalization condition (i) $E(\ln \theta_1) = 0$ holds if and only if $\mu_A = 0$. However, this can contradict the anchoring equation (20) whenever $E(Y) \neq 0$. Because re-normalization imposes $E(\ln \theta_T) = 0$, this implies that $\mu_A = E(Y)$. Because we interpret the anchoring as an attempt to give some specific scale and location to children skills, the anchoring measures could in general have a non-zero mean.[9] Of course, one can always de-mean the adult outcome Y, so that $E(Y) = 0$ by construction. But because the technologies are in general non-linear, with transformations such as these the anchored latent skills lose their specific meaning derived from the particular location and scale of the adult outcome.

5 Monte Carlo Exercises

To support our analytic results, we simulate some simple versions of the child development model to show that the remarks we derived hold in a simple data simulation and to quantify the potential biases in estimation.

5.1 Two Period Cobb-Douglas Model

The first example assumes the technology of skill formation is of the Cobb-Douglas form where the mean log-stationarity restriction, re-normalization (i), implicitly holds. In this example, we focus on the implications of the factor loading restriction, re-normalization (ii).

5.1.1 Data Generating Process

We consider a two period model $T = 2$, where the skill production technology is given by

$$\ln \theta_1 = \gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0, \tag{24}$$

and $\gamma_0 \in (0,1)$ is the production function parameter we want to estimate. The initial log children's skills and initial log investments are drawn from a Normal distribution:

[9]One example of an anchor in Cunha et al. (2010) is years of schooling. The sample mean of this variable in their sample is 13.38 years, implying that the estimate of μ_A would be 13.38. However, when Cunha et al. (2010) construct the anchor for log skills, they impose $\mu_A = 0$ (private email correspondence with Flavio Cunha).

$$\ln \theta_0 \sim N(0, \sigma_\theta^2) \ , \ \ \ln I_0 \sim N(0, \sigma_I^2), \tag{25}$$

where $\ln \theta_0$ and $\ln I_0$ are assumed independent. In these examples, we assume investment is also measured with error, and there are three latent variables: θ_0, θ_1, I_0. We have three measures for $\ln \theta_0$ $(Z_{0,1,\theta}, Z_{0,2,\theta}, Z_{0,3,\theta})$, three measures for $\ln \theta_1$ $(Z_{1,1,\theta}, Z_{1,2,\theta}, Z_{1,3,\theta})$, and three measures for $\ln I_0$ $(Z_{0,1,I}, Z_{0,2,I}, Z_{0,3,I})$. The measures take the following form:

$$Z_{t,m,\theta} = \mu_{t,m,\theta} + \lambda_{t,m,\theta} \ln \theta_t + \epsilon_{t,m,\theta}, \tag{26}$$

$$Z_{0,m,I} = \mu_{0,m,I} + \lambda_{0,m,I} \ln I_0 + \epsilon_{0,m,I}, \tag{27}$$

with $\epsilon_{t,m,\theta} \sim N(0, \sigma^2_{t,m,\theta})$ and $\epsilon_{0,m,I} \sim N(0, \sigma^2_{0,m,I})$ for both periods $t = \{0, 1\}$ and all the measures $m = \{1, 2, 3\}$. The measurement errors are assumed to be independent of each other. The full set of parameters we use are listed in tbe Appendix.

5.1.2 Estimation

We consider three estimators for γ_0. In order to eliminate differences in estimation results due to the other less central aspects of the estimation, all of the estimators are based on simulated method of moments. All three estimators impose the initial period normalization and compute the initial conditions in the same way, as described above. The estimators only differ in the assumptions imposed on the subsequent period.

- **Estimator 1 (Re-Normalization)** imposes re-normalization (ii): $\lambda_{1,1,\theta} = 1$. We then compute the remaining factor loadings for period 1 as follows

$$\lambda_{1,2,\theta} = \frac{Cov(Z_{1,2,\theta}, Z_{1,3,\theta})}{Cov(Z_{1,1,\theta}, Z_{1,3,\theta})} \text{ and } \lambda_{1,3,\theta} = \frac{Cov(Z_{1,3,\theta}, Z_{1,2,\theta})}{Cov(Z_{1,1,\theta}, Z_{1,2,\theta})}$$

- **Estimator 2 (Re-Normalization and Standardized Measures)** first age-standardizes all of the measures and then imposes re-normalization on the standardized measures following Estimator 1. Estimator 2 is then the same as Estimator 1 after age-standardizing the measures.

- **Estimator 3 (Initial Period Only Normalization)** does not impose the re-normalization assumption. Instead, for this estimator, we compute measurement parameters consistently with any technology parameter (γ_0) as follows:

$$\lambda_{1,m,\theta} = \frac{Cov(Z_{1,m,\theta}, Z_{0,1,\theta})}{Cov(\ln\theta_1, \ln\theta_0)} \text{ for all } m = 1, 2, 3, \tag{28}$$

  where $\ln\theta_1$ and hence $\lambda_{1,m,\theta}$ depend on the γ_0 parameter.

To isolate the role of the re-normalization restrictions, all of the simulated method of moments estimators use the same set of moments based on covariances between skill measures in $t+1$ and measures of inputs in period t (see Appendix). Each estimator for γ_0 minimizes the sum of the quadratic deviation between data and simulated moments, weighting each moment equally. We simulate a dataset of 1,000 observations, using 1,000 simulation draws for each estimator. We use a robust grid search over the parameter space to compute the estimator.

5.1.3 Results

In the first exercise (see Figure 1), we vary the true factor loading for one of the first period measures, $\lambda_{1,1,\theta}$. The true value of γ_0 is fixed at 0.5. For each value of the factor loading, we compute the estimate for the production parameter γ_0.

Several results are of note. First, the estimate of γ_0 from Estimator 3 (Initial Period Only Normalization), our preferred estimator, is invariant to the factor loading $\lambda_{1,1,\theta}$. This is because Estimator 3 computes the factor loading to be consistent for any value of the technology parameter γ_0. Estimator 3 is able to recover the true γ_0 estimate of 0.5.[10]

On the other hand, the estimate of γ_0 from Estimator 1 (Re-Normalization) varies depending on the true measurement parameter $\lambda_{1,1,\theta}$. Only when re-normalization restriction actually holds ($\lambda_{1,1,\theta} = 1$), do we see that that the γ_0 estimate is equal to the true value of γ_0. Estimator 2 uses the re-normalization assumption but first age-standardizes the measures. Figure 1 shows that estimates of γ_0 using age-standardized measures and re-normalization are also biased.[11]

[10] Note that because of the discrete grid search and finite number of simulation draws used to compute the simulated method of moments estimator, our estimates are not in all cases exactly equal to the true value. But the deviations of the estimate from the true value are very small.

[11] A fourth estimator one could consider is Estimator 3 (Initial Period Only Normalization) using age-standardized measures. In results not shown, but which directly follow our analytic results, this estimator is also able to recover the primitive parameter because this estimator is invariant to the location and scale of the period 1 measures.

In Figure 2 we show the results from a second simulation in which we fix the value of factor loading at a value different from 1 ($\lambda_{1,1,\theta} = 0.65$) and vary the true value of the production function parameter γ_0. It is clear that Estimator 3 (Initial Period Only Normalization) is able to recover the true value of γ_0 as we vary the true value of γ_0 over all possible values ($\gamma_0 \in (0,1)$). For Estimators 1 and 2, based on imposing re-normalization restrictions, these estimators cannot recover the true value of the primitive production function.

5.2 Two Period General CES Model

5.2.1 Data Generating Process

In this next set of exercises, we consider estimating a more general CES technology of skill formation:

$$\theta_1 = (\gamma_0 \theta_0^{\phi_0} + (1-\gamma_0) I_0^{\phi_0})^{1/\phi_0} \tag{29}$$

In this specification, there are two unknown production function parameters we would like to estimate, $\gamma_0 \in (0,1)$ and $\phi_0 \in (-\infty, 1]$.

5.2.2 Estimation

In this exercise, we maintain both the initial conditions and the measurement equations as in the previous exercise. We compute the same simulated method of moments estimators as above. We also consider a fourth estimator:

- **Estimator 4 (Internally Consistent Estimator)** is the estimator which is internally consistent with re-normalization (i).

This estimator maintains the re-normalization assumption of mean log-stationarity, imposing $E(\ln \theta_0) = E(\ln \theta_1) = 0$. As we show above, for this class of production technologies, this restriction implies that the complementarity parameter must be $\phi_0 = 0$ (Cobb-Douglas, log-linear).[12]

[12] As we discuss above, Estimator 4 is not really an estimator at all in the traditional sense, but simply presented to emphasized that re-normalization biases the elasticity of substitution to a particular value.

5.2.3 Results

In these exercises, we assess the ability of the estimators to recover the complementarity parameter ϕ_0. Results are shown in Figures 3 and 4. In Figure 3 we vary the factor loading parameters $\lambda_{1,1,\theta}$ and fix the true value of the complementarity parameter at $\phi_0 = 0.5$ (elasticity of substitution 2) and share parameter at $\gamma_0 = 0.7$. As in the first exercise, Estimator 3 (Initial Period Only Normalization) is invariant to the factor loading $\lambda_{1,1,\theta}$ value and is able to recover the true complementarity value of 0.5. Estimators 1 and 2 based on re-normalization, using either the raw or age-standardized measures, are generally biased.

Figure 4 conducts the reverse exercise and fixes the measurement factor loading at $\lambda_{1,1,\theta} = 0.65$ and varies the complementarity parameter ϕ_0 (keeping the factor share at $\gamma_0 = 0.7$). Again, Estimator 3 (Initial Period Only Normalization) recovers the true production function parameters, but the re-normalization based estimators (Estimators 1 and 2) are biased. Estimator 4 (Internally Consistent Estimator) maintains the value of $\phi_0 = 0$ as this estimator is constructed to satisfy the re-normalization assumption of mean log-stationarity. This estimator too, while being internally consistent, is clearly biased for true values of complementarity $\phi_0 \neq 0$.[13]

6 Conclusion

Dynamic latent factor models are an important tool for modeling the dynamics of skill development and incorporating the many varied and imperfect measures of skills available. As is well known, because latent variables have no natural units, these latent factor models require a normalization to fix the scale and location of the latent variables. However, additional normalizations beyond what is required are restrictions which can reduce the generality of the model. We show that the now common approach of "re-normalizing" latent skills each period in dynamic models of child development–treating skills in each period as separate factors–is both unnecessary in typical cases where the production technology already has a known location and scale and can cause important biases. Emphasizing the over-identification of these restrictions, we show that simple estimators which do not impose these re-normalizations restrictions can in fact identify the underlying parameters. We demonstrate our an-

[13]It should be noted that because of the moments selection we made, we are not fully assessing the implications of re-normalization (i) on the bias for Estimator 1 and 2. This is because the covariances we use in the moment conditions do not depend on the location of the latent skills, which is restricted by re-normalization (i). Other estimators using different moments or those using Maximum Likelihood Estimation, may produce even more biased estimators.

alytic results in a series of Monte Carlo experiments. In our related work Agostinelli and Wiswall (2016) we expand on these results, and characterize the conditions for identification of various kinds of skill development technologies.

References

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. *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.

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.

Gary Chamberlain. Education, income, and ability revisited. *Journal of Econometrics*, page 241257, 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*, 16:422449, 1975.

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.

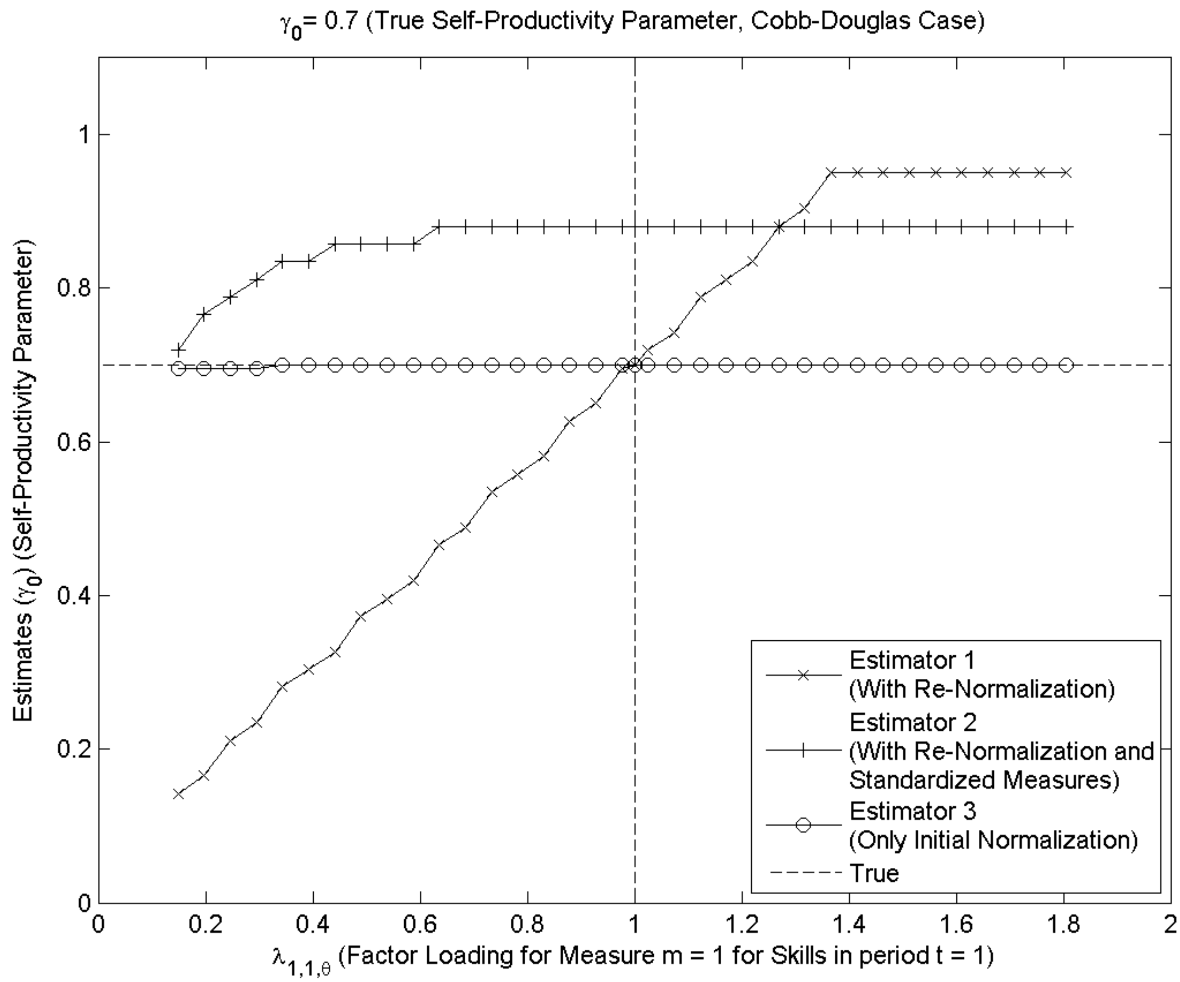
Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

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

Karl G. Jöreskog and Arthur S. Goldberger. Estimation of amodel withmultiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70:631639, 1975.

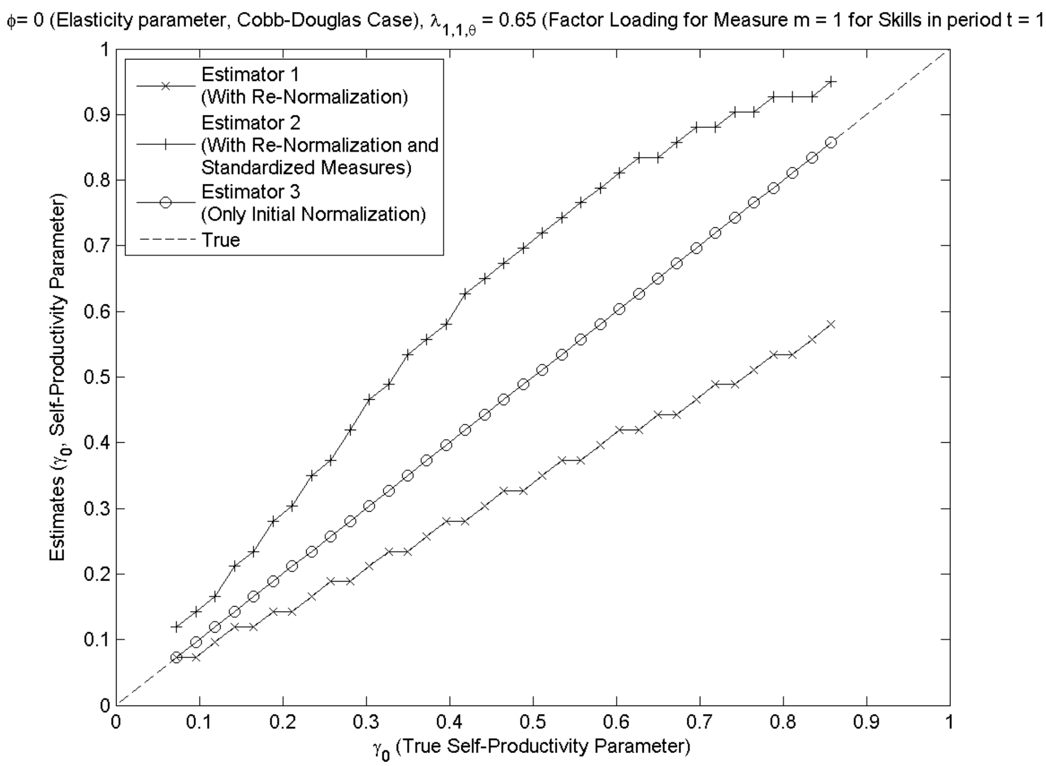
Ronni Pavan. On the production of skills and the birth order effect. *Working Paper*, 2015.

Figure 1: Monte Carlo Results for Estimates of γ (varying λ)



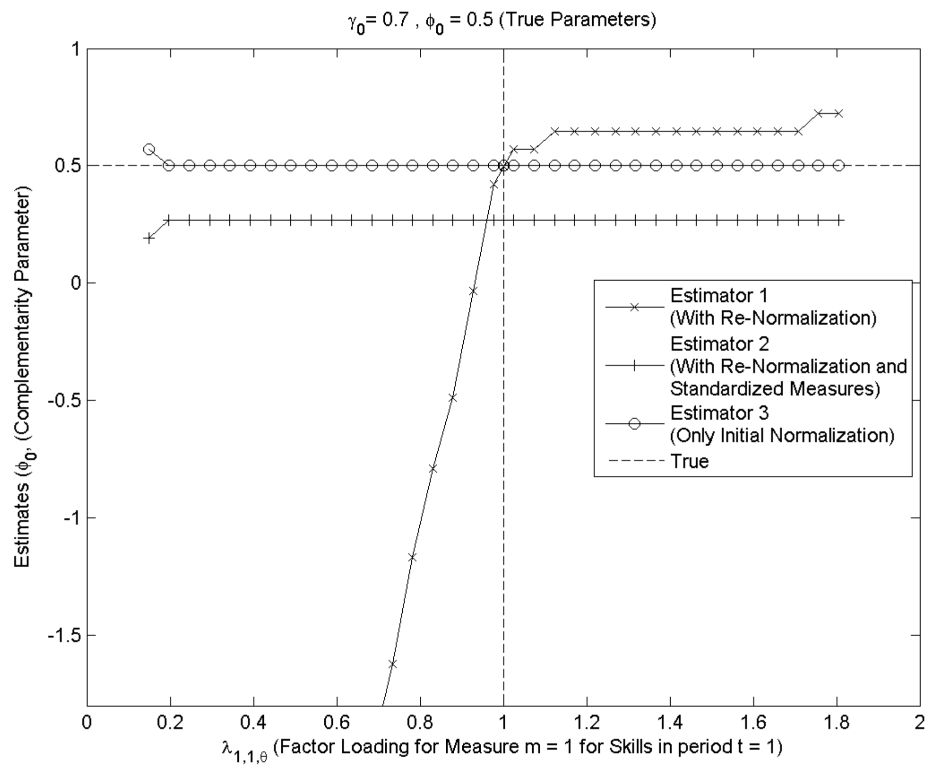
Notes: The dashed line represents the true value of the parameter. The true value and the Estimator 3 value are exactly equal and hence these lines overlap in the figure.

Figure 2: Monte Carlo Results for Estimates of γ



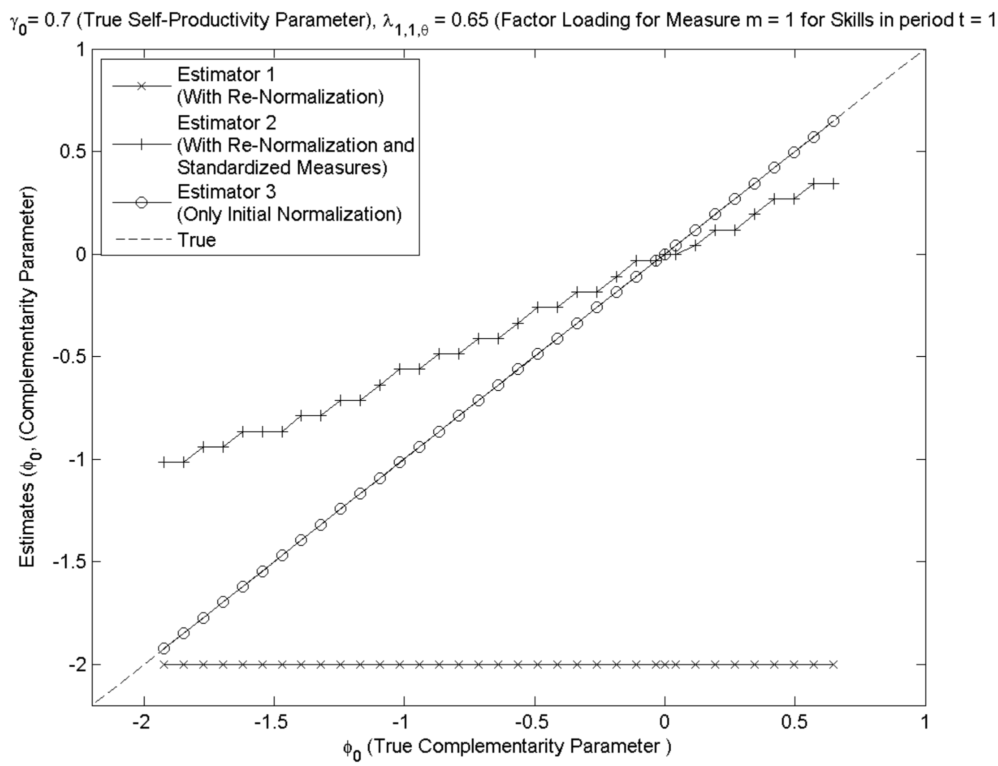
Notes: The 45 degree dashed line displays the true parameter. The true value and the Estimator 3 value are exactly equal and hence these lines overlap in the figure.

Figure 3: Monte Carlo Results for Estimates of ϕ (varying λ)



Notes: The dashed line represents the True value of the parameter. The true value and the Estimator 3 value are exactly equal and hence these lines overlap in the figure.

Figure 4: Monte Carlo Results for Estimates of ϕ



Notes: The 45 degree dashed line displays the true parameter. The true value and the Estimator 3 value are exactly equal and hence these lines overlap in the figure.

APPENDIX

A Monte Carlo Details

The moments used in all estimators is given by

$$
\begin{aligned}
Cov(Z_{1,m,\theta}, Z_{0,m',\theta}) &= \lambda_{1,m,\theta}\lambda_{0,m',\theta}Cov(\ln\theta_1, \ln\theta_0) \\
Cov(Z_{1,m,\theta}, Z_{0,m',I}) &= \lambda_{1,m,\theta}\lambda_{0,m',I}Cov(\ln\theta_1, \ln I_0) \\
Cov(Z_{1,m,\theta}, (Z_{0,m',\theta})^2) &= 2\mu_{0,m',\theta}\lambda_{0,m',\theta}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln\theta_0)+ \\
&\quad \lambda_{1,m,\theta}\lambda_{0,m',\theta}^2 Cov(\ln\theta_1, (\ln\theta_0)^2) \\
Cov(Z_{1,m,\theta}, (Z_{0,m',I})^2) &= 2\mu_{0,m',I}\lambda_{0,m',I}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln I_0)+ \\
&\quad \lambda_{1,m,\theta}\lambda_{0,m',I}^2 Cov(\ln\theta_1, (\ln I_0)^2) \\
Cov(Z_{1,m,\theta}, Z_{0,m'',\theta}\cdot Z_{0,m',I}) &= \mu_{0,m',I}\lambda_{0,m'',\theta}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln\theta_0)+ \\
&\quad \mu_{0,m'',\theta}\lambda_{0,m',I}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln I_0)+ \\
&\quad \lambda_{0,m'',\theta}\lambda_{0,m',I}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln\theta_0\cdot\ln I_0)
\end{aligned}
$$

Table A-1: Parameters for the Monte Carlo Exercises

Parameter		Value	
σ_θ	(Standard deviation initial skills)	1	
σ_I	(Standard deviation initial investment)	1	
$\mu_{t,m,\theta}$	(Constant for measurement equation)	1	$\forall\ t=\{0,1\},\ m=\{1,2,3\}$
$\mu_{0,m,I}$		1	$\forall\ m=\{1,2,3\}$
$\lambda_{0,m,\theta}$	(Factor loading for measurement equation)	1	$\forall\ m=\{1,2,3\}$
$\lambda_{1,m,\theta}$		1	$\forall\ m=\{2,3\}$
$\lambda_{0,m,I}$		1	$\forall\ m=\{1,2,3\}$
$\sigma_{0,m,\theta}$	(Standard deviation for measurement error)	0.15	$\forall\ m=\{1,2,3\}$
$\sigma_{1,m,\theta}$		0.15	$\forall\ m=\{1,2,3\}$
$\sigma_{0,m,I}$		0.15	$\forall\ m=\{1,2,3\}$

Notes: Table A-1 shows the values for the model parameters used in the Monte Carlo exercise in Section 5.

The full list of Monte Carlo parameters is given in the following table: