Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development

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

March, 2017

JEL Classification C38, J13

Keywords: Latent Factor Models; Measurement Error; Production Functions; Child Development; Skill Formation

*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, Western Ontario, Wisconsin, and Chicago for helpful comments. Previous version circulated as a NBER Working Paper. We are responsible for all errors.

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. Research following Cunha et al. (2010) estimate the production technology of skill formation by simultaneously "re-normalizing" the location and scale of the latent variables each period and assuming the production technology takes a restricted Constant Elasticity of Substitution (CES) form, which already has a known location and scale (do not need to be estimated). We show that simultaneously re-normalizing the model each period *and* imposing a particular parametric known location and scale technology places unnecessary over-identifying restrictions on the estimator and can bias the resulting estimates. In a series of Monte Carlo experiments replicating the Cunha et al. (2010) model and estimator, we show the biases can be substantial, and that simple alternative estimators, which do not impose these restrictions, can recover the underlying primitive parameters of the production technology.

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 develops 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. In a seminal contribution, Cunha et al. (2010) provide theoretical identification results proving non-parametric identification of the production technology up to a chosen normalization.

In the estimation of their model, Cunha et al. (2010) select a particular parametric class of constant returns to scale (CRS) and constant elasticity of substitution (CES) functions. Cunha et al. (2010) also estimate their model by "re-normalizing" the latent variables every period, imposing that the latent (log) stock of skills is mean 0 in every period, and fixing the scale of the latent skill stock in every period to a different measure. We show that the production function chosen by Cunha et al. (2010) already imposes a particular known location and scale and is part of a class of "known location and scale" functions analyzed in Agostinelli and Wiswall (2016). Re-normalizing the model is therefore unnecessary because the location and scale of the latent variables is already implicitly fixed by their chosen production function.

As shown in the identification theory of Cunha et al. (2010), *in principle*, the re-normalization approach is without loss of generality if one allows for a general enough production technology. However, we show that the parametric known location and scale CES technology estimated by Cunha et al. (2010) is already sufficiently restricted to allow identification without re-normalization. Simultaneously re-normalizing the model *and* imposing the particular parametric functional forms

[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).

for the technology is therefore over-identifying.

Not only are the re-normalization restrictions imposed in the estimation of Cunha et al. (2010) over-identifying, we also show that they can impose empirically relevant biases on the estimation of the production function. First, the re-normalization restrictions on the location of the latent variables implicitly impose a mean log-stationary restriction on the dynamics of the latent skill stock, and this restriction implies that the estimated technology must produce a mean log-stationary path of latent skill stocks. This restriction is consistent with log-linear (Cobb-Douglas) technologies but rules out many other types of production technologies, including at least some CES technologies displaying different degrees of substitutability than the Cobb-Douglas case. Second, re-normalization implicitly impose a restriction through re-normalizing the scale of the latent variables each period. We show that even in the case where the mean-log-stationary restriction is correctly imposed, these restrictions on the scale of the latent variables can still bias the estimates of the production function parameters. The resulting estimator linking the scale of the latent variables to different measures each period then makes the estimator of the production function primitives sensitive to the particular measures chosen. Simple alternative estimators for the Cunha et al. (2010) assumed technologies can be constructed which are invariant to the scale of the measures chosen, up to one initial period normalization.[3]

The remainder of the paper is organized as follows. We first present a simplified version of the Cunha et al. (2010) modeling assumption imposed in their estimation. We show that although some normalization is of course necessary, their estimation imposes over-identifying restrictions which can bias their estimator. 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 of the Cunha et al. (2010)'s assumed production technology.

[3]Please note: Following the circulation of earlier drafts of this paper, the authors of Cunha et al. (2010) have informed us that contrary to the estimation assumptions written in the paper (see p. 905, in particular, which indicates that the latent distribution of log skills is assumed to be mean 0 in all periods), they did not in fact impose the first of the two re-normalization restrictions in their estimation algorithm and allowed the mean of the latent variables to be unrestricted. We continue to analyze this case following the text of the Cunha et al. (2010) paper in order to provide a guide to future research. Cunha et al. (2010) have informed us that their estimation did in fact impose the second re-normalization restriction on the scale of the latent variables, and the issues we discuss here with respect to this restriction therefore still apply.

2 Cunha et al. (2010) Model

In this section we present a stylized model of skill formation and measurement, a simplified version of the influential model estimated in Cunha et al. (2010).

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 (birth in Cunha et al. (2010)) and $t = T$ is the final period of childhood (age 16 in Cunha et al. (2010)). 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 $G_1, \ldots, G_T$.

In their estimation, Cunha et al. (2010) assume the production technology takes the following Constant Returns to Scale (CRS), Constant Elasticity of Substitution (CES) form:

$$\theta_{i,t+1} = (\gamma_t \theta_{i,t}^{\phi_t} + (1-\gamma_t) I_{i,t}^{\phi_t})^{1/\phi_t}. \tag{2}$$

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 $1/(1 - \phi_t)$.

Given the later identification analysis is in terms of log variables, for convenience, we work with the log form of the production function, which we write as

$$\ln \theta_{i,t+1} = \ln[(\gamma_t \theta_{i,t}^{\phi_t} + (1-\gamma_t) I_{i,t}^{\phi_t})^{1/\phi_t}].$$

For our analysis here, the key characteristic of the Cunha et al. (2010) production function is that it has "known location and scale" (KLS), as defined in Agostinelli and Wiswall (2016). The log production technology (2) belongs to the class of KLS technologies because for any inputs $I_{i,t} = \theta_{i,t} = a > 0$, the output $\theta_{i,t+1} = a$. That is, for inputs which are known to be equal, we also know the output as well.[4] While the scale and location of the production function (2) 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.

In contrast, adding a free Total Factor Productivity (TFP) term and a unknown scale to the CES function now implies that the function no longer has a known scale and return to scale:

$$\theta_{i,t+1} = A_t(\gamma_t \theta_{i,t}^{\phi_t} + (1-\gamma_t) I_{i,t}^{\phi_t})^{\psi_t/\phi_t}. \tag{3}$$

Or in log-form:

$$\begin{aligned} \ln \theta_{i,t+1} &= \ln A_t + \psi_t \ln[(\gamma_t \theta_{i,t}^{\phi_t} + (1-\gamma_t) I_{i,t}^{\phi_t})^{1/\phi_t}] \\ &= \ln A_t + \psi_t \ln h_t(\theta_{i,t}, I_{i,t}). \end{aligned}$$

where $A_t > 0$ represents TFP, $\psi_t > 0$ is an unknown returns to scale parameter, and $h_t(\cdot)$ is the original Cunha et al. (2010) CRS CES function.[5] In log form, the location is determined by $\ln A_t$ and the scale by ψ_t. In this case, the production possibility set depends on the location and scale parameters (A_t,ψ_t). For any inputs $I_{i,t} = \theta_{i,t} = a > 0$, $\theta_{i,t+1} = A_t \cdot a^{\psi_t}$.

The Cunha et al. (2010) technology is a restricted form of (3) with $A_t = 1$ and $\psi_t = 1$ for all t. These restrictions implicitly fix the location and scale of the technology at known points, which then do not need to be estimated.

2.2 Measurement

Cunha et al. (2010) assume the stock of skills $\theta_{i,t}$ is not observed in data directly. We follow the estimation assumptions of Cunha et al. (2010) and assume a system of log linear measurement equations given by

[4] In this case, this result follows directly from the constant return to scale property. However, not all KLS technologies are necessarily CRS.

[5] Equivalently, one could write this function as $\theta_{i,t+1} = (\gamma_{1t}\theta_{i,t}^{\phi_t} + \gamma_{2t} I_{i,t}^{\phi_t})^{\psi_t/\phi_t}$, where $\gamma_{1t} + \gamma_{2t}$ does not equal a known constant.

$$Z_{i,t,m} = \mu_{t,m} + \lambda_{t,m} \ln \theta_{i,t} + \epsilon_{i,t,m} \tag{4}$$

The 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 $m = 1, \ldots, M_t$. The set of measures can differ period by period. $Z_{i,t,1}, \ldots, Z_{i,t,M_t}$ 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 homogeneous 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.[6] Also, note that sharing an index m is not intended to denote any relationship between measures across periods.

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 $\kappa_t \equiv E(X_t)$.

The theoretical identification results in Cunha et al. (2010) provide general conditions for identification with varying assumptions about the dependence of the measurement errors. In this paper, we are explicitly analyzing the assumptions Cunha et al. (2010) impose in the estimation. In the empirical specification in Cunha et al. (2010), the authors assume 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).[7]

Because latent skills are unobserved and have no natural units, some normalization is clearly necessary. Following Cunha et al. (2010), we normalize the initial

[6] 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.

[7] See Section 4.1 *Empirical Specification*, pp. 904-905. In their estimation, Cunha et al. (2010) impose further restrictions on the measurement errors and assume the marginal distributions of each $\epsilon_{t,m}$ are Normally distributed. We omit this assumption because it is not necessary to our analysis.

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

In this section we present the general identification problem following the empirical model estimated in Cunha et al. (2010).

3.1 Under-Identification

We analyze the under-identification of the model by first examining the general form of the production technology (with unknown location and scale) given by (3), in logs:

$$\begin{aligned}\ln\theta_{t+1} &= \ln A_t + \psi_t \ln[(\gamma_t \theta_{i,t}^{\phi_t} + (1-\gamma_t) I_{i,t}^{\phi_t})^{1/\phi_t}] \\ &= \ln A_t + \psi_t \ln h_t(\theta_t, I_t) \end{aligned} \tag{5}$$

Substituting the latent technology equation (5) into Cunha et al. (2010)'s assumed liner measurement equation (4), we have the following:

$$Z_{t+1,m} = \mu_{t+1,m} + \lambda_{t+1,m}[\ln A_t + \psi_t \ln h_t(\theta_t, I_t)] + \epsilon_{t+1,m}$$

Re-arranging,

$$\begin{aligned} Z_{t+1,m} &= (\mu_{t+1,m} + \lambda_{t+1,m} \ln A_t) + (\lambda_{t+1,m}\psi_t) \ln h_t(\theta_t, I_t) + \epsilon_{t+1,m} \\ &= a_t + b_t \ln h_t(\theta_t, I_t) + \epsilon_{t+1,m} \end{aligned} \tag{6}$$

From this expression, it is clear that we cannot separately identify the measurement parameters $\mu_{t+1,m}$, $\lambda_{t+1,m}$ from the production function parameters A_t, ψ_t.

3.2 Observationally Equivalent Models

We next examine 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.

The first model assumes the production technology is the restricted CRS CES (as assumed in the estimation in Cunha et al. (2010)), but leaves the latent variables and measurement parameters free:

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

The second model leaves the production function free (A_t, ψ_t free production function parameters), but restricts the latent variables and measurement parameters:

Model 2 *(i)* $\ln(\theta_{t+1}) = \ln A_t + \psi_t \ln h(\theta_t, I_t)$
(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 and cannot be separately identified. The two models present observationally equivalent normalizations for the under-identified general formulation (6). Model 1 normalizes the technology (sets $A_t = 1$ and $\psi_t = 1$) 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.[8]

3.3 Re-Normalization

Our main argument is that although some normalization in this class of latent factor models is clearly necessary, the estimation in Cunha et al. (2010) "over-normalize" the model: Cunha et al. (2010) simultaneously estimate parametric production technologies that already have a known location and scale *and* "re-normalize" the latent

[8] 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: $\mu_{t,m} = E(Z_{t,m})$. There are of course other possible normalizations. One possibility is to replace Model 2 (ii) with $\mu_t = 0$ for all t.

variables. We show that this approach imposes over-identifying restrictions and these over-identifying restrictions can bias the estimation.

We define *re-normalization* as:

Definition 1 *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 the Cunha et al. (2010) estimation, re-normalization is a set of assumptions with empirical content and testable restrictions.

As described in more detail below (see Section 3.7 for a detailed explanation), these re-normalization assumptions are distinct from using age-standardized measures, measures constructed in the *sample* to have a 0 mean and standard deviation 1 in each period. Using age-standardized measures does *not* imply the re-normalization assumption holds: re-normalization implies particular restrictions on the model primitives.

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, in combination with the restricted CES production function used by CHS in their estimation, each biases the estimates of the production function parameters.

3.4 Re-Normalization is Over-Identifying with Restricted Technologies

We now turn to our first result. If the production functions characterizing the skill development technology already has a known location and scale, then further restrictions to fix the location and scale are unnecessary. We classify this over-identified model as Model 3:

Model 3 *Restricted CES with Re-Normalization*
(i) $\ln(\theta_{t+1}) = \ln h_t(\theta_t, I_t)$ *(*$A_t = 1$*,* $\psi_t = 1$*)*
(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 (6), simultaneously restricting the technology *and* the latent variables and measurement parameters. This is the model estimated in Cunha et al. (2010).[9] 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 CES production technology has a known location or scale.*

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

Our previous result shows that re-normalization restrictions are over-identifying when the production technology takes on the CES form assumed by Cunha et al. (2010). We next show that re-normalizing the model can also impose specific and 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.*

Given the CES technology, the mean of log skills in period 1 is given by

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

The log-linear production technology ($\phi_0 \to 0$, Cobb-Douglas) is always consistent with re-normalization (i) 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

[9] As noted in the footnote, the authors of Cunha et al. (2010) have informed us that in their actual estimation, they impose (iii) $\lambda_{t,m} = 1$ for all $t > 0$ but not (ii) $E(\ln(\theta_t)) = 0$ for all $t > 0$.

initial period normalization). There may be other combinations of CES parameters and initial conditions which also generate a path of skill stocks which are mean-stationary, but the set of permissible technologies is restricted. 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) as imposed in Cunha et al. (2010) is therefore not without loss of generality.

3.6 Restrictions Implied by Re-Normalization: Constant Factor Loading

The second part of the re-normalization restriction imposed by Cunha et al. (2010), 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 restriction forces skills to "load onto" different measures in each period, where each of these normalizing measures can have different scales. This implies that estimators such as Cunha et al. (2010) that impose these assumptions are therefore sensitive to the scaling of the measures used. Simple alternative estimators, which relax the re-normalization assumptions, avoid this issue and are *invariant* to the scale of the measures used.

We show this result using a special case of the Cunha et al. (2010) CES production technology, the Cobb-Douglas case, for a two period $t = 0, 1$ model:

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

where $\gamma_0 \in (0, 1)$. To isolate the importance of the second re-normalization issue, we choose this example because it respects mean log-stationarity.

We summarize the potential bias to re-normalizing the scale of the latent variables in the following remark:

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

Consider estimation of the unknown production function parameter γ_0 using the regression of a period 1 skill measure $Z_{1,m}$ on observed log investment in period 0 $\ln I_0$. To simplify the derivation, we 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 regression coefficient is the covariance in the skill measure and observed investment, and 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 $\gamma_0 = 1 - Cov(Z_{1,m}, \ln I_0)$.

To see the problem with this re-normalization 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 implies a new value for the production function parameter: $\gamma_0' = 1 - Cov(Z_{1,m'}, \ln I_0) \neq \gamma_0$. 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. The simple alternative estimators we explore in the Monte Carlo exercises we discuss below, based on the identification results in Agostinelli and Wiswall (2016), avoid this issue and are actually *invariant* to the arbitrary location and scale of the measures. The Appendix provides another formulation of the problem using an "errors-in-variables" setup.

3.7 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 $S_{t,m}$:

$$S_{t,m} = \frac{Z_{t,m} - E(Z_{t,m})}{Var(Z_{t,m})^{1/2}} \tag{7}$$

$S_{t,m}$ has mean 0 and standard deviation 1, by construction.[10]

[10]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

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

$$S_{t,m} = \mu_{S,t,m} + \lambda_{S,t,m} \ln \theta_t + \epsilon_{S,t,m} \tag{8}$$

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 although $E(S_{t,m}) = 0$ 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.

construction.

3.8 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):

$$Y = \mu_A + \alpha_A \ln \theta_T + \epsilon_Y \tag{9}$$

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 $\theta_{A,t}$).

$$\ln \theta_{A,t} = \mu_A + \alpha_A \ln \theta_t \tag{10}$$

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

$$\ln \theta_t = -\frac{\mu_A}{\alpha_A} + \frac{1}{\alpha_A} \ln \theta_{A,t} \tag{11}$$

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

$$\ln \theta_{t+1} = -\frac{\mu_A}{\alpha_A} + \frac{1}{\alpha_A} \ln f_t(e^{\mu_A + \alpha_A \ln \theta_t}, I_t), \tag{12}$$

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:

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

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 (9) 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.[11] 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.

4 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.

4.1 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).

4.1.1 Data Generating Process

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

[11] 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).

$$\ln\theta_1 = \gamma_0 \ln\theta_0 + (1-\gamma_0)\ln I_0, \tag{13}$$

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:

$$\ln\theta_0 \sim N(0,\sigma^2_\theta)\ ,\quad \ln I_0 \sim N(0,\sigma^2_I), \tag{14}$$

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:

$$Z_{t,m,\theta} = \mu_{t,m,\theta} + \lambda_{t,m,\theta}\ln\theta_t + \epsilon_{t,m,\theta}, \tag{15}$$

$$Z_{0,m,I} = \mu_{0,m,I} + \lambda_{0,m,I}\ln I_0 + \epsilon_{0,m,I}, \tag{16}$$

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 the Appendix.

4.1.2 Estimation

We consider three estimators for γ_0. In order to eliminate differences in the estimation results due to the other less central aspects of the estimation, all of the estimators are based on simulated method of moments, using the same set of moments. All three estimators impose the initial period normalization and compute the initial conditions in the same way. 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:

$$\lambda_{1,m,\theta} = \frac{Cov(Z_{1,m,\theta}, Z_{0,1,\theta})}{Cov(\ln\theta_1, \ln\theta_0)} \text{ for all } m = 1, 2, 3, \tag{17}$$

  where $\ln\theta_1$ and hence $\lambda_{1,m,\theta}$ depend on the γ_0 parameter. In the Appendix, we show that this estimator identifies the production technology parameter, and is invariant to the measurement parameters. This identification analysis is a special case of the general analysis found in Agostinelli and Wiswall (2016).

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 A.1). 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 simulated samples for each estimation exercise. We use a robust grid search over the parameter space to compute the estimator.

4.1.3 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 0.7. 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 0.7.[12]

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

[12]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.

age-standardizes the measures. Figure 1 shows that estimates of γ_0 using age-standardized measures and re-normalization are also biased.[13]

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 ($\gamma_0 \in (0, 1)$). Estimators 1 and 2, based on imposing re-normalization restrictions as in Cunha et al. (2010), cannot recover the true value of the primitive production function.

4.2 Two Period General CES Model

4.2.1 Data Generating Process

In this next set of exercises, we consider estimating a more general CES technology of skill formation, as in Cunha et al. (2010):

$$\theta_1 = (\gamma_0 \theta_0^{\phi_0} + (1 - \gamma_0) I_0^{\phi_0})^{1/\phi_0} \tag{18}$$

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

4.2.2 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.

4.2.3 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

[13] 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.

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) as used by Cunha et al. (2010) are biased.[14]

5 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 and bias the estimation. We show that the approach used in the estimation of Cunha et al. (2010) of "re-normalizing" latent skills each period–treating skills in each period as separate factors–is unnecessary because their assumed restricted CES production technology already has a known location and scale. Further, these over-identifying restrictions can impose important biases. Emphasizing the problems with combining re-normalization and restricted technologies, we show that simple alternative estimators for these technologies can in fact identify the underlying parameters even without imposing these re-normalization restrictions. We demonstrate our analytic results in a series of Monte Carlo experiments.

References

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. *Working Paper*, 2016.

[14]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.

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.

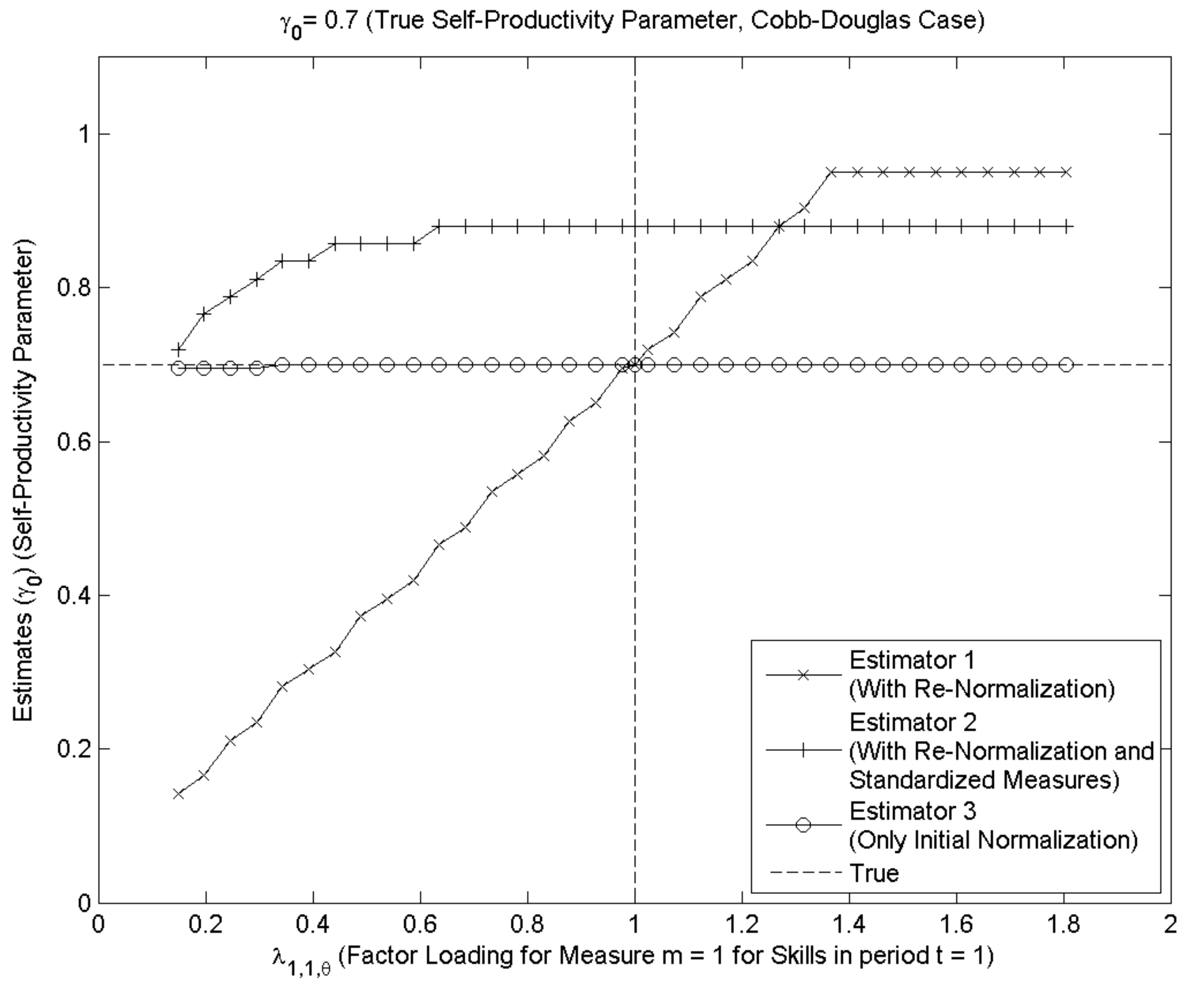
Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

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.

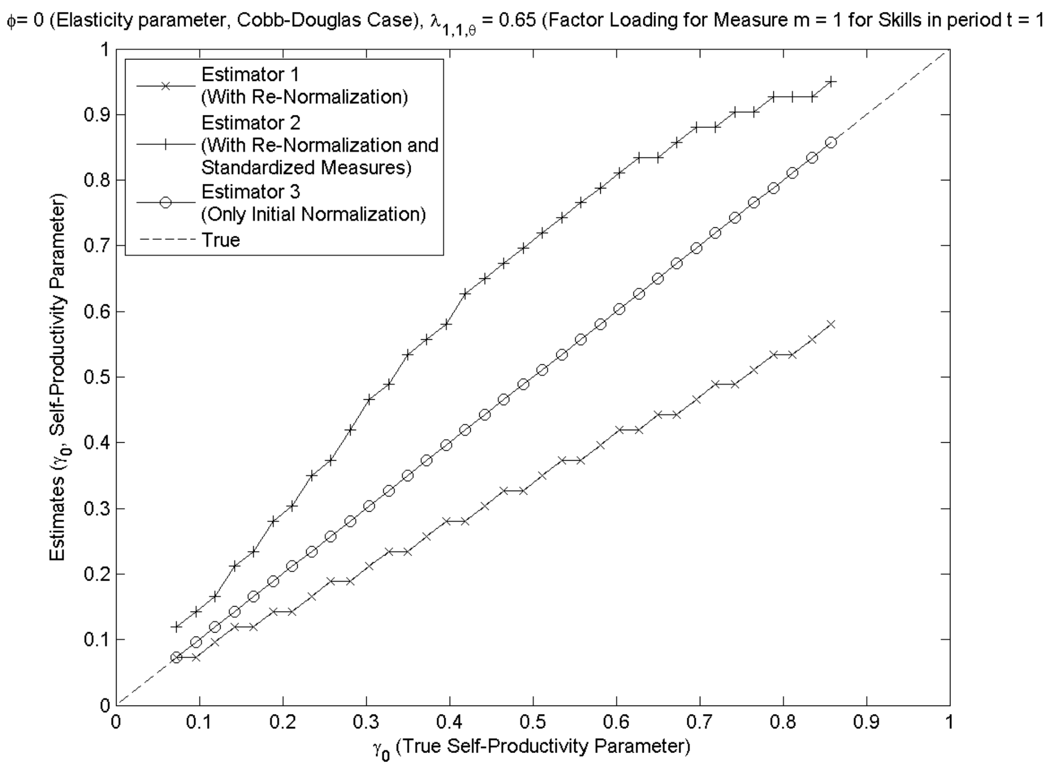
Ronni Pavan. On the production of skills and the birth order effect. *Working Paper*, 2015.

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



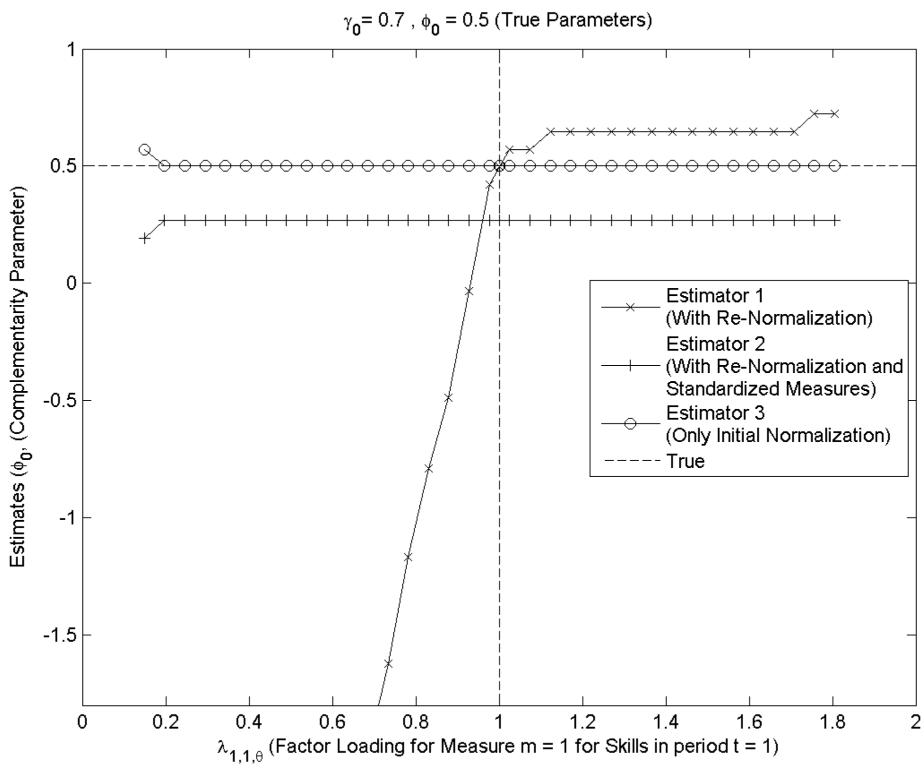
Notes: The dashed line represents the true value of the parameter. The true value and the Estimator 3 value are nearly equal and hence these lines overlap in the figure.

Figure 2: Monte Carlo Results for Estimates of γ



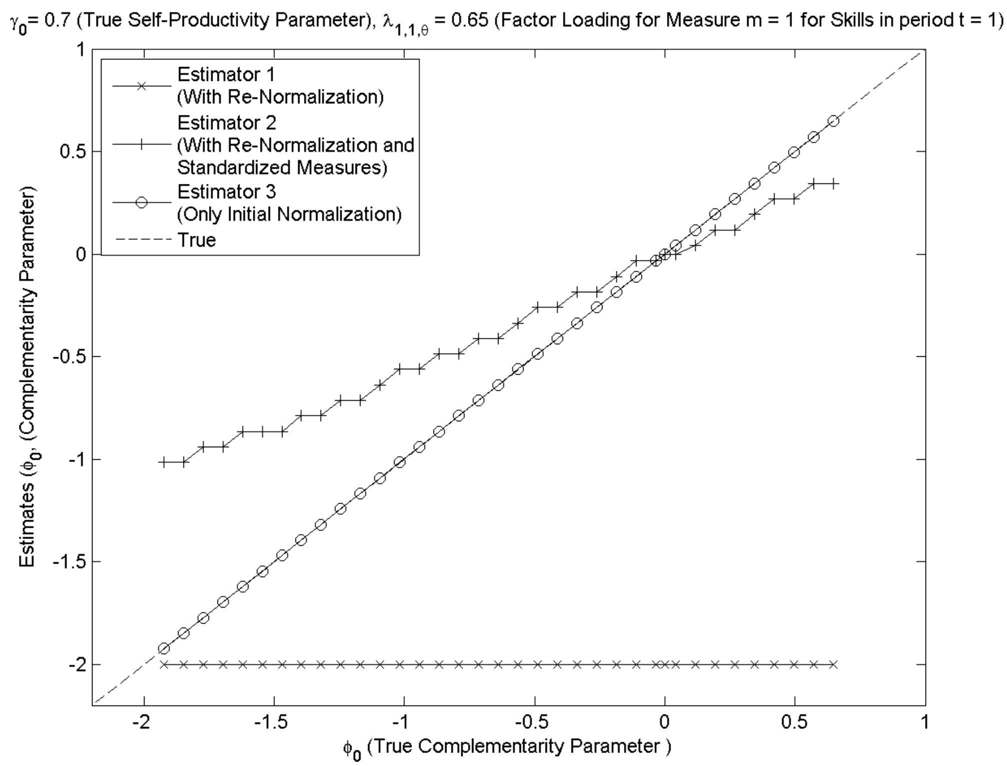
Notes: The 45 degree dashed line displays the true parameter. The true value and the Estimator 3 value are nearly equal and hence these lines overlap in the figure.

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 nearly 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 nearly equal and hence these lines overlap in the figure.

APPENDIX

A.1 Monte Carlo Details

The moments used in all estimators is given by

$$
\begin{aligned}
Cov(Z_{1,m,\theta}, Z_{0,m',\theta}) &= \lambda_{1,m,\theta}\lambda_{0,m',\theta}Cov(\ln\theta_1, \ln\theta_0) && \text{(A-1)}\\
Cov(Z_{1,m,\theta}, Z_{0,m',I}) &= \lambda_{1,m,\theta}\lambda_{0,m',I}Cov(\ln\theta_1, \ln I_0) && \text{(A-2)}\\
Cov(Z_{1,m,\theta}, (Z_{0,m',\theta})^2) &= 2\mu_{0,m',\theta}\lambda_{0,m',\theta}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln\theta_0)+ \\
&\quad \lambda_{1,m,\theta}\lambda^2_{0,m',\theta}Cov(\ln\theta_1, (\ln\theta_0)^2) && \text{(A-3)}\\
Cov(Z_{1,m,\theta}, (Z_{0,m',I})^2) &= 2\mu_{0,m',I}\lambda_{0,m',I}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln I_0)+ \\
&\quad \lambda_{1,m,\theta}\lambda^2_{0,m',I}Cov(\ln\theta_1, (\ln I_0)^2) && \text{(A-4)}\\
Cov(Z_{1,m,\theta}, Z_{0,m'',\theta}\cdot Z_{0,m',I}) &= \mu_{0,m',I}\lambda_{0,m'',\theta}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln\theta_0)+ \\
&\quad \mu_{0,m'',\theta}\lambda_{0,m',I}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln I_0)+ \\
&\quad \lambda_{0,m'',\theta}\lambda_{0,m',I}\lambda_{1,m,\theta}Cov(\ln\theta_1, \ln\theta_0\cdot\ln I_0) && \text{(A-5)}
\end{aligned}
$$

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 4.

The full list of Monte Carlo parameters is given in the Table.

A.2 Identification without Re-Normalization in the Cobb-Douglas Case

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.

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

$$\ln \theta_1 = \gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0 \tag{A-6}$$

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

$$Cov(Z_{1,m}, \ln I_0) = \lambda_{1,m} Cov(\ln \theta_1, \ln I_0)$$
$$= \lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, \ln I_0) + (1 - \gamma_0) Var(\ln I_0)] \quad \text{(A-7)}$$

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 $Z_{1,m}$:

$$Cov(Z_{1,m}, (\ln I_0)^2) = \lambda_{1,m} Cov(\ln \theta_1, (\ln I_0)^2)$$
$$= \lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, (\ln I_0)^2) + (1 - \gamma_0) Cov(\ln I_0, (\ln I_0)^2)] \quad \text{(A-8)}$$

The ratio of these two covariances is

$$\frac{Cov(Z_{1,m}, \ln I_0)}{Cov(Z_{1,m}, (\ln I_0)^2)} = \frac{\lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, \ln I_0) + (1 - \gamma_0) Var(\ln I_0)]}{\lambda_{1,m}[\gamma_0 Cov(\ln \theta_0, (\ln I_0)^2) + (1 - \gamma_0) Cov(\ln I_0, (\ln I_0)^2)]}$$
$$= \frac{\gamma_0 Cov(\ln \theta_0, \ln I_0) + (1 - \gamma_0) V(\ln I_0)}{\gamma_0 Cov(\ln \theta_0, (\ln I_0)^2) + (1 - \gamma_0) Cov(\ln I_0, (\ln I_0)^2)} \quad \text{(A-9)}$$

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, we identify the initial period moments, $Var(\ln \theta_0)$, $Var(\ln \theta_0)$, and $Cov(\ln \theta_0, \ln I_0)$. Solving (A-9) 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 used.

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

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} \cdot Cov(\ln \theta_1, \ln \theta_0)} \tag{A-10}$$

$$= \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} \cdot (\gamma_0 Var(\ln \theta_0) + (1-\gamma_0) Cov(\ln \theta_0, \ln I_0))}, \tag{A-11}$$

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.

A.3 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 (A-6). 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

$$\begin{aligned} Z_{1,m} &= \mu_{1,m} + \lambda_{1,m}\gamma_0\widetilde{Z}_{0,m} + \lambda_{1,m}(1-\gamma_0)\ln I_0 + (\epsilon_{1,m} + \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}) \\ &= \beta_0 + \beta_1\widetilde{Z}_{0,m} + \beta_2 \ln I_0 + \pi_{1,m} \end{aligned} \tag{A-12}$$

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 (A-12) now has the standard errors-in-variables form: (A-12) 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 (A-6) 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:

$$\ln\theta_1 = \gamma_{0,\theta}\ln\theta_0 + \gamma_{0,I}\ln I_0, \tag{A-13}$$

where $\gamma_{0,\theta}$ and $\gamma_{0,1}$ are free parameters and do not sum to 1, the structural parameters in equation A-13 $\gamma_{0,\theta}$ 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

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}\gamma_0 \widetilde{Z}_{0,m} + (\epsilon_{1,m} + \lambda_{1,m}\gamma_0 \widetilde{\epsilon}_{0,m}) \tag{A-14}$$

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.

The result above shows that assuming $\lambda_{1,m} = 1$ is an over-identifying restriction because we were able to point-identify the factor loading (and this factor loading need not be 1). For some more intuition, returning to the equation above, suppose we assume $\lambda_{1,m} = 1$ (as in re-normalization), and we then have this set of 2 equations:

$$\gamma_0 = \beta_1 \tag{A-15}$$

and

$$\gamma_0 = 1 - \beta_2 \tag{A-16}$$

In this way, we can use β_1, β_2 estimates to test the over-identifying restriction imposed by $\lambda_{1,m} = 1$, i.e. the null hypothesis $H_0 : \beta_1 + \beta_2 = 1$.