IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13714

A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation

Emilia Del Bono
Josh Kinsler
Ronni Pavan

SEPTEMBER 2020

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13714

A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation

Emilia Del Bono
ISER, University of Essex and IZA

Josh Kinsler
University of Georgia

Ronni Pavan
University of Rochester

SEPTEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

A Note on the Importance of Normalizations in Dynamic Latent Factor Models of Skill Formation*

In this paper we highlight an important property of the translog production function for the identification of treatment effects in a model of latent skill formation. We show that when using a translog specification of the skill technology, properly anchored treatment effect estimates are invariant to any location and scale normalizations of the underlying measures. By contrast, when researchers assume a CES production function and impose standard location and scale normalizations, the resulting treatment effect estimates are biased. Interestingly, the CES technology with standard normalizations yields biased treatment effect estimates even when age-invariant measures of the skills are available. We theoretically prove the normalization invariance of the translog production function and then produce several simulations illustrating the effects of location and scale normalizations for different technologies and types of skills measures.

JEL Classification: C13, C18, I38, J13, J24

Keywords: children, human capital, dynamic factor analysis, measurement, policy

Corresponding author:
Ronni Pavan
Department of Economics
University of Rochester
Harkness Hall
Rochester, NY 14627
USA

E-mail: ronni.pavan@rochester.edu

* We would like to thank Joachim Freyberger for his comments and feedback on our paper. We also thank Francesco Agostinelli and Matthew Wiswall for their helpful suggestions. The idea for this note originated from discussions with them. We acknowledge funding from the National Science Foundation, award no. 1725270. Emilia Del Bono received support from the ESRC Centre on Micro-Social Change (MiSoC), award no. ES/S012486/1.

1 Introduction

Dynamic latent factor models have become a popular tool for studying child development (Cunha *et al.* , 2010; Agostinelli & Wiswall, 2020; Pavan, 2016; Del Bono *et al.* , 2020; Attanasio *et al.* , 2017, 2020). In these models, unobserved child skills evolve dynamically according to a specified production technology, where the inputs can include past child skill levels, parental skills, and parental investments. Typically, the inputs and outputs of the technology are unobserved, but multiple noisy measures of each latent construct are available. Identification of these models is a challenge since latent skills have no natural units and lack a known location and scale. In order to jointly identify the production technology and measurement model, a location and scale normalization is required. Moreover, because the latent skills have no natural units, it can be useful to anchor them to an adult outcome such as earnings or completed schooling.

After choosing a parametric function for the production technology and a set of scale and location normalizations, researchers can estimate the parameters of the measurement system and production technology. The model can then be used to assess how altering skill inputs across different developmental stages influences final period child skills or adult outcomes. In this paper, we illustrate that the parametric form of the production technology has important implications for the generality of these estimated treatment effects. We prove that input treatment effects associated with a translog production function can be identified regardless of any scale and location normalization. The same is not true for the constant elasticity of substitution technology, another commonly used specification of the production function, where only specific normalizations produce unbiased treatment effects (Freyberger, 2020). Our findings indicate that there are important advantages to choosing a translog production function when estimating a dynamic latent factor model.

This paper builds on the recent and growing literature focused on the identification of dynamic latent factor models. Cunha & Heckman (2008) show that with a sufficient number of noisy measures relative to the latent skill and the appropriate scale and location normalizations, it is possible to identify a linear production technology that is anchored to an adult outcome. Cunha *et al.* (2010) extend this work, proving non-parametric identification of the production technology after applying a scale and location normalization to the measures. While Cunha *et al.* (2010) prove non-parametric identification of the production technology, for estimation they employ a constant returns to scale, constant elasticity of substitution (CES) production function. Agostinelli & Wiswall (2016) show that the estimated shape of this technology is not invariant to different location and scale normalizations applied to the measures, and Agostinelli & Wiswall (2020) demonstrates that if age invariant measures of the underlying skills are available, then identification of a more general production function with free returns to scale is possible.

Agostinelli & Wiswall (2020) claim that when knowledge of the production parameters is important, age invariance is a useful measurement property. An age invariant measure is one where differently aged children with identical latent skill levels attain the same value for the measure on average. There are two benefits associated with employing age invariant measures. First, the units of the latent factors are readily interpretable since they match those of the age invariant measures.[1] Second, age invariance requires that the intercept and loading factor for the age invariant measure are constant across time. As a result, once these parameters are normalized in the initial period, no additional normalizations are necessary. It then follows that changes in the location and scale of the skill distribution (through TFP dynamics and returns to scale production) are identified by changes in the location and scale of the age invariant measure. Many of the most recent dynamic latent factor models employ age-invariant measures (Agostinelli *et al.* , 2020; Attanasio *et al.* , 2019; Aucejo & James, 2019) to allow for time-varying TFP and free returns to scale.

However, relying on age-invariant measures to identify and estimate a production function that allows for TFP dynamics and free returns to scale is limiting in at least three dimensions. First, although skill measures may be designed to be age-invariant, there is no ex-post method to verify this property. Second, the location and scale of the age-invariant measure are typically normalized in the initial period. The estimated production parameters and treatment effects can be sensitive to this normalization, as noted in Attanasio *et al.* (2019). Finally, age invariant measures are not always available. This is especially

[1] Assuming the initial scale of the age invariant measure is set to 1, which is often the default.

true for child non-cognitive skills, implying that a researcher's flexibility in modeling joint skill dynamics is limited.

Rather than rely on age-invariant measures to identify flexible production functions and their associated treatment effects, we propose that researchers rely instead on a particular form for the production technology, the translog function (Christensen *et al.* , 1973). We show that treatment effects stemming from a translog production function with TFP dynamics and free returns to scale can be identified regardless of which location and scale normalization is implemented. This is not a general property of all production functions, and in fact the CES technology does not possess this feature.

Although the translog production function is a second-order Taylor polynomial approximation of the CES function (Kim, 1992), there are important differences between the two. The translog technology imposes no a priori restrictions on the cross-elasticities of substitution and, more importantly, it is linear in the parameters. Given that scale and location normalizations are linear transformations of the latent factors, any normalizations embedded in a linear measurement model will be counterbalanced by a change in the translog technology parameters. As a result, the implied marginal effects, when anchored appropriately, will be unaffected. The same is not true when the technology is non-linear in the parameters, as with the CES. For the CES, a specific set of normalizations is required to estimate unbiased treatment effects (Freyberger, 2020). A number of papers in the literature already employ variants of the translog production function because of its flexibility in capturing substitution patterns among inputs (Agostinelli *et al.* , 2020; Moroni *et al.* , 2019; Boneva & Rauh, 2017; Ronda, 2017).[2] Our results indicate that there is an important added benefit of working with a translog production technology.

Absent any location or scale normalization, the parameters of the translog production technology in a dynamic latent factor model cannot be interpreted without loss of generality. However, in the case of a translog technology, the production parameters can be difficult to interpret directly when higher-order terms are included. To understand the pro-

[2]The translog production function has also been used to study various industries in the US, including electric power generation (Christensen & Greene, 1976), manufacturing (Humphrey & Moroney, 1975; Hellerstein *et al.* , 1999), and public schooling (Figlio, 1999).

ductivity and substitutability of various inputs, a simple approach is to study the treatment effects of various inputs. We show that these treatment effects can be identified without any location or scale normalization. The fact that we only identify treatment effects is not particularly limiting since most papers in the literature use the estimated production function to examine how changes in a key input, such as household income, impact future child skills or adult outcomes (Agostinelli & Wiswall, 2020; Attanasio *et al.* , 2019; Aucejo & James, 2019).

In the first part of the paper (Section 2) we formally prove that appropriately anchored treatment effects based on a translog production function are identified absent any scale or location normalization. The setup of the model is similar in spirit to the ones employed in Cunha *et al.* (2010) and Agostinelli & Wiswall (2020). There is a single latent skill that evolves dynamically as a function of lagged latent skill and latent parental investment. Parental investment is determined by child skill and household income. Within this framework, we focus on identifying treatment effects associated with changes in family income across different periods of development. We show that if an adult outcome is available, income treatment effects can be anchored to this outcome. Alternatively, when an adult outcome is not available, income treatment effects can be anchored to the standard deviation of the child skill.

The second part of the paper (Section 3) provides simulation evidence in support of the theoretical proof. Specifically, we show that identification of treatment effects does not depend on location and scale normalizations when we adopt a translog production function, even when these normalizations are imposed to point estimate all parameters. For the simulation, we also expand on the theoretical framework by allowing for multiple latent skills, a link between skills and the evolution of household income, and a flexible translog technology. Finally, we illustrate that our result does not extend to other production technologies under commonly utilized normalizations, such as the CES production function. In that case the estimated treatment effects are sensitive to location and scale normalizations. Moreover, the availability of an age-invariant measure is not a solution since even an age-invariant measure requires an initial normalization that will impact the estimated treatment effects.

Our paper is the first to highlight that age-invariance is unnecessary when using a translog technology and not particularly useful even when adopting a CES technology. A contemporaneous paper by Freyberger (2020) also addresses issues of identification in dynamic models of latent skills formation, illustrating the types of policy relevant parameters that are identified absent location and scale normalizations and age-invariance for a broad class of production functions. Additionally, Freyberger (2020) shows that the adoption of a CES production function requires a very specific set of normalizations to both point identify the production parameters and produce unbiased estimates of treatment effects. However, these normalizations have never been used in the literature on skills formation to the best of our knowledge.

2 Location and Scale Invariance of Translog Treatment Effects

The setup of our theoretical framework builds off the models utilized in Cunha *et al.* (2010) and Agostinelli & Wiswall (2020). There is one latent child skill (θ_t) that evolves over time (t) and a latent level of parental investment (I_t) that also varies over time. Investment depends on household income (Y_t) and child skill, and income evolves stochastically. An adult outcome (Q) is available as an anchor for child skills. Multiple noisy measures of child skill (Z_{θ,t,m_t}) and parental investment (Z_{I,t,l_t}) are available each period, where m_t and l_t index the specific skill and investment measures employed in period t.

The following equations describe the key components of the model. Skill dynamics are determined by the following translog technology

$$\ln \theta_{t+1} = A_t + \psi_t \left(\gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + (1 - \gamma_{1t} - \gamma_{2t}) \ln \theta_t \cdot \ln I_t\right) + \eta_{\theta t}.$$

In our simulations we extend the function to allow for square terms. Parental investment is determined by

$$\ln I_t = \alpha_{0t} + \alpha_{1t} \ln \theta_t + \alpha_{2t} \ln Y_t + \eta_{It}.$$

We assume and adult outcome is available and given by

$$Q = \mu_Q + \alpha_Q \ln \theta_T + \eta_Q,$$

while household income transitions according to

$$\ln Y_{t+1} = \rho_{0t} + \rho_{1t} \ln Y_t + \eta_{Yt}.$$

Finally, multiple measures of latent skill and parental investment are available each period. These measures take the following form[3]

$$\begin{aligned} Z_{\theta,t,m_t} &= \mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t} \ln \theta_t + u_{\theta,t,m_t} \\ Z_{I,t,l_t} &= \mu_{I,t,l_t} + \lambda_{I,t,l_t} \ln I_t + u_{I,t,l_t}. \end{aligned}$$

While Cunha *et al.* (2010) allow for more general forms of measurement equations, our framework requires that the measures are log-linear in the latent variables. We maintain the orthogonality assumptions outlined in Agostinelli & Wiswall (2020). In particular, we assume that $\eta_{\theta t}$ is mean zero, i.i.d, and independent of the current stock of skills and investment. η_{It} is mean zero, i.i.d, and independent of the current stock of skills and household income. η_{Yt} is mean zero, i.i.d, and independent of all latent variables. Finally, all measurement errors (including η_Q) are assumed independent of each other (across measures and over time), and all measurement errors are assumed independent of

[3]While the skill measures can be understood as the raw measures found in a survey, researchers often employ age-standardized measures. Age-standardizing the measures has no impact on our theoretical result. Assume, for example, that the raw skill measures are given by

$$Z^R_{\theta,t,m_t} = \mu^R_{\theta,t,m_t} + \lambda^R_{\theta,t,m_t} \ln \theta_t + u^R_{\theta,t,m_t},$$

and then apply a linear transformation such that: $Z_{\theta,t,m_t} = \frac{Z^R_{\theta,t,m_t} - a_{m,t}}{b_{m,t}}$. We can then re-define $\mu_{\theta,t,m_t} = \frac{\mu^R_{\theta,t,m_t} - a_{m,t}}{b_{m,t}}$, $\lambda_{\theta,t,m_t} = \lambda^R_{\theta,t,m_t}/b_{m,t}$ and $u_{\theta,t,m_t} = u^R_{\theta,t,m_t}/b_{m,t}$ and proceed accordingly.

the latent variables, household income, and the structural shocks ($\eta_{\theta t}$, η_{It}, η_{Yt}). Specific distributional assumptions for these shocks are not necessary for this discussion, although we do require these distributions to be continuous and have finite first moments.

While the above equations define the dynamic features of the latent factor model, the initial conditions must be specified. We assume

$$\begin{aligned} \Omega &= (\ln \theta_1, \ln Y_1) \\ \Omega &\sim F(\mu_\Omega, \Sigma_\Omega) \end{aligned}$$

The correlation between initial skill and income means that income and skill will remain correlated throughout. Thus, if one is interested in estimating the impact of an income boost early in life on adult outcomes Q, it will be critical to account for latent skills of the child.

The remainder of this section shows that properly anchored treatment effects based on the technology described in point (1) are identified absent any location and scale normalization of the underlying latent factors. The proof proceeds by identifying components of the signal, investment, and production models. We then show how to combine these components to recover properly anchored treatment effects using either the available adult outcome or the standard deviation of latent skill.

2.1 Identifying Skill and Investment Signals

In the first part of our identification proof, we show that the joint density of all signal components (the measures minus the idiosyncratic noise) can be identified without imposing any normalization. We use this result in Section 2.5 to derive the distribution of treatment effects.

Assume that two measures of child skill and parental investment are available each period, and three measures are available at least once. Consider first the measurement equations associated with child skill. Notice that

$$\frac{cov\left(Z_{\theta,t,m_t}, Z_{\theta,\tau,m_t''}\right)}{cov\left(Z_{\theta,t,m_t'}, Z_{\theta,\tau,m_t''}\right)} = \frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}$$

and $E\left(Z_{\theta,t,m_t}\right) = \mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t}\mu_{\theta t}$ (where $\mu_{\theta t} = E(\ln\theta_t)$) are directly identified from the observable measures. Multiplying the mean of measure m_t', $E\left(Z_{\theta,t,m_t'}\right)$, by the ratio of the factor loadings from measures m_t and m_t' leads to

$$\frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}E\left(Z_{\theta,t,m_t'}\right) = \frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}\mu_{\theta,t,m_t'} + \lambda_{\theta,t,m_t}\mu_{\theta t}.$$

Taking the difference between the above and $E\left(Z_{\theta,t,m_t}\right)$ yields

$$\frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}E\left(Z_{\theta,t,m_t'}\right) - E\left(Z_{\theta,t,m_t}\right) \quad = \tfrac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}\mu_{\theta,t,m_t'} - \mu_{\theta,t,m_t}.$$

The left hand side of the above equation is identified, meaning that $\breve{Z}_{\theta,t,m_t'} = \frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}Z_{\theta,t,m_t'} - \frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}E\left(Z_{\theta,t,m_t'}\right) + E\left(Z_{\theta,t,m_t}\right)$ is also identified. Therefore, we can derive the joint density of Z_{θ,t,m_t} and $\breve{Z}_{\theta,t,m_t'}$.

Stacking the equations for Z_{θ,t,m_t} and $\breve{Z}_{\theta,t,m_t'}$,

$$\begin{aligned} Z_{\theta,t,m_t} &= \mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t}\ln\theta_t + u_{\theta,t,m_t} \\ \breve{Z}_{\theta,t,m_t'} &= \mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t}\ln\theta_t + \frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}u_{\theta,t,m_t'}, \end{aligned}$$

reveals that they are noisy measures of the same underlying signal, $\mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t}\ln\theta_t$. Given our previous assumptions, $\mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t}\ln\theta_t$, u_{θ,t,m_t}, and $\frac{\lambda_{\theta,t,m_t}}{\lambda_{\theta,t,m_t'}}u_{\theta,t,m_t'}$ are mutually independent, continuous, have finite first moments, and $E(u_{\theta,t,m_t}) = 0$. Then, using Kotlarski's lemma (Kotlarski, 1967), we can identify the density of $\mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t}\ln\theta_t$.

Taking an identical approach, we can also identify the density of $\mu_{I,t,l_t} + \lambda_{I,t,l_t}\ln I_t$. Stacking these signals within and across periods, we can identify the joint density of the child skill and investment signals.

2.2 Identifying Reduced-Form Investment Parameters

To identify the parameters of the parental investment function, consider Z_{I,t,l_t}. This measure is a noisy signal of parental investment, $\ln I_t$. Substituting the parental investment function for $\ln I_t$ in Z_{I,t,l_t} yields

$$Z_{I,t,l_t} = \mu_{I,t,l_t} + \lambda_{I,t,l_t}\alpha_{0t} + \lambda_{I,t,l_t}\alpha_{1t}\ln\theta_t + \lambda_{I,t,l_t}\alpha_{2t}\ln Y_t + \lambda_{I,t,l_t l_1}\eta_{It} + u_{I,t,l_t}$$

In the above equation, $\ln\theta_t$ is unobserved. To replace it, notice that

$$\ln\theta_t = \frac{Z_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}} - \frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}} - \frac{u_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}}$$

Substituting for $\ln\theta_t$ using the measure m_t results in

$$\begin{aligned} Z_{I,t,l_t} &= \mu_{I,t,l_t} + \lambda_{I,t,l_t}\alpha_{0t} - \tfrac{\lambda_{I,t,l_t}\alpha_{1t}\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}} \\ &\quad + \tfrac{\lambda_{I,t,l_t}\alpha_{1t}}{\lambda_{\theta,t,m_t}} Z_{\theta,t,m_t} + \lambda_{I,t,l_t}\alpha_{2t}\ln Y_t \\ &\quad - \tfrac{\lambda_{I,t,l_t}\alpha_{1t}}{\lambda_{\theta,t,m_t}} u_{\theta,t,m_t} + \lambda_{I,t,l_t}\eta_{It} + u_{I,t,l_t} \end{aligned}$$

where Z_{θ,t,m_t} is correlated with the error term. Using Z_{θ,t,m'_t} as an instrument we identify three coefficients:

$$\begin{aligned} \beta^0_{I,t}(l_t, m_t) &= \mu_{I,t,l_t} + \lambda_{I,t,l_t}\alpha_{0t} - \frac{\lambda_{I,t,l_t}\alpha_{1t}\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}} \\ \beta^1_{I,t}(l_t, m_t) &= \frac{\lambda_{I,t,l_t}\alpha_{1t}}{\lambda_{\theta,t,m_t}} \\ \beta^2_{I,t}(l_t, m_t) &= \lambda_{I,t,l_t}\alpha_{2t} \end{aligned} \tag{1}$$

Throughout the paper, we use β to indicate known regression coefficients. The precise notation for the reduced form investment coefficients is $\beta^j_{I,t}(m_t, l_t)$ where j represents the coefficient number and the arguments reflect the child skill and investment measures utilized in the regression. For ease of presentation, we suppress the arguments of the $\beta^j_{I,t}$ coefficients when possible.

2.3 Identifying Reduced-Form Production Parameters

Similar to our approach for identifying the investment function, first consider a $t+1$ period measure of child skill, $Z_{\theta,t+1,m_{t+1}}$. This is a noisy measure of $\ln\theta_{t+1}$. However, we can substitute for $\ln\theta_{t+1}$ using the production function. This yields

$$\begin{aligned} Z_{\theta,t+1,m_{t+1}} &= \mu_{\theta,t+1,m_{t+1}} + \lambda_{\theta,t+1,m_{t+1}} A_t + \lambda_{\theta,t+1,m_{t+1}} \psi_t \gamma_{1t} \ln\theta_t + \lambda_{\theta,t+1,m_{t+1}} \psi_t \gamma_{2t} \ln I_t \\ &\quad + \lambda_{\theta,t+1,m_{t+1}} \psi_t \left(1-\gamma_{1t}-\gamma_{2t}\right) \ln\theta_t \cdot \ln I_t + \lambda_{\theta,t+1,m_{t+1}} \eta_{\theta t} + u_{\theta,t+1,m_{t+1}}. \end{aligned}$$

In the above equation, both $\ln\theta_t$ and $\ln I_t$ are unobserved, but we can substitute for them using their respective measurement equations, Z_{θ,t,m_t} and Z_{I,t,l_t}. Making these substitutions and organizing terms appropriately yields

$$\begin{aligned} Z_{\theta,t+1,m_{t+1}} &= \mu_{\theta,t+1,m_{t+1}} + \lambda_{\theta,t+1,m_{t+1}} A_t - \lambda_{\theta,t+1,m_{t+1}} \psi_t \gamma_{1t} \frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}} + \\ &- \lambda_{\theta,t+1,m_{t+1}} \psi_t \gamma_{2t} \frac{\mu_{I,t,l_t}}{\lambda_{I,t,l_t}} + \lambda_{\theta,t+1,m_{t+1}} \psi_t \left(1-\gamma_{1t}-\gamma_{2t}\right) \frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}} \frac{\mu_{I,t,l_t}}{\lambda_{I,t,l_t}} \\ &+ \frac{\lambda_{\theta,t+1,m_{t+1}}}{\lambda_{\theta,t,m_t}} \psi_t \left(\gamma_{1t} - \left(1-\gamma_{1t}-\gamma_{2t}\right) \frac{\mu_{I,t,l_t}}{\lambda_{I,t,l_t}}\right) Z_{\theta,t,m_t} + \\ &+ \frac{\lambda_{\theta,t+1,m_{t+1}}}{\lambda_{I,t,l_t}} \psi_t \left(\gamma_{2t} - \left(1-\gamma_{1t}-\gamma_{2t}\right) \frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}}\right) Z_{I,t,l_t} + \\ &+ \frac{\lambda_{\theta,t+1,m_{t+1}}}{\lambda_{\theta,t,m_t} \lambda_{I,t,l_t}} \psi_t \left(1-\gamma_{1t}-\gamma_{2t}\right) Z_{\theta,t,m_t} \cdot Z_{I,t,l_t} \\ &+ \epsilon_{\theta t+1} \end{aligned}$$

where $\epsilon_{\theta t+1}$ is a mean zero error term correlated with Z_{θ,t,m_t}, Z_{I,t,l_t} and $Z_{\theta,t,m_t} \cdot Z_{I,t,l_t}$, meaning that we cannot recover the reduced-form parameters by regressing $Z_{\theta,t+1,m_{t+1}}$ on these three observed variables. Instead, we employ Z_{θ,t,m'_t}, Z_{I,t,l'_t}, and $Z_{\theta,t,m'_t} \cdot Z_{I,t,l'_t}$ as instruments, where $l'_t \neq l_t$ and $m'_t \neq m_t$. This allows us to recover the following reduced form parameters

$$
\begin{aligned}
\beta^1_{\theta,t}\left(m_{t+1},m_t,l_t\right) &= \mu_{\theta,t+1,m_{t+1}}+\lambda_{\theta,t+1,m_{t+1}}A_t-\lambda_{\theta,t+1,m_{t+1}}\psi_t\gamma_{1t}\frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}}+\\
&- \lambda_{\theta,t+1,m_{t+1}}\psi_t\gamma_{2t}\frac{\mu_{I,t,l_t}}{\lambda_{I,t,l_t}}+\lambda_{\theta,t+1,m_{t+1}}\psi_t\left(1-\gamma_{1t}-\gamma_{2t}\right)\frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}}\frac{\mu_{I,t,l_t}}{\lambda_{I,t,l_t}}\\
\beta^2_{\theta,t}\left(m_{t+1},m_t,l_t\right) &= \frac{\lambda_{\theta,t+1,m_{t+1}}}{\lambda_{\theta,t,m_t}}\psi_t\left(\gamma_{1t}-\left(1-\gamma_{1t}-\gamma_{2t}\right)\frac{\mu_{I,t,l_t}}{\lambda_{I,t,l_t}}\right)\\
\beta^3_{\theta,t}\left(m_{t+1},m_t,l_t\right) &= \frac{\lambda_{\theta,t+1,m_{t+1}}}{\lambda_{I,t,l_t}}\psi_t\left(\gamma_{2t}-\left(1-\gamma_{1t}-\gamma_{2t}\right)\frac{\mu_{\theta,t,m_t}}{\lambda_{\theta,t,m_t}}\right)\\
\beta^4_{\theta,t}\left(m_{t+1},m_t,l_t\right) &= \frac{\lambda_{\theta,t+1,m_{t+1}}}{\lambda_{\theta,t,m_t}\lambda_{I,t,l_t}}\psi_t\left(1-\gamma_{1t}-\gamma_{2t}\right).
\end{aligned}
$$

Similar to the investment coefficients, $\beta^j_{\theta,t}\left(m_{t+1},m_t,l_t\right)$ indicates known regression coefficients for the cognitive skills regression, where again j is the coefficient number and the arguments represent the measures utilized for the regression. While it does not matter which measure we use for a given latent variable, it is useful to consistently employ the same measure when referring to a latent variable from the same period.

Finally, it will be useful in the next section to have the parameters of the production technology defined in terms of the the known regression coefficients. Doing so yields,

$$
\begin{aligned}
\lambda_{\theta,t+1,m_{t+1}}\psi_t &= \beta^4_{\theta,t}\left(\lambda_{\theta,t,m_t}\lambda_{I,t,l_t}+\mu_{I,t,l_t}\lambda_{\theta,t,m_t}+\mu_{\theta,t,m_t}\lambda_{I,t,l_t}\right)+\left(\lambda_{\theta,t,m_t}\beta^2_{\theta,t}+\lambda_{I,2,l_t}\beta^3_{\theta,t}\right)\\
\psi_t\gamma_{1t} &= \frac{\left(\mu_{I,t,l_t}\lambda_{\theta,t,m_t}\beta^4_{\theta,t}+\lambda_{\theta,t,m_t}\beta^2_{\theta,t}\right)}{\lambda_{\theta,t+1,m_{t+1}}}\\
\psi_t\gamma_{2t} &= \frac{\left(\mu_{\theta,t,m_t}\lambda_{I,t,l_t}\beta^4_{\theta,t}+\lambda_{I,t,l_t}\beta^3_{\theta,t}\right)}{\lambda_{\theta,t+1,m_{t+1}}} \qquad (2)
\end{aligned}
$$

2.4 Identifying the Marginal Effect of Income on Log Skill

As noted in the introduction, it is common for researchers use the estimated technology to simulate how child skills are impacted by interventions in various periods of development. In this section we show that given the parameters already identified, these policy effects are identified up to a normalizing constant.

Consider the marginal change in period $t+k$ skill given a marginal change in household income in period t, holding income in all other periods fixed.[4] This derivative can be written

$$\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t} = \frac{\partial \ln \theta_{t+k-1}}{\partial \ln Y_t} \left(\frac{\partial \ln \theta_{t+k}}{\partial \ln \theta_{t+k-1}} + \frac{\partial \ln \theta_{t+k}}{\partial \ln I_{t+k-1}} \frac{\partial \ln I_{t+k-1}}{\partial \ln \theta_{t+k-1}} \right)$$

The three derivatives inside the parenthesis can be written as

$$\begin{aligned}
\frac{\partial \ln \theta_{t+k}}{\partial \ln \theta_{t+k-1}} &= \psi_{t+k-1}\gamma_{1t+k-1} + \psi_{t+k-1}\left(1-\gamma_{1t+k-1}-\gamma_{2t+k-1}\right)\ln I_{t+k-1} \\
\frac{\partial \ln \theta_{t+k}}{\partial \ln I_{t+k-1}} &= \psi_{t+k-1}\gamma_{2t+k-1} + \psi_{t+k-1}\left(1-\gamma_{1t+k-1}-\gamma_{2t+k-1}\right)\ln \theta_{t+k-1} \\
\frac{\partial \ln I_{t+k-1}}{\partial \ln \theta_{t+k-1}} &= \alpha_{1t+k-1}
\end{aligned}$$

Using the results from equations (1) and (2), we can show that

$$\begin{aligned}
\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t} &= \frac{\partial \ln \theta_{t+k-1}}{\partial \ln Y_t} \frac{\lambda_{\theta,t+k-1,m_{t+k-1}}}{\lambda_{\theta,t+k,m_{t+k}}} \Big[\left(\beta^2_{\theta,t+k-1} + \beta^4_{\theta,t+k-1}\left(\mu_{I,t+k-1,l_{t+k-1}} + \lambda_{I,t,l_{t+k-1}} \ln I_{t+k-1}\right)\right) \\
&+ \beta^1_{I,t+k-1}\left(\beta^3_{\theta,t+k-1} + \beta^4_{\theta,t+k-1}\left(\mu_{\theta,t+k-1,m_{t+k-1}} + \lambda_{\theta,t+k-1,m_{t+k-1}} \ln \theta_{t+k-1}\right)\right) \Big]
\end{aligned}$$

This defines $\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t}$ as a function of known regression coefficients, signals, and a scaled version of $\frac{\partial \ln \theta_{t+k-1}}{\partial \ln Y_t}$. We can then continue recursively, replacing $\frac{\partial \ln \theta_{t+k-1}}{\partial \ln Y_t}$ using a similar formula until

$$\frac{\partial \ln \theta_{t+1}}{\partial \ln Y_t} = \frac{\beta^2_{I,t}}{\lambda_{\theta,t+1,m_{t+1}}} \left(\beta^3_{\theta,t} + \beta^4_{\theta,t}\left(\mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t} \ln \theta_t\right)\right)$$

We can express the recursive structure of the marginal effect of income in the following

[4] A temporary change in household income in period t that does not impact future household income is akin to a one-time government transfer that has been investigated in the previous literature (Agostinelli & Wiswall, 2020). However, our framework can easily accommodate permanent income changes.

way

$$
\begin{aligned}
\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t} &= \prod_{s=1}^{k-1} \Big[\left(\beta^2_{\theta,t+s} + \beta^4_{\theta,t+s}\left(\mu_{I,t+s,l_{t+s}} + \lambda_{I,t,l_{t+s}} \ln I_{t+s}\right)\right) \\
&\times \; \beta^1_{I,t+s}\left(\beta^3_{\theta,t+s} + \beta^4_{\theta,t+s}\left(\mu_{\theta,t+s,m_{t+s}} + \lambda_{\theta,t+s,m_{t+s}} \ln \theta_{t+s}\right)\right) \Big] \\
&\times \; \frac{\beta^2_{I,t}}{\lambda_{\theta,t+k,m_{t+k}}} \left(\beta^3_{\theta,t} + \beta^4_{\theta,t}\left(\mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t} \ln \theta_t\right)\right) \qquad (3)
\end{aligned}
$$

Because the reduced form β's are known and the joint distribution of the skills and investment signals are known, the distribution of $\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t}$ is identified up to a scaling factor, $\frac{1}{\lambda_{\theta,t+k,m_{t+k}}}$.

2.5 Identifying Appropriately Anchored Treatment Effects

The final step is to show that we can identify appropriately anchored treatment effects. There are two possibilities, anchor to an observable adult outcome or to the standard deviation of the latent child log skill.[5]

2.5.1 Anchoring to an adult outcome

In this case, the derivative of interest becomes $\frac{\partial Q}{\partial \ln Y_t} = \alpha_Q \frac{\partial \ln \theta_T}{\partial \ln Y_t}$. Because we can choose $m_T = Q$, $\lambda_{\theta,T,m_T} = \alpha_Q$. Substituting for $\frac{\partial \ln \theta_T}{\partial \ln Y_t}$ with equation (3) and choosing $m_T = Q$ means that the α_Q terms cancel and we can identify the distribution of $\frac{\partial Q}{\partial \ln Y_t}$ absent any scaling factor. We can then use this derivative to construct the appropriate average treatment effect. Suppose, for example, that we are interested in estimating the average impact of a one-time increase in income of $\delta \times 100\%$ during $t = 1$. Assuming δ is small, we can approximate percent changes with log changes, i.e. $\Delta \ln Y_1 = \delta$. The resulting

[5] Notice that in both cases the treatment effects are linear transformations of log skills. Freyberger (2020) shows that if we instead consider outcomes that depend on the level of skills, the resulting treatment effects would depend on the normalizations. Given that in our framework adult outcomes and all measures are linear functions of log skills, we find it natural to consider treatment effects that utilize the same units. It should also be noted that while Freyberger (2020) considers treatment effects anchored to the adult outcomes or in terms of quantiles of the skill distribution, utilizing standardized log skills is a unique feature of this paper.

expected increase in skills measured in adult outcome units is

$$E(\Delta Q|\Delta \ln Y_1 = \delta) \quad = \quad \int \delta \times \frac{\partial Q(m_1,...,m_T,l_1,...,l_{T-1})}{\partial \ln Y_1} dF(m_1,...,m_T,l_1,...,l_{T-1})$$

where $F(m_1,...,m_T,l_1,...,l_{T-1})$ is a compact notation to represent the joint density of the signal content of the measures, whose identification was shown in section 2.1. $\frac{\partial Q(m_1,...,m_T,l_1,...,l_{T-1})}{\partial \ln Y_1}$ is the derivative previously identified, where we make explicit its dependence on a particular set of measures utilized in the analysis. The treatment effect is a linear function of δ since the derivative of the adult outcome with respect to log-income is not a function of income. Although we only show the average treatment effect of a % increase in income, we can also identify the treatment effects associated with different changes in income for different populations.[6]

It should be noticed that even in the absence of an adult outcome, we can still identify the distribution of treatment effects if we express the impact in terms of any of the measures available. For example, we can measure the impact of a change in income in terms of the measure m_T if we calculate $\lambda_{\theta,T,m_T} \frac{\partial \ln \theta_T}{\partial \ln Y_t}$, which can be identified using a similar strategy to the above.

2.5.2 Identifying Treatment Effects in terms of Standard Deviations

If an adult outcome is not available and it is inconvenient to use any of the existing measures to normalize the units of the treatment effect of interest, it is still possible to identify the distribution of treatment effects expressed in terms of the standard deviation of the underlying variables, in this case log skills. Using three measures of child skill we can identify:

$$\lambda_{\theta,t,m}\sigma_{\ln\theta,t} = \sqrt{\frac{cov\left(Z_{\theta,t,m}, Z_{\theta,t,m'}\right) cov\left(Z_{\theta,t,m}, Z_{\theta,t,m''}\right)}{cov\left(Z_{\theta,t,m'}, Z_{\theta,t,m''}\right)}}$$

Knowledge of the above in combination with equation (3) allows us to identify the ratio $\frac{\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t}}{\sigma_{\ln\theta,t+k}}$. This ratio is the marginal effect of a change in income on "standardized" skills,

[6] Additional details available upon request.

where we define standardized skills as $\ln\tilde{\theta}_t = \frac{\ln\theta_t}{\sigma_{\ln\theta,t}}$. From this it follows that $\frac{\partial\ln\tilde{\theta}_T}{\partial\ln Y_1} = \frac{\frac{\partial\ln\theta_T}{\partial\ln Y_1}}{\sigma_{\ln\theta,T}}$. As before, if we are interested in estimating the average impact of a temporary increase in income of $\delta \times 100\%$ during $t = 1$, we can write the expected increase as:

$$E(\Delta\ln\tilde{\theta}_T|\Delta\ln Y_1 = \delta) = \int \delta \times \frac{\partial\ln\tilde{\theta}_T(m_1,...,m_T,l_1,...,l_{T-1})}{\partial\ln Y_1} \times dF(m_1,...,m_T,l_1,...,l_{T-1})$$

where again we make explicit the dependence of the derivative on the underlying measures.

3 Simulation Evidence

So far we have shown that without making any location or scale assumptions on either the measurement model or production function, we are able to show that the distribution of relevant policy effects are identified in a fairly standard dynamic latent factor model where we assume that the production technology takes the translog form. While the policy effects can be identified without a precise location and scale normalization, these normalizations will still be needed when researchers want to separately estimate all the production technology and measurement parameters.

In this section we present a series of simulations to support our theoretical findings. Specifically, we show that when using a translog technology the estimated treatment effects are correct regardless of the specific location and scale normalization. The same is not true when using a CES production function. For the CES, different location and scale normalizations can lead to different estimated treatment effects, all of which can be incorrect. We focus on the CES production function because it is a commonly used specification in the literature on skill formation (Cunha *et al.* , 2010; Attanasio *et al.* , 2019; Aucejo & James, 2019). Additionally, we only consider normalizations that are typically implemented in empirical applications. In particular, we focus on initial scale normalizations and those implied by age-invariance of the measures.[7]

[7]See Freyberger (2020) for the specific types of normalizations which are truly without loss of generality in the CES case.

3.1 Setup

To show that our theoretical result is not driven by the particular setup of Section 2 (other than the translog technology), we adopt a more general framework in our simulations. We consider a model with two time varying latent skills, and allow one of these skills to affect the evolution of income. Additionally, we assume that the adult outcome depends on both latent skills. The inclusion of the second latent skill is motivated by the literature on skill development, where it is common to model the evolution of child cognitive skills together with either parental skills and/or a second child trait, such as non-cognitive skills.

Agents are characterized by two latent skills θ_{1t} and θ_{2t}, parental investments I_t, and family income Y_t. Parental investments are determined by skills, income, and a shock according to

$$\ln I_t = \alpha_{0t} + \alpha_{1t} \ln \theta_{1t} + \alpha_{2t} \ln \theta_{2t} + \alpha_{3t} \ln Y_t + \eta_{It}.$$

Define the vector $\Theta_t = (\theta_{1t}, \theta_{2t}, I_t)$. The translog skill technology then takes the following form

$$\ln \theta_{jt+1} = A_{jt} + \sum_{k=1}^{3} \gamma_{jkt} \ln \Theta_{kt} + \sum_{k=1}^{3} \gamma_{j(k+3)t} \ln \theta_{jt} \cdot \ln \Theta_{kt} + \eta_{\theta jt}$$

where $j = (1, 2)$ and $\ln \Theta_{kt}$ is the log of the k^{th} argument of the vector Θ_t.

The adult outcome is allowed to be a function of both latent skills and can be written as

$$Q = \mu_Q + \alpha_{1Q} \ln \theta_{1T} + \alpha_{2Q} \ln \theta_{2T} + \eta_Q.$$

Household income evolves as

$$\ln Y_{t+1} = \rho_{0t} + \rho_{1t} \ln Y_t + \rho_{2t} \ln \theta_{2t} + \eta_{Yt}.$$

The initial draw of $(\ln \theta_{11}, \ln \theta_{21}, \ln Y_1)$ comes from a mixture of two joint normals, while all other shocks are from independent normal distributions. Lastly, we assume that skill and investment measures take the following form

$$\begin{aligned} Z_{\theta_j,t,m_{jt}} &= \mu_{\theta_j,t,m_{jt}} + \lambda_{\theta_j,t,m_{jt}} \ln \theta_{jt} + u_{\theta_j,t,m_{jt}} \text{ for } j = 1,2 \\ Z_{I,t,l_t} &= \mu_{I,t,l_t} + \lambda_{I,t,l_t} \ln I_t + u_{I,t,l_t}. \end{aligned}$$

Our theoretical proof relies on the availability of multiple skills and investment measures to identify the distribution of signals (Section 2.1) and to eliminate endogeneity in the reduced form production and investment equations (Sections 2.2 and 2.3). Since these identification issues are not relevant to demonstrate the importance (or not) of location and scale normalizations, we simplify the simulation exercise and assume that one error-free measure of each latent factor is available according to:

$$\begin{aligned} Z_{\theta_j,t} &= \mu_{\theta_j,t} + \lambda_{\theta_j,t} \ln \theta_{jt} \text{ for } j = 1,2 \\ Z_{I,t} &= \mu_{I,t} + \lambda_{I,t} \ln I_t. \end{aligned}$$

Given the results from Section 2.1, the joint distribution of $Z_{\theta_j,t}$ and $Z_{I,t}$ is identified, and thus our approach is without loss of generality.

In addition to the above model, we also simulate and estimate a version where we assume a CES production technology. In this case the production function is given by

$$\theta_{jt+1} = A_{jt} \left(\gamma_{j1t} \theta_{1t}^{\phi_{jt}} + \gamma_{j2t} \theta_{2t}^{\phi_{jt}} + (1 - \gamma_{j1t} - \gamma_{j2t}) I_t^{\phi_{jt}} \right)^{\frac{\psi_{jt}}{\phi_{jt}}} e^{\eta_{\theta jt}}.$$

All other model components are as described above.

3.2 Simulation and Results

We simulate data for 500,000 individuals and $T = 4$ periods, generating latent skills, income, and measures according to the equations above. We create different datasets, varying the form of the true production technology (translog or CES) and whether the true measures are age-invariant. For each scenario, we estimate income treatment effects employing different assumptions about the shape of the production technology and different normalizations for the location and scale parameters. Table 1 anchors the estimated treatment effects to an adult outcome, while Table 2 anchors treatment effects to the standard deviation of final period log skills. The tables present the bias in the estimated treatment effects relative to the truth for the various technology and normalization assumptions. It should be stressed that we do not calibrate the model parameters, so in what follows we focus primarily on the presence of bias rather than its magnitude.

To initiate the simulation, we select reasonable parameter values for the production function, investment function, income dynamics, and initial conditions.[8] We begin by assuming that the true measures of skills are not age invariant (Panel A of both tables). We do this by choosing values for $\mu_{\theta_j,t}$ and $\lambda_{\theta_j,t}$ for $j = 1, 2$ that change over time.[9] In the second part of our simulation exercise (Panel B of both tables), we consider the case in which age invariant measures for both skills are available, and set $\mu_{\theta_j,t} = 2$ and $\lambda_{\theta_j,t} = 5$ for $j = 1, 2$ and all periods. Within Panels A and B, we simulate data for four different production technologies. First, we use a translog production function (column 1), then we adopt a CES production function with decreasing, constant and increasing returns to scale, respectively (columns 2-4).

Once we simulate the data, we estimate the model parameters and treatment effects imposing two alternative location and scale normalizations:

1. $\tilde{\mu}_{\theta_j,t} = 0$ and $\tilde{\lambda}_{\theta_j,t} = 1$ for $j = 1, 2$ and all t
2. Technology restrictions: for the translog technology, we set $\tilde{A}_{jt} = 0$ and $\sum_{k=1}^{6} \tilde{\gamma}_{jkt} = 1$, while for the CES technology, we set $\tilde{A}_{jt} = 0$ and $\tilde{\psi}_{jt} = 1$.[10]

We use tildes to emphasize that these are normalizations made for estimation purposes and they do not necessarily reflect the true underlying parameter values.

[8] The precise values for the technology parameters can be found in appendix Table A1. In the appendix, we also replicate the main results simulating 200 data sets, each using randomly chosen set of parameter values from a reasonable range. There is no change when using the translog production function. The CES function continues to produce biased treatment effects.

[9] The loading factors for both skills take the values 5, 2, 10, and 7, while the mean takes the values 0, 1, 2 and 3. For simplicity we always set $\mu_{I,t} = 0$ and $\lambda_{I,t} = 1$ both in the simulation and in the normalization.

[10] Agostinelli & Wiswall (2020) use these normalizations and define the resulting technologies as Known Location and Scale (KLS).

The first normalization restricts the location and scale of the measures to be the same across all periods and is common in the literature on the dynamics of skills formation. In Panel A of Tables 1 and 2, we assume that the true measures *are not* age invariant, even though the researcher is imposing this property with the first normalization. In Panel B, the true measures *are* age invariant, but the researcher picks an arbitrary initial location and scale for the factors. This is justified by the idea that although a researcher may know that a measure's mean and loading are constant over time, they will not know their true values, as noted in Attanasio *et al.* (2019). Thus, it makes sense to examine the effect of picking an incorrect location and scale even when age invariance is satisfied. The second normalization restricts the functional form of the technology in order to pin down the location and scale of the measures and shows the effect of making the wrong technology assumptions. Notice that in this second scenario, the researcher still needs to normalize the initial location and scale parameters, which we set to $\tilde{\mu}_{j,1} = 0$ and $\tilde{\lambda}_{j,1} = 1$, in line with the first normalization.[11]

Looking at the first column of Table 1, we see that for a translog technology the treatment effect associated with a one standard deviation increase in initial log income is unbiased under each of the normalizations, regardless of whether an age-invariant measure is available. When a CES technology is used instead, the treatment effect is biased under all normalizations, although no conclusion about the magnitude and sign of the bias should be drawn given we arbitrarily chose the parameters for the purpose of this simulation. Clearly, the biases would be smaller the closer the chosen location and scale normalizations are to the truth.

Importantly, the bias that emerges when we adopt a CES technology is present both in Panel A, where no age invariant measure is present, and in Panel B, where we assume both skills have an age invariant measure. The bias in Panel B stems from the fact that although the researcher correctly identified an age-invariant measure, the true mean and loading are

[11]As mentioned earlier, an interested reader should look at Freyberger (2020), where the details of the CES normalizations that are truly without loss of generality are derived. Specifically, he shows that setting the location of the unobservables and the loading factor of the adult outcome leads to point identification of the relevant parameters without biasing treatment effects.

unknown. Similarly, bias arises even when we impose a CRS normalization and the true technology is CRS (column 3). This is because we still need an initial normalization of the measures and the values chosen can be different from the true values.

In Table 2 we calculate the impact of a one standard deviation increase in initial log income on final log skills ($\ln \theta_{1T}$) anchored to the standard deviations of those log skills. This is particularly useful when researchers do not have access to adult outcomes. The same considerations apply to this set of results as for Table 1. We do not see biases when adopting a translog specification and different normalizations, but these biases emerge when using a CES technology even when age invariant measures are available.

4 Concluding Remarks

A growing interest in early childhood skill development combined with enhanced data on child skill measures has spawned a rich literature that employs latent factor models to study skill dynamics. Identifying the technology of skills formation when skills are unobserved relies on the availability of multiple noisy measures of child skills each period. While these measures aid in identification, it is still the case that the location and scale of the latent skills needs to be pinned down in order to estimate the production function and measurement parameters.

In this paper we show that in contrast, no location or scale normalizations are necessary to identify certain policy-relevant treatment effects when employing a translog production technology. Moreover, the estimated treatment effects are invariant to the actual scale and location normalization implemented to point identify the model. This property does not generalize, and in particular does not hold for the most common production function used in the literature, the CES. In the case of the CES, the estimated treatment effects are sensitive to the particular normalization chosen even when an age-invariant measure is available. Recent work by Freyberger (2020) shows that some normalizations do not distort treatment effect estimates when using the CES, but these normalizations are quite different from the ones implemented in the literature thus far.

The key takeaway from our paper is that employing a translog production function when estimating a dynamic latent factor model of skill development confers two benefits. First, the translog technology puts no a priori restrictions on the cross-elasticities of substitution, a well-known result in the literature. Second, treatment effect estimates are unaffected by location and scale normalizations, an added benefit relative to other widely used parametric forms for skill technology.

References

Agostinelli, Francesco, & Wiswall, Matthew. 2016. *Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development*. NBER Working Paper 22441.

Agostinelli, Francesco, & Wiswall, Matthew. 2020. *Estimating the Technology of Children's Skill Formation*. NBER Working Paper 22442.

Agostinelli, Francesco, Saharkhiz, Morteza, & Wiswall, Matthew. 2020. *Home and School in the Development of Children*. NBER Working Paper 26037.

Attanasio, Orazio, Meghir, Costas, Nix, Emily, & Salvati, Francesca. 2017. Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of Economic Dynamics*, **25**, 234–259.

Attanasio, Orazio, Meghir, Costas, & Nix, Emily. 2019. Human capital development and parental investment in India. *Review of Economic Studies*. forthcoming.

Attanasio, Orazio, Cattan, Sara, Fitzsimons, Emla, Meghir, Costas, & Rubio-Codina, Marta. 2020. Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia. *American Economic Review*, **110**(1), 48–85.

Aucejo, Esteban M., & James, Jonathan. 2019. *The Path to College Education: The Role of Math and Verbal Skills*.

Boneva, Teodora, & Rauh, Christopher. 2017. *Human Capital and Parental Beliefs*.

Christensen, Laurits R., & Greene, William H. 1976. Economies of Scale in U.S. Electric Power Generation. *Journal of Political Economy*, **84**(4), 655–676.

Christensen, Laurits R., Jorgenson, Dale W., & Lau, Lawrence J. 1973. Transcendental Logarithmic Production Frontiers. *The Review of Economics and Statistics*, **55**(1), 28–45.

Cunha, Flavio, & Heckman, James J. 2008. Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, **43**(4), 738–782.

Cunha, Flavio, Heckman, James J., & Schennach, Susanne M. 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, **78**(3), 883–931.

Del Bono, Emilia, Kinsler, Josh, & Pavan, Ronni. 2020. *Skill Formation and the Trouble with Child Non-Cognitive Skill Measures*. Working Paper.

Figlio, David N. 1999. Functional form and the estimated effects of school resources. *Economics of Education Review*, **18**, 241–252.

Freyberger, Joachim. 2020. *Normalizations and misspecification in skill formation models*. unpublished manuscript.

Hellerstein, Judith K., Neumark, David, & Troske, Kenneth R. 1999. Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations. *Journal of Labor Economics*, **17**(3), 409–446.

Humphrey, David B., & Moroney, J.R. 1975. Substitution among Capital, Labor, and Natural Resource Products in American Manufacturing. *Journal of Political Economy*, **83**(1), 57–82.

Kim, H. Youn. 1992. The Translog Production Function and Variable Returns to Scale. *The Review of Economics and Statistics*, **74**(3), 546–552.

Kotlarski, I.I. 1967. On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics*, **20**, 69–76.

Moroni, Gloria, Nicoletti, Cheti, & Tominey, Emma. 2019. *Child Socio-Emotional Skills: The Role of Parental Inputs*. IZA Discussion Paper 12432.

Pavan, Ronni. 2016. On the Production of Skills and the Birth Order Effect. *Journal of Human Resources*, **51**(3), 699–726.

Ronda, Victor. 2017. *The Effect of Maternal Psychological Distress onChildrens Cognitive Development*. unpublished manuscript.

Table 1: % Bias in Adult Outcome Effect

Panel A: Time Varying μ_t and λ_t

	translog	CES		
		DRS	CRS	IRS
Normalizations:				
$\tilde{\mu}_t = 0$, $\tilde{\lambda}_t = 1$ for all t	0%	-12.3%	-12.1%	-10.6%
Tech. restrictions + *initial normalization*	0%	-16.6%	-16.7%	-15.3%

Panel B: Age invariant $\mu_t = 2$ and $\lambda_t = 5$

	translog	CES		
		DRS	CRS	IRS
Normalizations:				
$\tilde{\mu}_t = 0$, $\tilde{\lambda}_t = 1$ for all t	0%	-12.3%	-12.3%	-11.3%
Tech. restrictions + *initial normalization*	0%	-13.6%	-14.5%	-13.3%

Notes: In this table we show the impact of an increase of one standard deviation in initial log income on an adult outcome, according to the model specified in section 3. In panel A, the true loading factors for both skills take the values 5, 2, 10 and 7 over time, while the mean takes the values 0, 1, 2 and 3. In panel B, we impose age-invariance and set the true loading factors for both skills according to $\lambda_t = 5$ and $\mu_t = 2$ for all t. All other parameters have been chosen arbitrarily but are reasonable given the estimates in the literature. See appendix Table A1. When estimating the model under technology restrictions, we impose an initial normalization of $\tilde{\mu}_1 = 0$ and $\tilde{\lambda}_1 = 1$.

Table 2: % Bias in Standardized Skills

Panel A: Time Varying μ_t and λ_t

	translog	CES		
		DRS	CRS	IRS
Normalizations:				
$\tilde{\mu}_t = 0$, $\tilde{\lambda}_t = 1$ for all t	0%	-15.7%	-15.5%	-14.9%
Tech. restrictions + *initial normalization*	0%	-54.5%	-63.4%	-60.3%

Panel B: Age invariant $\mu_t = 2$ and $\lambda_t = 5$

	translog	CES		
		DRS	CRS	IRS
Normalizations:				
$\tilde{\mu}_t = 0$, $\tilde{\lambda}_t = 1$ for all t	0%	-15.6%	-15.5%	-15.1%
Tech. restrictions + *initial normalization*	0%	-29.6%	-46.8%	-46.6%

Notes: In this table we show the impact of an increase of one standard deviation in initial log income on the standardized final level of the first skill, according to the model specified in section 3. In panel A, the true loading factors for both skills take the values 5, 2, 10 and 7 over time, while the mean takes the values 0, 1, 2 and 3. In panel B, we impose age-invariance and set the true loading factors for both skills according to $\lambda_t = 5$ and $\mu_t = 2$ for all t. All other parameters have been chosen arbitrarily but are reasonable given the estimates in the literature. See appendix Table A1. When estimating the model under technology restrictions, we impose an initial normalization of $\tilde{\mu}_1 = 0$ and $\tilde{\lambda}_1 = 1$.

Appendix

A Monte Carlo Exercise

In this appendix we present the results of a simple Monte Carlo exercise. The purpose is to illustrate that the biases (or lack thereof) in Tables 1 and 2 are not specific to the parameter values selected. We repeat the same simulation exercise utilized in Tables 1 and 2 two-hundred times and vary at each iteration the value of the true parameters, randomly drawing them from a wide range of possible values. The ranges for the technology parameters specifically are shown in Table A1. The model is identical to the one outlined in section 3 and utilized for Tables 1 and 2, with the only difference that at each iteration we only simulate 100,000 individuals. For each iteration, we fix the measurement parameters in the age-varying and age-invariant cases equal their counterparts from the simulation exercises exhibited in Tables 1 and 2.

Table A2 reports the average and the standard deviation of the absolute value of the bias across all iterations for each normalization and true technology. The results are consistent with the findings in Table 1. The translog technology yields unbiased estimates of the impact a one standard deviation change in initial income on adult outcomes regardless of the normalization imposed. In contrast, the CES yields biased estimates for all normalizations, regardless of whether an age-invariant measure is available. This is the result of incorrectly normalizing the initial scale parameter.

Table A1: Parameter Values for the Simulation

Parameters	Tables 1 and 2 $t=1$	$t=2$	$t=3$	Monte Carlo All t
Translog for θ_1				
A_{1t}	1	2	3	[0 , 3]
γ_{11t}	0.6	0.7	0.8	[0.5 , 1.5]
γ_{12t}	0.3	0.2	0.2	[0.2 , 0.6]
γ_{13t}	0.3	0.3	0.2	[0.2 , 0.8]
γ_{14t}	-0.1	-0.1	-0.1	[-0.1 , 0.2]
γ_{15t}	0.1	0.2	0.2	[-0.1 , 0.2]
γ_{16t}	0.1	0.1	-0.1	[-0.1 , 0.2]
$\sigma_{\theta_{1t}}$	0.6	0.5	0.4	[0.3 , 0.6]
Translog for θ_2				
A_{2t}	2	2	1	[0 , 2]
γ_{21t}	0.1	0.1	0.1	[0.1 , 0.4]
γ_{22t}	0.6	0.7	0.8	[0.5 , 1.5]
γ_{23t}	0.5	0.4	0.3	[0.2 , 0.6]
γ_{24t}	0.1	0.1	0.2	[-0.1 , 0.2]
γ_{25t}	-0.1	-0.1	-0.1	[-0.1 , 0.2]
γ_{26t}	0.1	0.2	0.2	[-0.1 , 0.2]
$\sigma_{\theta_{2t}}$	0.6	0.5	0.4	[0.3 , 0.6]
Investment Function				
α_{0t}	1	2	3	[0 , 3]
α_{1t}	0.2	0.3	0.4	[-0.01 , 0.4]
α_{2t}	0.2	0.2	0.1	[-0.01 , 0.3]
α_{3t}	0.3	0.3	0.3	[0.1 , 0.5]
σ_{I_t}	0.3	0.3	0.4	[0.2 , 0.5]
CES for θ_1				
A_{1t}	1	2	1	[0 , 2]
γ_{11t}	0.6	0.7	0.6	[0.33 , 0.88]
γ_{12t}	0.2	0.1	0.2	[0.05 , 0.45]
γ_{13t}	0.2	0.2	0.2	[0.05 , 0.45]
ϕ_{1t}	0.7	0.2	-1.0	[-1.0 , 0.8]
$\sigma_{\theta_{1t}}$	0.5	0.5	0.4	[0.3 , 0.5]
ψ_{1t} for DRS	0.7	0.7	0.7	[0.6 , 0.8]
ψ_{1t} for IRS	1.3	1.7	2.3	[1.1 , 2]
CES for θ_2				
A_{2t}	2	2	1	[0 , 2]
γ_{21t}	0.1	0.1	0.1	[0.07 , 0.42]
γ_{22t}	0.6	0.7	0.8	[0.37 , 0.82]
γ_{23t}	0.3	0.2	0.1	[0.07 , 0.42]
ϕ_{1t}	-0.5	-0.5	-0.5	[-1.0 , 0.7]
$\sigma_{\theta_{2t}}$	0.5	0.4	0.3	[0.3 , 0.5]
ψ_{2t} for DRS	0.7	0.7	0.7	[0.6 , 0.8]
ψ_{2t} for IRS	2	2	2	[1.1 , 2]

Notes: This table contains the parameter values utilized for the simulations in Table 1 and 2 and the Monte Carlo exercise of appendix Table A2. The intervals for the Monte Carlo column represent the range of the parameter values from which the parameter was drawn at each iteration. For brevity we do not report the parameter values for the initial distribution of the random variables, for the income function, and the adult outcome equation.

Table A2: % Bias in Adult Outcome Effect - A Monte Carlo Exercise

Panel A: Time Varying μ_t and λ_t

	translog	CES		
		DRS	CRS	IRS
$\tilde{\mu}_t = 0$, $\tilde{\lambda}_t = 1$ for all t				
Average Absolute Value Bias	0%	25%	22%	23%
Standard Deviation Absolute Value Bias	0%	32%	27%	27%
Tech. restrictions + *initial normalization*				
Average Absolute Value Bias	0%	29%	28%	27%
Standard Deviation Absolute Value Bias	0%	35%	30%	29%

Panel B: Age invariant $\mu_t = 2$ and $\lambda_t = 5$

	translog	CES		
		DRS	CRS	IRS
$\tilde{\mu}_t = 0$, $\tilde{\lambda}_t = 1$ for all t				
Average Absolute Value Bias	0%	24%	22%	23%
Standard Deviation Absolute Value Bias	0%	30%	26%	26%
Tech. restrictions + *initial normalization*				
Average Absolute Value Bias	0%	25%	25%	24%
Standard Deviation Absolute Value Bias	0%	30%	27%	27%

Notes: In this table we show the average impact of an increase of one standard deviation in initial log income on an adult outcome, according to the model specified in section 3. We repeat the exercise 200 times and report the average and the standard deviation of the absolute value of the bias. At each iteration we vary the model parameters, picking them randomly from a reasonable range. The ranges for the technology parameters are presented in appendix Table A1. The measurement parameters do not vary across iterations. In panel A, the true loading factors for both skills take the values 5, 2, 10 and 7 over time, while the mean takes the values 0, 1, 2 and 3. In panel B, we impose age-invariance and set the true loading factors for both skills according to $\lambda_t = 5$ and $\mu_t = 2$ for all t. When estimating the model under technology restrictions, we impose an initial normalization of $\tilde{\mu}_1 = 0$ and $\tilde{\lambda}_1 = 1$.