Received: 27 March 2021 | Revised: 31 May 2022 | Accepted: 3 June 2022

DOI: 10.1002/jae.2929

WILEY Journal of APPLIED ECONOMETRICS

RESEARCH ARTICLE

Identification of dynamic latent factor models of skill formation with translog production

Emilia Del Bono[1] | **Josh Kinsler**[2] | **Ronni Pavan**[3]

[1]Institute for Social and Economic Research, University of Essex, Colchester, UK

[2]University of Georgia, Athens, Georgia, USA

[3]University of Rochester, Rochester, New York, USA

Correspondence
Ronni Pavan, Economics Department, University of Rochester, Rochester, NY 14627 USA.
Email: rpavan@ur.rochester.edu

Funding information
National Science Foundation, Grant/Award Number: 1725270; Economic and Social Research Council, Grant/Award Number: ES/S012486/1

Summary

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 vary with the chosen normalizations. Access to age-invariant measures does not solve this problem since arbitrary scale and location restrictions are still imposed in the initial period. We theoretically prove the normalization invariance of the translog production function and then complete several empirical exercises illustrating the effects of location and scale normalizations for different technologies and types of skills measures.

KEYWORDS

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

1 | INTRODUCTION

Dynamic latent factor models have become a popular tool for studying child development (Agostinelli & Wiswall, 2020; Attanasio et al., 2017, 2020; Cunha et al., 2010; Del Bono et al., 2020; Pavan, 2016). In these models, unobserved child skills evolve dynamically according to a specified production technology, where the inputs often include lagged child skills, parental skills, and 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 point identify the measurement model and production technology, a set of location and scale normalizations is required. Once these normalizations have been implemented, researchers can estimate the measurement and technology parameters and assess how altering inputs across different developmental stages influences the accumulation of skills (Cunha & Heckman, 2008; Cunha et al., 2010).

Two recent papers, Agostinelli and Wiswall (2016) and Agostinelli and Wiswall (2020), make important contributions to this literature. Agostinelli and Wiswall (2016) show that setting the scale and location of the latent variables each period, a practice commonly considered a normalization, results in biased estimates of the technology parameters and marginal effects when a constant elasticity of substitution (CES) production function exhibiting constant return to scale is employed. Agostinelli and Wiswall (2020) expands on the previous paper by illustrating that, under standard assumptions, it is not possible to identify a technology allowing for TFP dynamics and free returns to scale because the location and scale of the technology cannot be separately identified from the location and scale of the measures. They also suggest a possible solution to this identification problem: exploit age-invariant measures of skills when possible.[1] However, as we discuss below, for age-invariance to solve the identification issues, the *initial* scale and location of latent skills must be known, which is not possible since skills are by definition unobserved.

In this paper, we highlight two important results that can help shape future developments in this literature. First, many of the concerns raised by Agostinelli and Wiswall (2020) in relation to the use of location and scale normalizations do not apply when adopting a translog skill technology in combination with a log-linear measurement system. Specifically, location and scale normalizations affect the estimated translog production parameters but have no impact on the implied marginal effects on the long-run outcomes of interest. Therefore, the latter can be consistently recovered even when age invariant measures are not available.

Second, we show that when using a CES production technology, the availability of age invariant measures does not result in skill technology parameters or marginal effects that are insensitive to initial scale and location restrictions. The problem is that using age invariant measures eliminates the need to restrict the scale and location of a latent variable each period, but the initial scale and location are still arbitrarily set by the researcher.

The current literature has not paid sufficient attention to this problem and instead assumes that the scale and location of initial skill are simply set at the correct value. For example, in the Monte Carlo exercise of Attanasio et al. (2020), the initial scale and location of latent skill are set at a value assumed to be the true value, though the authors note that the estimated production parameters and treatment effects can be sensitive to this assumption. A better understanding of the implications of these initial assumptions is important given that many recent dynamic latent factor models employ a CES technology with age invariant measures (Agostinelli et al., 2020; Attanasio et al., 2020; Aucejo & James, 2021) to allow for time-varying TFP and free returns to scale. As we show here, the translog production function does not suffer the same limitations, and thus provides researchers with a practical alternative.

In the first part of the paper (Section 2), we 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 to the ones employed in Cunha et al. (2010) and Agostinelli and 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 child log skill. The fact that we only identify treatment effects is arguably not an issue for researchers, as most studies use the estimated production function parameters 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., 2020; Aucejo & James, 2021).

Although the translog function (Christensen et al., 1973) is a second-order Taylor polynomial approximation of the CES function (Kim, 1992), there are meaningful differences between the two. In contrast to the CES, the translog technology imposes no a priori restrictions on the cross-elasticities of substitution and is therefore a more flexible approach when the true data generating process is unknown.[2] Importantly, the translog technology is also 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 second part of the paper (Section 3) extends our theoretical result for the translog model in a number of important directions. First, using the National Longitudinal Study of Youth 1979 Child and Young Adult Data (CNLSY), we confirm that regardless of which location and scale normalization is implemented to point identify the model, the implied treatment effects in the translog case are unaffected. In contrast, treatment effect estimates using a CES specification are

[1]An age invariant measure is one where differently aged children with identical latent skill levels attain the same value for the measure on average. In other words, an age-invariant measure implies a restriction on the location and scale of the latent skill over different time periods.

[2]The basic translog specification can be extended to include higher order polynomials of log inputs, meaning that it can approximate any underlying data generating process. In Section 3.2, we discuss that even in an extended translog model, treatment effects are invariant to location and scale normalizations.

sensitive to the chosen location and scale restrictions.[3] Importantly, the availability of an age-invariant skill measure, the PIAT math score, does not eliminate treatment effect sensitivity to the initial scale and location normalization for the CES. Second, we demonstrate through simulation that our theoretical result using a simple translog specification holds for more flexible translog models and can accommodate multiple child skills. Finally, we discuss that even when the true underlying skill process follows a CES, a model based on a translog technology yields more accurate treatment effect estimates than a model based on a CES function with incorrect scale and location normalizations.

In a contemporaneous and independently developed paper, Freyberger (2021) also studies identification of normalization free parameters in the context of dynamic latent factor models. His results are more general along a number of dimensions. First, with respect to the translog specification, Freyberger (2021) expands the set of policy relevant parameters that can be identified absent scale and location restrictions. Second, Freyberger (2021) shows that normalization free treatment effects can be identified even in the case of a CES production function as long as the scale of latent skill is estimated along with the technology and measurement parameters. Finally, Freyberger (2021) extends these results to a general nonparametric model for skill production and measurement.

Our paper is narrower in focus but speaks more directly to applied researchers interested in estimating dynamic latent factor models. In particular, we focus on the advantages of adopting a translog specification versus the more commonly used CES, pointing out shortcomings of the latter, which have not been fully appreciated in the empirical literature thus far. A key advantage of the translog approach is that it easily expands to include higher order polynomials to closely approximate any underlying skill technology, including a CES. We illustrate this point explicitly in Section 3.3. Importantly, we show that when using a flexible translog model, treatment effect estimates remain invariant to location and scale normalizations and that this is also true with multiple latent skills. Relative to Freyberger (2021), we prove that it is possible to anchor treatment effects to the standard deviation of latent skills, which allows for easy comparisons with the broader child development literature. Finally, we demonstrate that normalization free parameters can be recovered even when researchers follow a multi-step estimation approach, which has become the norm in the field.

2 | INVARIANCE OF TRANSLOG TREATMENT EFFECTS

The setup of our theoretical framework builds from the models in Cunha et al. (2010) and Agostinelli and 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:

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

Parental investment is determined by

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

We assume an adult outcome is available and given by

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

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

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

[3] The fact that the estimates of interest are affected by the scale of the inputs when adopting a CES production function has also been noted in the dynamic macroeconomic literature. Klump et al. (2012) review this literature and show the importance of normalizing the inputs of a CES in an economically meaningful way using a baseline or reference point. This shows that the identification issues we explore in this paper extend beyond the estimation of latent factor models and apply to any type of scaling problem related to input variables.

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 and 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. Finally, all measurement errors (including η_Q) are assumed to be mean zero and independent across measures, across individuals, and over time. Additionally, all measurement errors are assumed independent of the latent variables, household income, and the structural shocks $(\eta_{\theta t}, \eta_{It})$. Specific distributional assumptions for these shocks are not necessary at this point.

While the above equations define the dynamic features of the factor model, the initial conditions must be specified. Defining $\Omega = (\ln \theta_1, \ln Y_1)$, we assume $\Omega \sim F(\mu_\Omega, \Sigma_\Omega)$. The correlation between initial skill and income means that income and skill will be 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 Equation (1) are identified absent any location and scale normalization of the underlying latent factors. Treatment effects can be anchored using either the available adult outcome or the standard deviation of latent skill.

2.1 | Identifying signal and reduced form components

Our proof begins by identifying key components of the signal, investment, and production models.

Theorem 1. *Assuming that two measures of child skill and parental investment are available each period, the joint densities of* $Z^*_{\theta,t,m_t} = \mu_{\theta,t,m_t} + \lambda_{\theta,t,m_t} \ln \theta_t$ *and* $Z^*_{I,t,l_t} = \mu_{I,t,l_t} + \lambda_{I,t,l_t} \ln I_t$ *are identified given the previous assumptions and additional technical restrictions.*

Proof. See the Online Appendix Section A1. □

For this result, we rely primarily on the i.i.d. nature of the errors in the measurement equations both across measures and over time. Note that we do not separately identify the mean and loading for each measure, simply the joint density of all signal components.

Next, we recover several reduced form parameters describing the investment and production models. Consider the investment model first. We begin with the measurement equation for Z_{I,t,l_t} and substitute on the right hand side for $\ln I_t$ using Equation (2). This provides a link between the observed investment measure, child skill, and family income as a function of the investment parameters. Child skill is unobserved, so we substitute further, replacing $\ln \theta_t$ with an expression derived from the measurement equation Z_{θ,t,m_t}. Ultimately, we obtain the following reduced form equation:

$$Z_{I,t,l_t} = \beta^0_{I,t}(m_t, l_t) + \beta^1_{I,t}(m_t, l_t) Z_{\theta,t,m_t} + \beta^2_{I,t}(m_t, l_t) \ln Y_t + \epsilon_I(m_t, l_t),$$

linking an observed investment measure with a skill measure and parental income. The reduced-form coefficients ($\beta_{I,t}$'s) are functions of the underlying investment model parameters, as well as the measurement parameters.[4] Since Z_{θ,t,m_t} will be correlated with the error term, to identify the $\beta_{I,t}$'s, we follow Agostinelli and Wiswall (2020) and use Z_{θ,t,m'_t} as an instrument for Z_{θ,t,m_t}. The arguments of the $\beta^j_{I,t}$ reflect the specific investment and child skill measures utilized in the regression. Going forward, we will suppress these arguments when possible.

We take a similar approach to identify a reduced-form model of skill production. In this case, we begin with a measurement equation for child skill at time $t+1$ ($Z_{\theta,t+1,m_{t+1}}$) and substitute for $\ln \theta_{t+1}$ using Equation (1). Substituting further for child skill and investment at time t using expressions derived from their associated measures yields

$$\begin{aligned} Z_{\theta,t+1,m_{t+1}} = {} & \beta^1_{\theta,t}(m_{t+1}, m_t, l_t) + \beta^2_{\theta,t}(m_{t+1}, m_t, l_t) Z_{\theta,t,m_t} \\ & + \beta^3_{\theta,t}(m_{t+1}, m_t, l_t) Z_{I,t,l_t} + \beta^4_{\theta,t}(m_{t+1}, m_t, l_t) Z_{\theta,t,m_t} \cdot Z_{I,t,l_t} + \epsilon_\theta(m_{t+1}, m_t, l_t), \end{aligned}$$

[4]The precise mapping between the investment reduced-form and structural parameters is derived in the Online Appendix Section A2.

where the $\beta_{\theta,t}$'s are functions of the underlying skill production and measurement equation parameters.[5] ϵ_θ is a mean zero error term correlated with Z_{θ,t,m_t}, Z_{I,t,l_t}, and $Z_{\theta,t,m_t} \cdot Z_{I,t,l_t}$. Using 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 $m'_t \neq m_t$, we can recover the $\beta^j_{\theta,t}(m_{t+1}, m_t, l_t)$'s as in Agostinelli and Wiswall (2020). Again, the arguments of $\beta^j_{\theta,t}$ reflect the specific child skill and investment measures utilized in the regression, which we will suppress when possible.

2.2 | Identifying the marginal effect of income

As noted in Section 1, it is common for researchers to use the estimated technology to simulate how child skills are impacted by interventions in various periods of development. We first 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.[6] 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 above equation allows us to implement a recursive strategy to rewrite the marginal effect of income in a useful way.

Theorem 2. *Recursively substituting using the derivatives of Equations (1) and (2) and the previously defined coefficients $\beta^j_{I,t}(m_t, l_t)$ and $\beta^j_{\theta,t}(m_{t+1}, m_t, l_t)$, we can write $\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t}$ as*

$$\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t} = \prod_{s=1}^{k-1}\left[\left(\beta^2_{\theta,t+s} + \beta^4_{\theta,t+s}Z^*_{I,t+s,l_{t+s}}\right) \times \beta^1_{I,t+s}\left(\beta^3_{\theta,t+s} + \beta^4_{\theta,t+s}Z^*_{\theta,t+s,m_{t+s}}\right)\right] \times \frac{\beta^2_{I,t}}{\lambda_{\theta,t+k,m_{t+k}}}\left(\beta^3_{\theta,t} + \beta^4_{\theta,t}Z^*_{\theta,t,m_t}\right). \quad (3)$$

Proof. See the Online Appendix Section A4. □

Because the reduced form β's are known and the joint distribution of the skills and investment signals is known (from Theorem 1), the above result implies that the distribution of $\frac{\partial \ln \theta_{t+k}}{\partial \ln Y_t}$ is identified up to a scaling factor, $\frac{1}{\lambda_{\theta,t+k,m_{t+k}}}$. This scaling factor is the inverse of the loading factor for $\ln \theta_{t+k}$ and can be any non-zero number. As a result, it is not particularly interpretable.

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.[7,8] Cunha and Heckman (2008) and Cunha et al. (2010) advocate anchoring to adult outcomes of the child, such as earnings or completed schooling, which have a well defined cardinal scale. However, this is not always possible, and our second approach provides a useful alternative.

2.2.1 | Anchoring to an adult outcome

In this case, the derivative of interest becomes $\frac{\partial Q}{\partial \ln Y_t} = \alpha_Q \frac{\partial \ln \theta_T}{\partial \ln Y_t}$. Using the results from Theorems 1 and 2, and because we can choose the adult outcome as a measure of the child skill (and therefore $\lambda_{\theta,T,m_T} = \alpha_Q$), we can write

[5]The precise mapping between the production function reduced-form and structural parameters is derived in the Online Appendix Section A3.

[6]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.

[7]Notice that in both cases, the treatment effects are functions of log skills. Freyberger (2021) shows that if we instead consider outcomes that depend on the level of skills, the resulting treatment effects would depend on the scale and location restrictions. 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.

[8]It should also be noted that while Freyberger (2021) considers treatment effects anchored to adult outcomes or in terms of quantiles of the skill distribution, utilizing standardized log skills is a unique feature of this paper.

$$\frac{\partial Q}{\partial \ln Y_t} = \prod_{s=1}^{k-1} \left[\left(\beta^2_{\theta,t+s} + \beta^4_{\theta,t+s} Z^*_{I,t+s,l_{t+s}} \right) \times \beta^1_{I,t+s} \left(\beta^3_{\theta,t+s} + \beta^4_{\theta,t+s} Z^*_{\theta,t+s,m_{t+s}} \right) \right] \times \beta^2_{I,t} \left(\beta^3_{\theta,t} + \beta^4_{\theta,t} Z^*_{\theta,t,m_t} \right), \tag{4}$$

which means that the distribution of the marginal effect of income on the adult outcome is identified. 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 expected increase in skills measured in adult outcome units is

$$E(\Delta Q | \Delta \ln Y_1 = \delta) = \int \delta \times \frac{\partial Q(Z^*_{\theta,m}, Z^*_{I,l})}{\partial \ln Y_1} dF(Z^*_{\theta,m}, Z^*_{I,l}),$$

where $F(Z^*_{\theta,m}, Z^*_{I,l})$ is a compact notation to represent the joint density of the signal content of the measures and $\frac{\partial Q(Z^*_{\theta,m}, Z^*_{I,l})}{\partial \ln Y_1}$ is the derivative previously identified. 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 percentage increase in income, we can also identify the treatment effects associated with different changes in income for different populations.[9,10]

2.2.2 | Anchoring to the standard deviation of skill

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

$$\sigma_{\ln \theta,t} = \frac{1}{\lambda_{\theta,t,m}} \sqrt{\frac{cov\left(Z_{\theta,t,m_t}, Z_{\theta,t,m'_t}\right) cov\left(Z_{\theta,t,m_t}, Z_{\theta,t+1,m_{t+1}}\right)}{cov\left(Z_{\theta,t,m'_t}, Z_{\theta,t+1,m_{t+1}}\right)}}$$

for any t. In other words, we can identify the standard deviation of latent log-skills at time $t + k$ up to the same scaling factor that we identified the marginal effect of $\ln Y_t$ on log-skills at time $t + k$ (see Equation (3)). Combining these two results 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, where we define standardized skills as $\ln \tilde{\theta}_t = \frac{\ln \theta_t}{\sigma_{\ln \theta,t}}$.[11] 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

$$E(\Delta \ln \tilde{\theta}_T | \Delta \ln Y_1 = \delta) = \int \delta \times \frac{\partial \ln \tilde{\theta}_T(Z^*_{\theta,m}, Z^*_{I,l})}{\partial \ln Y_1} dF(Z^*_{\theta,m}, Z^*_{I,l}),$$

where again the expectation is taken with respect to the joint density of the signal content of the measures.

Expressing the impact of a percent change in income in terms of standard deviations of the underlying log skill is particularly useful because it is a common procedure in this and similar literatures (e.g., Dahl & Lochner, 2012).[12]

3 | EMPIRICAL ILLUSTRATIONS AND EXTENSIONS

In the previous section, we proved that when using a translog production function, appropriately anchored treatment effects can be identified without imposing any location and scale normalizations. We now discuss extensions of this

[9] Additional details available upon request.

[10] It should be noted 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.

[11] In a model with multiple latent skills, researchers can anchor to the skill that is appropriate for the counterfactual of interest. For example, if the object of interest is the impact of a change in income on non-cognitive skill, we could simply anchor to the standard deviation of non-cognitive skill.

[12] A disclaimer is that most of the literature looks at the standard deviations of skills while we look at the standard deviations of log skills, but this is semantic because our measures are functions of log skills while, in these other cases, the measures are the skills.

TABLE 1 Income treatment effects

	$T_Q(1sd)$	$T_Q(10k)$	$T_\sigma(1sd)$	$T_\sigma(10k)$
Translog				
Age Invariance + $\lambda_{PIAT,1} = 1$ & $\mu_{PIAT,1} = 0$	0.0427	0.0173	0.0492	0.0199
Age Invariance + $\lambda_{PIAT,1} = 10$ & $\mu_{PIAT,1} = 5$	0.0427	0.0173	0.0492	0.0199
Rescaled Unit + $\lambda_{PIAT,1} = 1$ & $\mu_{PIAT,1} = 0$	0.0427	0.0173	0.0492	0.0199
No Age Invariance	0.0427	0.0173	0.0492	0.0199
CES				
Age Invariance + $\lambda_{PIAT,1} = 1$ & $\mu_{PIAT,1} = 0$	0.0483	0.0181	0.0534	0.0200
Age Invariance + $\lambda_{PIAT,1} = 10$ & $\mu_{PIAT,1} = 5$	0.0432	0.0167	0.0461	0.0178
Rescaled Unit + $\lambda_{PIAT,1} = 1$ & $\mu_{PIAT,1} = 0$	0.0429	0.0166	0.0457	0.0177
No Age Invariance	0.0488	0.0182	0.0536	0.0200

Note: Income treatment effects are obtained simulating 100,000 child outcomes using production and investment parameters estimated from the CNSLY. We consider increasing income by 1 standard deviation (1sd) or $10,000 (10k) in the initial period. The first two columns anchor treatment effects to an adult outcome (Q), while the final two columns anchor treatment effects to the standard deviation of log skills in the final period (σ). Details on the estimation sample and models are provided in Section 3.1 and Online Appendix Section B1. Each row is estimated separately, where the underlying scale and location normalizations vary across rows. The Age Invariance rows assume the PIAT math measure loading and mean are fixed over time. The Rescaled Unit rows standardize the PIAT math raw scores according to the first period mean and standard deviation and also imposes age invariance. The No Age Invariance row forces the PIAT math measure loading and mean to vary over time.

theoretical result in a number of important directions. Here, we focus on summarizing the key findings, leaving most of the details to the Online Appendix Sections B1–B3.

3.1 | An application of the translog and CES models

Our first empirical exercise uses the National Longitudinal Study of Youth 1979 Child and Young Adult Data (CNLSY) to estimate a series of dynamic latent factor models of cognitive skill development. The baseline model follows Agostinelli and Wiswall (2020) and is a translog specification quite similar to the one outlined in Section 2. We then alter the model to allow for a CES skill production function, a commonly used specification in the literature on skill formation (Attanasio et al., 2020; Aucejo & James, 2021; Cunha et al., 2010). For both types of skill technologies, we investigate the sensitivity of estimated income treatment effects to different location and scale normalizations imposed in estimation. Additional details on the model and data are provided in the Online Appendix Section B1.

We summarize the key results from this exercise in Table 1. In the top panel, we show that estimated treatment effects when using a translog specification are identical under four different location and scale normalizations. The first two normalizations impose the age-invariance assumption for the PIAT math score but set the initial location and scale to different values. The third normalization is identical to the first except that instead of using the raw PIAT math scores, we standardize the scores according to the mean and variance from the initial period. The fourth normalization sets different scale and location normalizations each period for the PIAT math measure, mimicking a situation where no age-invariant measure is available.[13] The first two columns of Table 1 show the impact on the adult outcome (years of schooling) when we increase initial family income by one standard deviation or by $10,000. The final two columns reflect the same treatments but anchor to the standard deviation of log skills in the final period.[14] Given our theoretical result, it is not surprising that the implied treatment effects in the translog case are unaffected by the specific location and scale normalization implemented to point identify the model. When we use a CES production function, however, different

[13]Without age-invariance, researchers still typically set the scale and location normalizations to be constant, despite the possibility that the true scale and location of skill are varying over time. Thus, there is a disconnect between a skill's actual location and scale and the normalizations selected. The situation depicted in the fourth row of the upper panel is equivalent to this scenario.

[14]Treatment effects are calculated by simulating outcomes for 100,000 children using the estimated model parameters. We anchor to the adult outcome, Q, and standardized log-skills according to

$$T_Q = E\left[\alpha_Q \times (\ln\theta'_T - \ln\theta_T)\right],$$

$$T_\sigma = E\left[\frac{\ln\theta'_T - \ln\theta_T}{\sigma_{\ln\theta_T}}\right],$$

where $\ln\theta'_T$ is the counterfactual level of log skill when we either increase initial family income by one standard deviation or $10,000. α_Q is the loading in the adult outcome equation.

location and scale restrictions lead to different estimated treatment effects. This result is illustrated in the second panel of Table 1.

Even when an age invariant measure is available (PIAT math), the CES estimates are sensitive to different normalizations of the *initial* scale and location. A 1 order of magnitude difference in the initial scale, for example, can generate income treatment effects on adult outcomes or standardized skill that vary by more than 10%.[15] So while a researcher may know a measure has a constant mean and loading over time, they will not know the true values. Treatment effect estimates based on a CES will then be sensitive to ad hoc restrictions imposed by the researcher.[16] The sensitivity of CES treatment effect estimates will depend on how different the initial location and scale restrictions are and the true elasticity of substitution. When the true elasticity of substitution is close to one, the CES model converges to a Cobb-Douglas, which is a special case of the translog technology.

Agostinelli and Wiswall (2020) advocate for using age invariant measures as a way to produce interpretable production function estimates, whether it be a translog or CES. However, they do not consider the variation induced by the initial scale and location normalizations, operationalized in their paper by setting the PIAT math loading to one and the mean of initial skill to zero. Changing the value of these normalizations will produce different production function estimates for both the translog and CES, as shown in the Online Appendix Section B1.

3.2 | A more flexible translog

In both our theoretical proof and analysis using the CNLSY, we specify a simple translog production function for one type of child skill. One concern might be that the location and scale invariance properties of this model will not translate to more complicated environments. In the Online Appendix Section B2, we use simulations to show that the treatment effect invariance of the translog specification persists when we analyze a more complicated multi-skill model where the production function also includes third-order interactions and polynomials.

Our motivation in extending the results beyond the basic translog model is driven in part by the link between the translog and the CES functions. The translog production function was developed as an approximation of the CES function by a second-order Taylor polynomial. It is therefore natural to analyze a third-order expansion, which may be a better approximation of the true data generating process.

Using a third-order expansion, we continue to find that appropriately anchored treatment effects are invariant to scale and location normalizations implemented during estimation. Two key features drive this result. First, our extended translog function is linear in the parameters. Second, the arguments are power functions of how skills enter the observed measures. As a consequence, we could change $\ln \theta_t$ in all equations of the model with any continuous and monotonic transformation $f(\theta_t)$ and obtain the same invariance result.

3.3 | Investigating misspecification

Although the translog production function, or a generalization of it, may have the important property of producing identical treatment effects regardless of the chosen normalizations, the estimated treatment effects can be biased if the true data generating process is misspecified. While this is true generally, if a researcher is convinced that the true form of the data generating process follows a CES function, they may be willing to estimate a CES specification even if different normalizations will imply different estimates. In the Online Appendix Section B3, we investigate this situation.

We simulate data from a model where we assume the true underlying production function is a CES.[17] We then estimate a CES specification and show that if the initial location and scale normalizations are sufficiently far away from the truth, the bias induced by this untestable, ad hoc assumption can be larger than the misspecification bias generated by

[15]The variability of the estimated treatment effects is not simply an issue of interpretation, which arises when we estimate the effect of height on wages, for example, and the estimate changes whether height is expressed in meters or feet. Treatment is defined as the impact of a $10,000 increase in initial income on accumulated years of schooling. Neither income or years of education are directly affected by the scale of the cognitive skill. Any sensitivity of the treatment effect to these seemingly innocuous and often hidden scaling choices is undesirable.

[16]An interested reader should look at Freyberger (2021), where the details of the CES restrictions that are truly normalizations, i.e., 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.To the best of our knowledge, these specific normalizations have never been implemented in the literature.

[17]The parameters are chosen to be similar to the ones we estimated using the CNSLY.

instead estimating a translog model. Indeed, using the generalized translog technology discussed in the previous section produces estimated treatment effects that are very close to the truth. Of course, the poor performance of the CES model will depend on the specific parametrization. If the true elasticity of substitution is very close to one, the CES is similar to a Cobb Douglas and therefore also to a translog where the interaction term has no effect. In this case, the chosen normalization would have little impact on the estimated treatment effects. Additional details are available in the Online Appendix Section B3.

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 need 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 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 scale and location chosen even when an age-invariant measure is available. Recent work by Freyberger (2021) shows that some identifying restrictions do not distort treatment effect estimates when using the CES, but these normalizations are quite different from the restrictions implemented in the literature thus far and require that the measurement parameters and the technology are jointly estimated.

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, treatment effect estimates are unaffected by location and scale normalizations, contrary to other widely used parametric forms for skill technology. Second, the translog technology, or its generalization, may provide a good description of the data even when a researcher is unsure about the true shape of the DGP.

ACKNOWLEDGMENTS

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 work originated from discussions with them. We acknowledge funding from the National Science Foundation, Award No. 1725270. Emilia Del Bono also acknowledges funding from the Economic and Social Research Council Grant ES/S012486/1.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [http://qed.econ.queensu.ca/jae/datasets/del_bono001/].

DATA AVAILABILITY STATEMENT

The data and code are available in the JAE data archive and can be found at http://qed.econ.queensu.ca/jae/datasets/del_bono001/.

REFERENCES

Agostinelli, F., Saharkhiz, M., & Wiswall, M. (2020). Home and school in the development of children. NBER Working Paper 26037.

Agostinelli, F., & Wiswall, M. (2016). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. NBER Working Paper 22441.

Agostinelli, F., & Wiswall, M. (2020). Estimating the technology of children's skill formation. NBER Working Paper 22442.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., & Rubio-Codina, M. (2020). Estimating the production function for human capital: Results from a randomized control trial in Colombia. *American Economic Review*, *110*(1), 48–85.

Attanasio, O., Meghir, C., & Nix, E. (2020). Human capital development and parental investment in india. *Review of Economic Studies*, *87*(6), 2511–2541.

Attanasio, O., Meghir, C., Nix, E., & Salvati, F. (2017). Human capital growth and poverty: Evidence from ethiopia and peru. *Review of Economic Dynamics*, *25*, 234–259.

Aucejo, E. M., & James, J. (2021). The path to college education: The role of math and verbal skills. *Journal of Political Economy*, *129*(10), 2905–2946.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1973). Transcendental logarithmic production frontiers. *The Review of Economics and Statistics*, *55*(1), 28–45.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, *43*, 738–782.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), 883–931.

Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, *102*(5), 1927–56.

Del Bono, E., Kinsler, J., & Pavan, R. (2020). Skill formation and the trouble with child non-cognitive skill measures. IZA Discussion Papers 13713.

Freyberger, J. (2021). Normalizations and misspecification in skill formation models. *arXiv preprint arXiv:2104.00473*.

Kim, H. Y. (1992). The translog production function and variable returns to scale. *The Review of Economics and Statistics*, *74*(3), 546–552.

Klump, R., McAdam, P., & Willman, A. (2012). The normalized CES production function: Theory and empirics. *Journal of Economic Surveys*, *26*, 769–799.

Pavan, R. (2016). On the production of skills and the birth order effect. *Journal of Human Resources*, *51*(3), 699–726.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of the article.

How to cite this article: Del Bono, E., Kinsler, J., Pavan, R. (2022). Identification of dynamic latent factor models of skill formation with translog production. *Journal of Applied Econometrics, 37*(6), 1256–1265. https://doi.org/10.1002/jae.2929