Childhood Skill Formation
and
Intergenerational Mobility

Emre Enes Yavuz *

November 23, 2022

Click for the latest version

Abstract

I estimate a childhood skill formation function for the US without making restrictive functional form or distributional form assumptions using the PSID Child Development Supplement dataset. I rely on the nonparametric identification results in the literature and use orthogonal polynomials and quantile regressions to achieve flexibility in functional and distributional forms. I use a simulated EM algorithm for the estimation. I also estimate a CES skill formation function, typically assumed in the literature, for comparison. The estimation results of the flexible case are different from the CES case. In particular, the skill investment is a substitute for the current skill level, which can result from low investment at earlier ages. The skill investment is a complement to parents' education, so more educated parents are more productive. In addition, the uncertainty around the skill function is more negatively skewed for children of high-educated parents. These features cannot be captured by the CES case altogether. I highlight the role of skill formation function in the implications of rising inequality for intergenerational mobility. My results suggest that the functional form of childhood skill formation function can potentially explain a flat mobility trend despite rising inequality in the data.

*I am grateful to Matthias Deopke, Marti Mestieri and Nezih Guner for their exceptional support and guidance. I am also grateful to Manuel Arellano, Joel Mokyr, Marios Angeletos, Serdar Ozkan, Serdar Birinci, George-Levi Gayle, Oksana Leukhina, Matthew Rognlie and all participants of Northwestern Macro Lunch and Federal Reserve Bank of St. Louis Disertation Interns Workshop. All errors are my own.

1 Introduction

Childhood skill formation in the family is where human capital production starts and is likely to have long-term consequences for adulthood outcomes such as income, health, and overall welfare. Therefore it is crucial to understand how the skills are produced in the family to improve the lives of future generations. For example, how inequality is transmitted across generations crucially depends on the childhood skill formation process. The rising inequality increases the dispersion in the parental inputs in skill formation, e.g., in parental time and expenditures across families. However, how this is reflected in children's skills, therefore, in the next generation's adult outcomes, depends on the skill formation function.

Estimating a childhood skill formation function is challenging because the skill is an unobservable abstract concept without a natural measurement unit, and we have several noisy measures, such as cognitive test scores. Although there are nonparametric identification results in the literature, the functional form is typically assumed to be either Cobb-Douglas or constant elasticity of substitution (CES) function with log-additive noise (Cunha, Heckman and Schennach, 2010; Attanasio et al., 2020). These assumptions force a few parameters to capture different features of the skill formation process, such as substitution patterns and returns of skill investment.

In this paper, I estimate a childhood skill production function without a functional form or distributional form assumption using the data from the Panel Study of Income Dynamics Child Development Supplement (PSID CDS) for the US. The data has information about how much time parents spend with children, child-related expenditures, and cognitive test scores for children over three waves for 1997, 2002, and 2007. I rely on the nonparametric identification results in the literature (Cunha, Heckman and Schennach, 2010) and use a simulated EM algorithm with quantile regressions similar to the one developed by Arellano, Blundell and Bonhomme (2017). I allow functional form flexibility using orthogonal polynomials of log inputs and estimate coefficients with quantile regressions to achieve distributional form flexibility. I also estimate a CES case for the sake of comparison.

I found several differences between the estimation results of flexible and CES cases. A child's skill level in the next period is a function of skill investment consisting of parental time and expenditure, current skill level, and parents' education levels. In the flexible case, the investment and the current skill level are substitutes, i.e., the investment is more productive for disadvantaged children with low skill levels, which is the function of past investment. In other words, any missing investment at earlier ages can be substituted at later ages. This result differs from the findings of Cunha, Heckman and Schennach (2010), where they estimate a CES skill formation function using a different dataset, the National Longitudinal Survey of Youth (NLSY). The findings of Agostinelli and Wiswall (2021) align with my results, where they estimate a translog, i.e., Cobb-Douglas, plus an interaction term between investment and the current skill level, also using the NLSY dataset.

Parents' skill level and investment are complements, i.e., high-educated parents are more productive than low-educated parents. However, it is interesting that the returns are decreasing faster for high-educated parents. High-educated parents hit the flat or highly concave part of the skill formation function quicker than low-educated parents.

The uncertainty in the childhood skill production function is more negatively skewed for high-educated families. On average, children of more educated parents acquire higher skills, but because of negative skewness, they are more likely to end up at a lower level of skill distribution.

Agostinelli and Wiswall (2021) is another paper in the literature improving the functional form in estimating childhood skill formation. They estimate a translog production function which consists of a Cobb-Douglas and an interaction term between skill investment and current skill level using the NLSY dataset. This paper uses second-order orthogonal polynomials of log inputs with all interaction terms and estimates the coefficients for different quantiles instead of assuming a log additive noise. This flexibility allows higher moments, such as skewness, to depend on the inputs, not just the mean. The results align where the specifications are common in both papers, although they use a different dataset and estimation technique.

An advantage of the dataset I use (PSID CDS) is having direct measures for the skill investments in terms of parental time and expenditure. In NLSY, only some proxy measures are available, and the papers using this data assume an abstract skill investment. This advantage allows for more tangible policy discussions, such as the effects of monetary subsidies or policies that permit parents to spend more time with their children. Also, the estimated function can be easily used in quantitative macroeconomic models because time and expenditure have direct counterparts in the model settings.

As a potential application of my estimates, I focus on the role of childhood skill formation in intergenerational mobility. High-income families devote more time and expenditure to their children's education and skill development than low-income families, and the gap has been getting wider along with rising inequality (Corak, 2013). One may be worried about lower intergenerational mobility because children of higher-income families can be even more likely to have high income because of the abundance of opportunities, while children of low-income families could be stuck in poverty. However, I showed in a simple theoretical model similar to the one in Becker et al. (2018) but with a general skill production function that the implication of higher inequality for intergenerational mobility depends on the functional form of childhood skill formation. The intuition is straightforward: we see more input dispersion, but the output dispersion, i.e., skills, depends on the functional form of the skill production.

To see how intergenerational mobility moves along with rising inequality, I estimate a trend in intergenerational elasticity in earnings using PSID for 1968-2019. The intergenerational elasticity of earnings seems fixed over time despite rising inequality. My results align with the literature that uses other data or methodologies for a similar time (Lee and Solon, 2009; Chetty et al., 2014; Song et al., 2020).

The flexible estimation results can potentially explain the flat mobility trends. Inequality has been rising because of higher returns to skills, and families reacted to this by increasing their investment in childhood skills. However, the returns are higher for disadvantaged children, which may help them catch up. On the other hand, high-educated parents hit the flat region of the skill formation function. Hence, even if they drastically increase their investment, their children might not benefit significantly. The children of low-educated parents may be able to catch up because of slowly decreasing returns. Also, the more negative risk for more educated parents can contribute to the mean reversion mechanism. Blanden, Doepke and Stuhler (2022) highlights the findings showing that the dispersion in test scores between high and low-income families has not risen much in contrast to the one in the skill inputs, which is consistent with potential implications of my estimation results.

Section 2 describes the data, empirical model, and estimation algorithm and provides the results. Section 3 highlights the crucial role of childhood skill formation in intergenerational mobility using a simple simulation and a theoretical model. It also estimates the trends in intergenerational mobility in earnings and comments on the implications of the estimation results.

2 Childhood Skill Formation Function

2.1 Data and Sample Selection

I use the survey data from Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS). PSID is the longest panel data that follows a nationally representative sample of families in the US since 1968. As supplementary data, CDS collects information about the children at years 1997, 2002 and 2007. At each wave, they started to interview the children under age 12 and their parent and keep following them at later waves.

CDS data has detailed information on the skill investment for children both in terms of parental time and expenditure. Children and parents fills time diaries for a weekday and weekend. The diaries has entries covering entire 24 hours with information about the activity and any adult presence during the activity. I collect the activities that parents participate actively as time inputs to skill formation function. I create a measure of weekly parental time by weighted average by weighting the weekday by 5/7 and the weekend by 2/5. The time diary for weekend is available either for Sunday or Saturday, I normalize them by using the ratio of mean parental time across all children.

The second main input of the skill formation function is child related expenditures such as child care, books, toys, extra curricular activities, summer camps etc. Unfortunately, the first wave 1997 has information only about childcare. In the estimation, I specify the first period expenditure as a latent variable. Intuitively, the later periods expenditures inform what could be the expenditure investment in the first period.

On the output side of skill formation function, there are three cognitive tests available. In the Letter-Word Identification test, as the name suggests, children are asked to identify letters and words. Applied Problems test has basic mathematics questions. Lastly, Passage Completion test provides a sentence with a missing word or phrase and child is expected to choose the best fit among provided options. First two tests are provided for ages above 3, which covers all children, but Passage Completion is only for age 6 and above because the child should have enough language development to be able to understand the questions. There for ages between 3-5, I use two tests as measures of skill. These tests are specially designed by experts such that it measures cognitive skills consistently at any age. Any two children with same underlying skill level but at different ages should have the same expected scores. This is partly achieved by increasingly more difficult questions as the test progress.

CDS is connected to main PSID where I can have a lot of information about all household and parents. I use parents' education as proxy of their skill level. I gather the household income using all income variables such as wages, asset returns, business income. Children's years of education when they become adults are also available and I use it to estimate a mapping from skill level.

I select the children in intact families living with biological parents. I drop observations if it is not available for two consecutive periods or if any variable of interest is missing. After selection I have 679 children in the final dataset.

2.2 Empirical Model

The childhood skill formation function has three main inputs. The first one is skill investment which consists of parental time and child-related expenditures. The second input is the current level of child skill to allow for any dynamic complementarity. Lastly, it depends on the parents' education level to allow heterogeneity in parental productivity.

Let θ_{it+1} be skill level of child i at time $t+1$,

$$\ln\theta_{t+1} = F(\ln I_t, \ln\theta_t, \ln\theta_P, u_t), \tag{2.1}$$

where I_t is investment, aggregate of parental time and child-related expenditure, θ_P is an aggregated term for mother's and father's education and lastly u_{it} is random variable normalized to uniform distribution. Notice that $F(\cdot)$ is in fact the quantile function skills conditional on inputs and any distribution can be expressed in this way.[1]

The aggregate investment consists of parental time and monetary expenditure for child development. It is given by Cobb-Douglas aggregator,

$$\ln I_t = \ln Time_t^{mom} + \gamma_I^{dad} \ln Time_t^{dad} + \gamma_I^{exp} \ln Exp_t, \tag{2.2}$$

where $Time_t^{mom}$ and $Time_t^{dad}$ are time measures for activities with children with active participation of mother and father.

The aggregator of parental education is given by,

$$\ln\theta_P = \ln\theta_{mom} + \gamma_\theta^{dad} \ln\theta_{dad} + \gamma_\theta^{int.} \ln\theta_{dad} \ln\theta_{mom}, \tag{2.3}$$

where the interaction coefficient γ_θ^{int} determines if there is super modularity between parents' skills and allows to analyze any effect of sorting between couples.

In both aggretor equations, the coefficients in the first terms are omitted because it is not going to identified. The coefficients in the second terms are informative only relative to omitted first term. For example, in equation 2.2, γ_I^{dad} tells how much father's time is productive with respect to the mother's time.

Initial distribution of skills are also specified by a quantile function,

$$\ln\theta_0 = F_0(\ln\theta_P^0, age_0, u_0) \tag{2.4}$$

where θ_P^0 is aggregate parents skill with the same functional form in the equation 2.3 but with different coefficients to be estimated and age_0 is the age of the child in the first period of the data set. In the first wave the children are at different ages because the survey started to follow children below age 12. Therefore, earlier periods are missing and it is necessary to include age to control for that. I also include parental education

[1] For example if we had an additive normal noise, i.e.

$$\theta_{t+1} = \tilde{F}(\theta_t, I_t, \theta_P) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1)$$

quantile function representation is given by,

$$\theta_{t+1} = F(\theta_t, I_t, \theta_P, u_t) = \tilde{F}(\theta_t, I_t, \theta_P) + \Phi^{-1}(u_t),$$

where $\Phi^{-1}(\cdot)$ is the inverse of standard normal c.d.f.

because initial skill level can also be interpreted as ability and genetic transmission certaionly plays a role. The interaction terms between parents' education and age allows heteregenous paths for children. For example, a child of high educated parent might be able obtain more skills aver the same age period.

Time and expenditure policy functions are specified in a similar way,

$$\begin{aligned} \ln Time_t^k &= F_k(\ln y_t, \ln\theta_t, \ln\theta_P^k, u_t^k), \quad k \in \{mom, dad\}, \\ \ln Exp_t &= F_k(\ln y_t, \ln\theta_t, \ln\theta_P^{Exp}, u_t^{Exp}), \\ \text{with} \quad & u_t^{mom}, u_t^{dad}, u_t^{Exp} \sim U(0,1), \end{aligned} \tag{2.5}$$

where y_t is total household income and θ_P^{tmom} is aggregate parents skill. Specifying policy function in this flexible reduced way allows them to be consistent with multiple models. Skill investments are allowed be endogenous and they can depend on the all relevant state variables, however it is assumed that there is no underlying unobservable heterogeneity that affects investment desicion. It is somewhat strong assumption but possible to be addressed this kind endogeneity using instrumental variable approzch for example using past household income as an instrument for investment.

The skills are not directly observable and we have only proxy measures of them. In the data, there are three different cognitive tests for children, namely Letter-Word Identification, Applied (Math) Problems and Passage Completion tests. Each test is designed to provide an age consistent skill measure, i.e. two children with same underlying skills but at different ages should have the same expected score.

Question level information is available. So it is possible to see whether a child answered a question correct or not. Let $Prob_{mq}$ be the probability of answering question i correct in test m,

$$Prob_{mq} = \frac{exp\,(\alpha_m + \beta_m\theta - d_q)}{1 + exp\,(\alpha_m + \beta_m\theta - d_{mq})}, \tag{2.6}$$

where θ is the underlying skill level, α_m and β_m are location and scale parameters of the test m and d_{mq} is the difficulty level of the question q. There is no natural unit of mesaurement for skills so I can set a measurement unit by normalizing one of the tests' parameters. I normalize $\alpha_m = 0$, $\beta_m = 1$ for Letter-Word Identification test and $d_{mq} = 0$ for a question the test.

Lastly, I estimate a mapping between skill levels in the last period and final years of education of the children when they become adults. I specify that the conditional distribution of years of education as binomial distribution with probability p,

$$p_{edu} = \frac{exp\,(f(\ln\theta_T, \ln\theta_P, age_T))}{1 + exp\,(f(\ln\theta_T, \ln\theta_P, age_T))}, \tag{2.7}$$

where θ_T is the skill level of child and age_T is the age of the child in the last period T. Binominal distribution provides a symmetric discrete distribution around a mean years of education conditional on the skill level, age and parental education.

The identification results in the literatre, for example in Cunha, Heckman and Schennach (2010), works in this case as well for identification of the non-parametric production function with multiple measures available for skills, e.g. test scores. The intuition is that for every period there multiple measures of skills that contains information on skills. Across measures the underlying skill level is common but the uncertainty is not common. This identidy the joint skill distribution for each period. In a recent paper, Agostinelli and Wiswall (2021) showed that there is a trade off between restrictions on the shape of function and measurement equation. If the measurement equation is not restricted, any change in measurement over time could come from either change in underlying skill or change in the measurement. For example, the test could be getting easier with age even if the skill level is same. I put restrictions on measurement by assuming parameters in equation 2.6 are fixed for all ages.

2.3 Estimation

The aim is to estimate the skill formation function in the equation 2.1 without restricting functional and distribution form assumptions. For the functional for flexibility, I use hermite polynomials to approximate the unknown true function. For the distributional form flexibility, I estimate the parameters for orthogonal polynomials for a set of quantiles.

$$F(\ln I_t, \ln \theta_t, \ln \theta_P, u_t) = \sum_{k=0}^{K_\theta} a_k(u_t)\varphi_k(\ln I_t, \ln \theta_t, \ln \theta_P), \tag{2.8}$$

where $\varphi_k(\cdot)$'s are orthogonal polynomials and $a_k(u_t)$ are coefficients specific to the quantile $u_t \in (0,1)$. In practice, I choose all interactions of second order polynomials of each input. I take a grid of quantiles $\{u^0, u^1, \ldots, u^L\} \in (0,1)$ and estimate coefficients with quantile regression. I choose $L = 7$ with equidistant grid points in unity. For the off the grid quantiles, I use linear interpolation. Lastly, I fit an exponential tail to avoid plat tails in the distribution and it is estimated with maximum likelihood estimator. The complete quantile function is given by,

$$
\hat{F}(\cdot, u_t) = \begin{cases} \sum_{k=0}^{K_\theta} a_k(u_1)\varphi_k(\cdot) + \frac{\ln(u_t/u^1)}{\lambda_1} & \text{if} \quad u_t \leq u^1 \\ \sum_{k=0}^{K_\theta} \frac{u_t - u^l}{u^{l+1}-u^l}[a_k(u^{l+1}) - a_k(u^l)]\varphi_k(\cdot) + a_k(u^l)\varphi_k(\cdot) & \\ \qquad\qquad\qquad \text{if} \quad u^l < u_t \leq u^{l+1} & , \\ \sum_{k=0}^{K_\theta} a_k(u^L)\varphi_k(\cdot) - \frac{\ln((1-u_t)/(1-u^L))}{\lambda_L} & \text{if} \quad u_t > u^L \end{cases} \tag{2.9}
$$

where $\cdot$ is place-holder for the inputs of the production function, λ_1 and λ_L are parameters of exponential tails.

I follow the same approach for the investment policy functions (2.5) and the initial skill distribution (2.4).

The skill measures (2.6) and the mapping from final skills to the years of education (2.7) are estimated with maximum likelihood estimator given their parametric specification.

Also in the CES specification, I replace the skill formation function with the CES production function with a normal noise.

$$
\begin{aligned} \ln \theta_{t+1} &= F_{CES}(I_t, \theta_t, \theta_P) + \epsilon_t \\ &= \ln \left[A \left(\alpha_I I_t^\phi + \alpha_\theta \theta_t^\phi + \alpha_{\theta_P} \theta_P^\phi \right)^{\frac{1}{\phi}} \right] + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\theta), \end{aligned} \tag{2.10}
$$

with $\alpha_\theta + \alpha_I + \alpha_{\theta_P} = 1$. The parameters of CES production function is estimated with maximum likelihood based on the Normality assumption.

As the estimation algorithm, I follow Arellano and Bonhomme (2016); Arellano, Blundell and Bonhomme (2017) and use a simulated EM algorithm. Both quantile regressions and maximum likelihood estimation requires to solve some kind of optimization problem with an objective function depending on the parameters, the data and unobservable skills. Let X, denote the all the data, Λ all parameters and Θ underlying unobservable skill level. Then the true parameter values solves a set of optimization problems,

$$
\Lambda = \arg\min_{\tilde{\Lambda}} \mathbb{E}\left[R(\Theta, X, \tilde{\Lambda}) \right] \tag{2.11}
$$

where $R(\Theta, X, \tilde{\Lambda})$ includes the moment conditions of quantile regressions and minus log likelihood for parametric parts of empirical model such as skill measurement in equation 2.6. It is not possible to minimize the sample counterpart of the objective function because skills are not observable and should be integrated out. However, that is not possible to because distribution of skills are unspecified and numerical methods are too costly. Even if in the CES case, if I make a normality assumption for initial skills the distribution int he following periods will not be normal because the skill function is nonlinear. Cunha, Heckman and Schennach (2010) uses mixture of normals to approximate the skill distribution and use Kalhman filter to evaluate likelihood. EM algorithm relies on the following alteration of objective function. Applying the law of iterated expectations will give,

$$\Lambda = \arg\min_{\tilde{\Lambda}} \mathbb{E}_X \left[\mathbb{E}_{\Theta|X,\tilde{\Lambda}} \left(R(\Theta, X, \tilde{\Lambda}) \right) \right]. \tag{2.12}$$

Now the dual role of parameters is visible. The first role was already there, the objective function depends on parameters. The second role is in the inner expectation with respect to the distribution of the skills conditional on the data and parameters denoted by $f(\Theta \mid X, \Lambda)$. The main idea is if we fix the parameter values in the inner expectation, we can simulate skills using its conditional distribution and this step is called Expectation (E) step. It is enough to be able to evaluate $f(\Theta \mid X, \Lambda)$ up to a constant to use Markov chain Monte Carlo methods for simulation. With simulated skills, now it is possible to minimize the sample counter part of objective function and get a new set of parameters and this step is called Maximization (M) step. Iteration of these two steps gives a sequence of parameter values converging to the true values.

EM algorithm starts with a set of guessed parameters, say $\hat{\Lambda}^0$ and let the s be the iteration index,

E Step: Simulate M many sample of Θ using conditional density $f(\Theta \mid X, \hat{\Lambda}^s)$ using MCMC.

$$\Theta_{im} \sim f(\Theta \mid X_i, \hat{\Lambda}^s) \quad \text{for} \quad i = 0, \dots, N \quad m = 0, M,$$

where i is index for child. Hence for each child, I draw a sample of skills with size M. Notice that it is possible to derive a density function from skill production function 2.1 since it is a quantile function whose inverse is a cumulative density function whose derivative is density function. The conditional density function also include all parts of the empirical model. In practice I use ensemble MCMC sampler with 100 steps which seems enough to get a random sample after checking autocorrelation of MCMC steps.

M Step: Update the parameters by minimizing the sample counterpart of the objective function using the simulated skills.

$$\hat{\Lambda}^{s+1} = \arg\min_{\tilde{\Lambda}} \frac{1}{NM} \sum_{i=0}^{N} \sum_{m=0}^{M} R(\Theta_{im}, X_i, \tilde{\Lambda}).$$

This iteration gives a sequence converging to the true values but fluctuating around them because of sampling error in the E step. I repeat the iteration $S = 500$ times and the use the last mea of last 250 steps as final estimates, i.e. $\hat{\Lambda} = \frac{1}{S/2} \sum_{s=S/2}^{S} \hat{\Lambda}^s$. If the objective function is well behaved, i.e. convexity and continuity, the convergence is

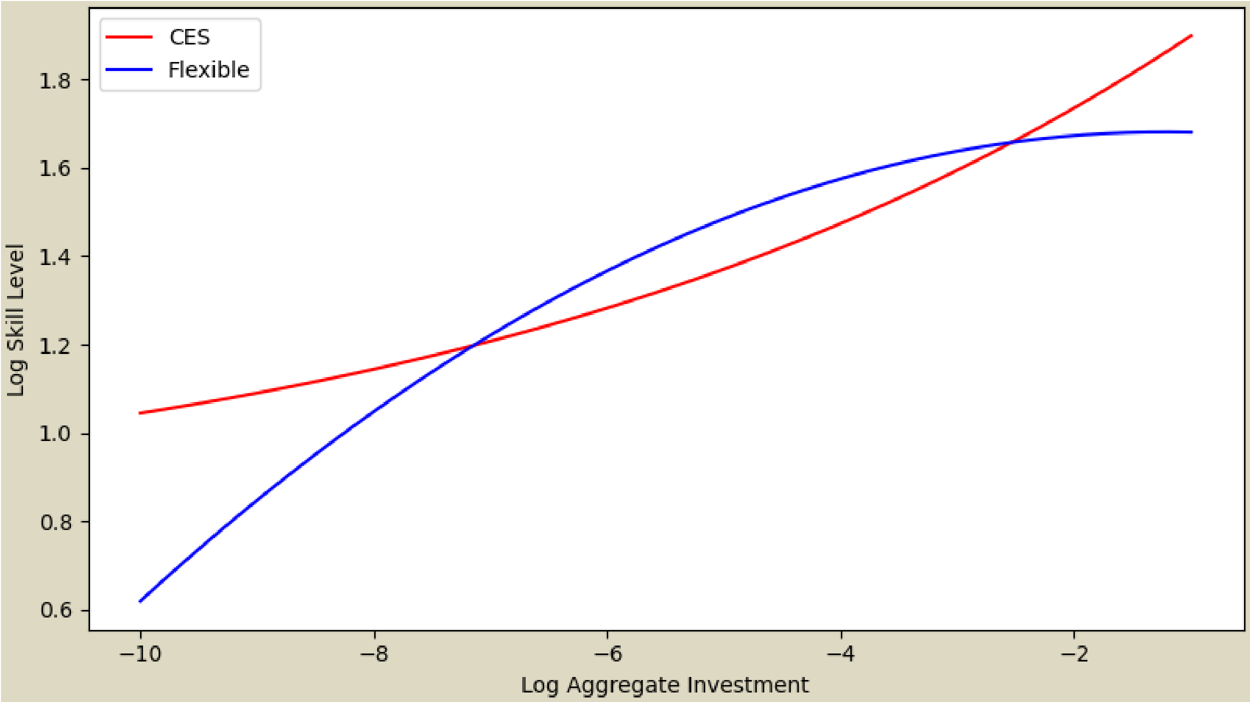


Figure 1: Childhood skill function at mean with respect to aggregate investment. The other two inputs, child's lag skill level and parents' skill level are fixed at their median value.

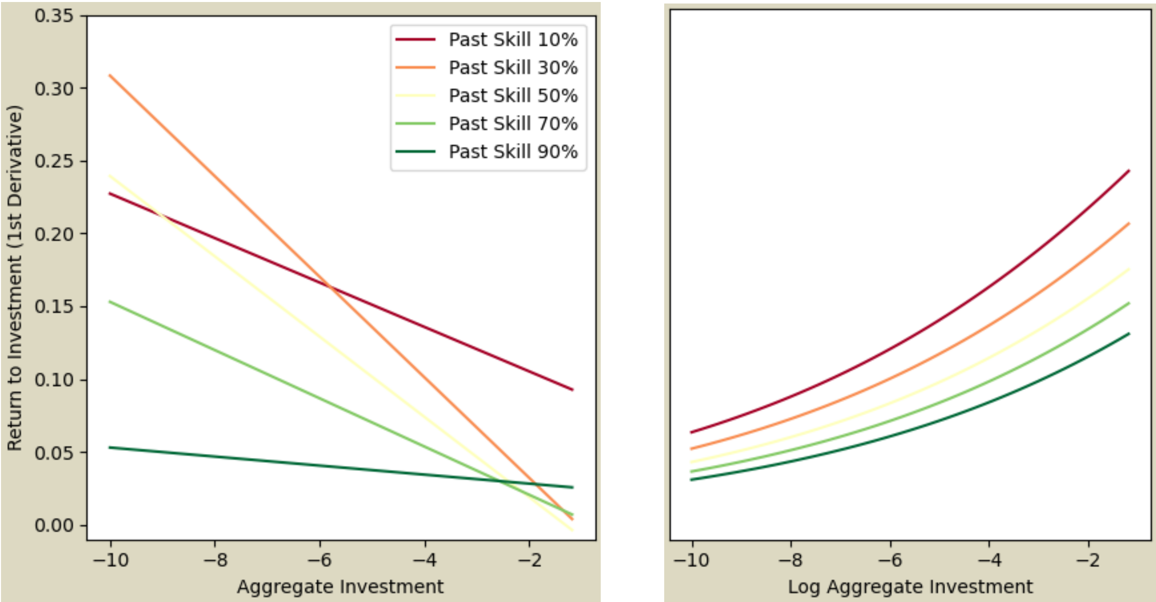
guaranteed to the local minima. I started the algorithm from a wide range of parameter values to be sure that the final estimates are true minimizers of moment conditions.

Standard errors can be obtained by nonparametric bootstrap however this version of the paper does not include them yet because of required computational time, hence the results are preliminary and taken with caution.

2.4 Results

I compare the estimation results of flexible and CES cases by using several plots. Figure 1 plots the skill function at its mean with respect to log of aggregate investment, i.e. $\mathbb{E}_{\theta_t,\theta_P,u_t}\left[F(I_t,\theta_t,\theta_P,u_t)\right]$, expectation taken over the current skill level, parents education level and the uncertainty. While the flexible case is concave in log-log scale, a somewhat strong form of concavity. This means that the returns are extremely small at the high levels of investment. However the CES case is convex because the elasticity of substitution parameter governs this feature and it is in the substitutes region i.e. $\phi > 0$.

In Figure 2, I plot the 1st derivative of the skill function with respect to log aggregate investment for different levels of past skill levels to see the patterns in the return of skill investment, i.e. $\mathbb{E}_{\theta_P,u_t}\left[\frac{\partial F(I_t,\theta_t,\theta_P,u_t)}{\partial \ln I_t}\right]$. Notice that this is also the elasticity of child



(a) Flexible (b) CES with Normal Noise

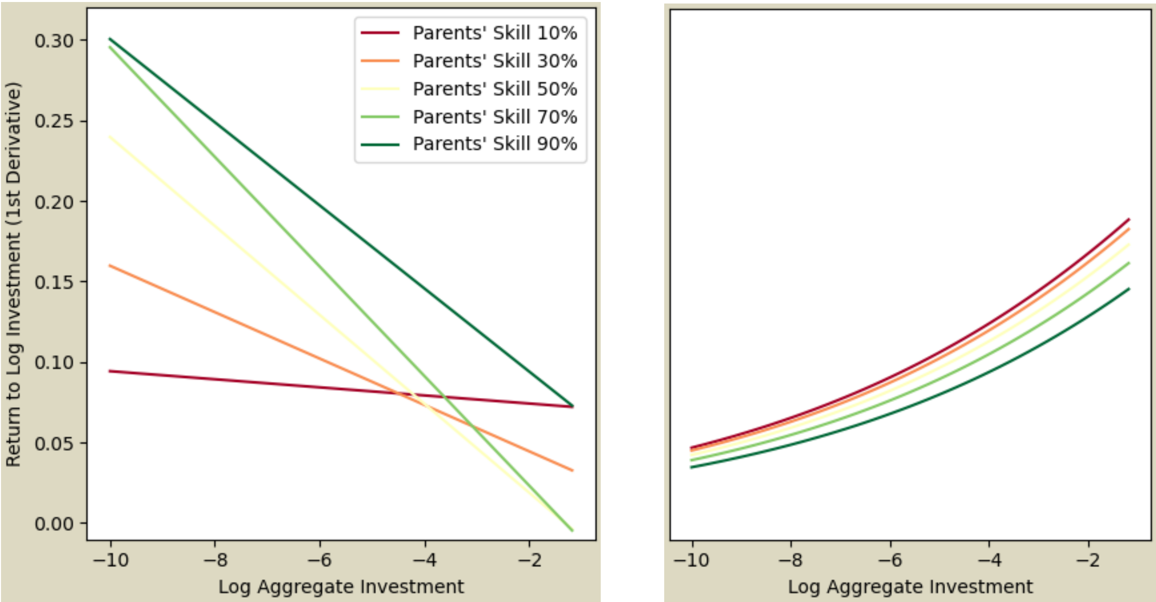
Figure 2: Return (Elasticity) of Childhood Skills with respect to log aggregate investment for different levels of past skill level. Parents' skill level is fixed to its median value.

skills with respect to investment because I use log scale. In both cases the returns are higher for children with lower levels of past skills and this means that the skill level and investment are substitutes.

$$\mathbb{E}_{\theta_P, u_t}\left[\frac{\partial^2 F(I_t, \theta_t, \theta_P, u_t)}{\partial \ln I_t \partial \ln \theta_t}\right] < 0 \tag{2.13}$$

The investment is more productive for currently low skilled children. Current skills level is function of past investments hence this results also means that it is possible to substitute the missing early investment with later investment. This is different from the finding of Cunha, Heckman and Schennach (2010) where they assume CES functional form for the skill formation. However this results in line with Agostinelli and Wiswall (2021) where they include an interaction term between investment and the current skill level on top of the Cobb-Douglas and the estimated coefficient of interaction term is negative. When we look at the pattern of the return, we see that they are decreasing in the flexible case with different speeds for different levels of past skill while it is increasing in CES case.

In Figure 3, I plot a similar graph but this time for different levels of parents' skill, i.e. $\mathbb{E}\left[\frac{\partial F(I_t, \theta_t, \theta_P, u_t)}{\partial \ln I_t}\right]$. We see that in the flexible case more educated parents are more productive hence the parents' skill and investment are complements.



(a) Flexible (b) CES with Normal Noise

Figure 3: Return (Elasticity) of Childhood Skills with respect to log aggregate investment for different levels of parents' skill level. Past skill level is fixed to its median value.

$$\mathbb{E}\left[\frac{\partial^2 F(I_t, \theta_t, \theta_P, u_t)}{\partial \ln I_t \partial \ln \theta_P}\right] > 0 \tag{2.14}$$

However this is not the case in CES, because all inputs are forced to be either complements or substitutes all together. It is interesting to see that in the flexible case, the returns are decreasing for all parents but at a faster for more educated parents.

$$\mathbb{E}\left[\frac{\partial^3 F(I_t, \theta_t, \theta_P, u_t)}{\partial^2 \ln I_t \partial \ln \theta_P}\right] < 0 \tag{2.15}$$

This means that the more educated parents hit the flat part of the skill function more quickly than low educated parents.

Figure 4 plots the skewness in the childhood skill formation with respect to parents' skill level. I use a quantile based skewness measure, defined as,

$$Skewness_t(\theta_P, u) = \mathbb{E}\left[\frac{\frac{F(I_t,\theta_t,\theta_P,1-u)+F(I_t,\theta_t,\theta_P,u)}{2} - F(I_t, \theta_t, \theta_P, 0.5)}{\frac{F(I_t,\theta_t,\theta_P,1-u)-F(I_t,\theta_t,\theta_P,u)}{2}}\right], \tag{2.16}$$

where $u \in (0, 0.5)$ and I set $u = u_1$ the first quantile grid point in the estimation. The skewness measure finds the mid point between an upper $(1-u\ )$ and lower quantile (u) and normalize its distance to median. By construction the measure stays in between $(-1, 1)$.

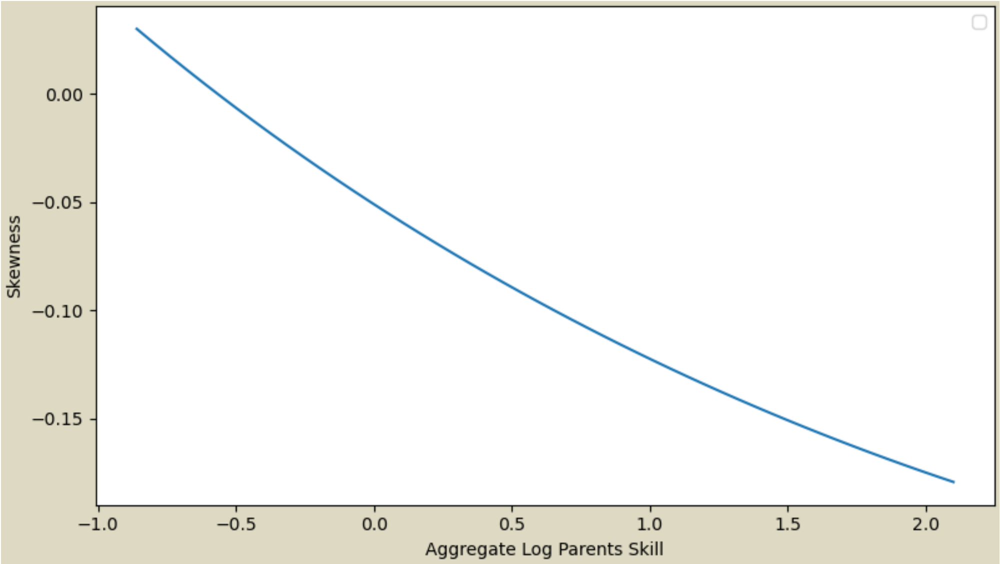


Figure 4: Skewness of childhood skill formation function conditional on parents' skill level.

It is more negatively skewed for more educated parents, i.e. children of more educated parents are more subject to negative risk while the risk is more symmetric of children of low educated parents.

$$\frac{\partial Skewness_t(\theta_P, u)}{\partial \theta_P} < 0. \tag{2.17}$$

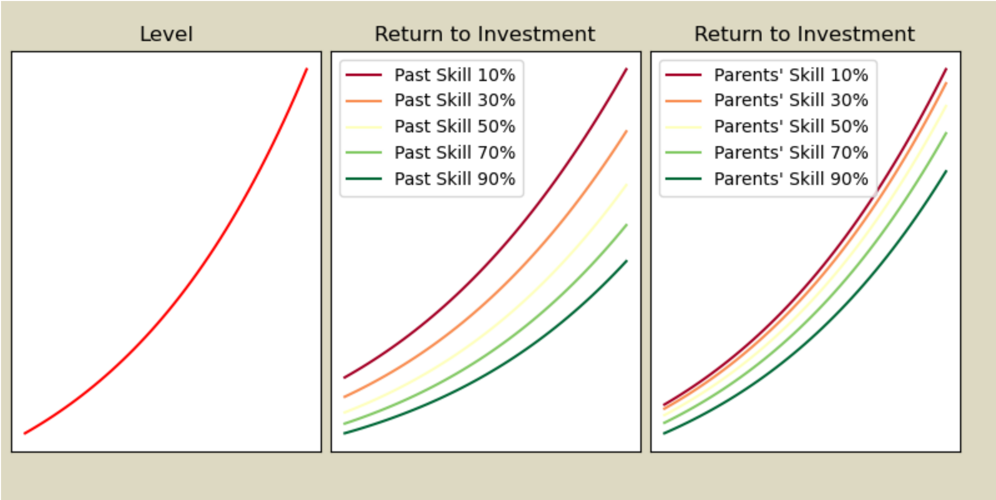
In CES case the skewness is zero by construction because of normality assumption and the normal distribution is symmetric.

I would like to point out that how CES is restrictive because the substitution parameter governs substitution patterns of all inputs, sign of monotonicity and the concavity/convexity in log scale. Figure 5 have the plot for the same features of CES skill function with different values of elasticity of substitution parameters. Figure 5a repeats the baseline case where the estimation results suggest that the inputs are substitutes. Figure 5b plots the Cobb-Douglas case where the returns in logs are constant for all past skill level or parents' skill level. Lastly, Figure 5c illustrates the complement case. When the parameter of elasticity substitution switched from the region of substitutes to complements the shape of function, the order of returns changes all together.

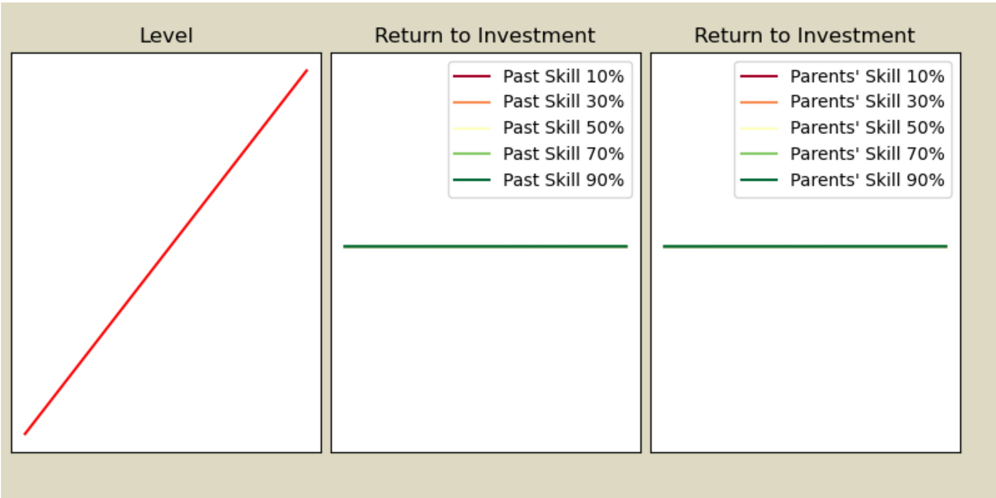
If the elasticity of substitution parameter is higher than zero, substitutes region, i.e. $\phi > 0$,

$$\frac{\partial^2 F_{CES}(I_t, \theta_t, \theta_P)}{\partial \ln I_t^2} > 0, \frac{\partial^2 F_{CES}(I_t, \theta_t, \theta_P)}{\partial \ln I_t \partial \ln \theta_t} < 0, \frac{\partial^2 F_{CES}(I_t, \theta_t, \theta_P)}{\partial \ln I_t \partial \ln \theta_P} < 0$$

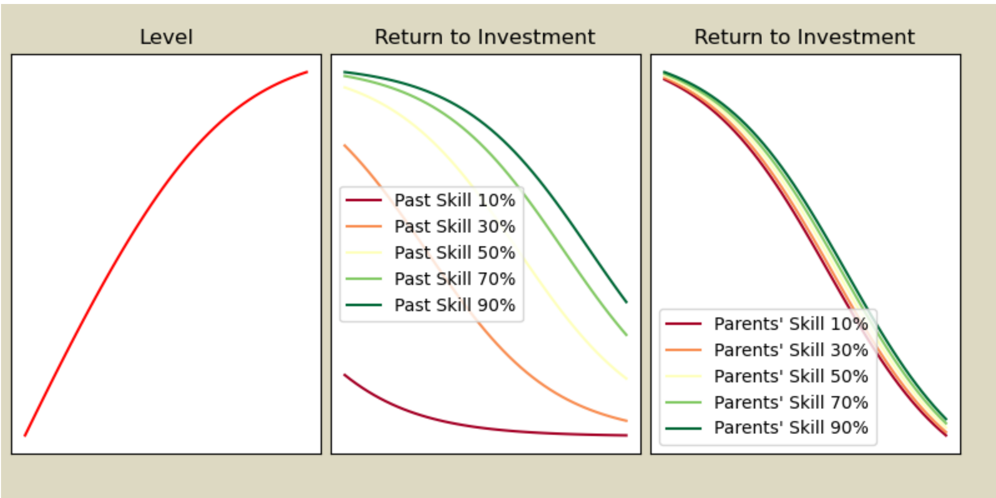
However if $\phi < 0$, then all signs would switch,



(a) Baseline: Substitutes ($\phi = 0.17$)



(b) Cobb-Douglas ($\phi = 0$)



(c) Complements ($\phi = -0.5$)

Figure 5: CES skill function with different substitution parameters.

$$\frac{\partial^2 F_{CES}(I_t, \theta_t, \theta_P)}{\partial \ln I_t^2} < 0, \frac{\partial^2 F_{CES}(I_t, \theta_t, \theta_P)}{\partial \ln I_t \partial \ln \theta_t} > 0, \frac{\partial^2 F_{CES}(I_t, \theta_t, \theta_P)}{\partial \ln I_t \partial \ln \theta_P} > 0$$

There is no reason for these features to be connected and results of flexible case estimation results are in the regions that are different than the ones CES forces them.

3 Intergenerational Mobility

This section is about the implications of childhood skill formation function for intergenerational mobility and inequality. Inequality has been rising substantially and it might be worrying for the intergenerational mobility through childhood skill formation channel.

High income families spends devote more resources for their children both in terms of time and expenditure than low income families and the gap has been getting wider over time (Corak, 2013). As the inequality rises, the resources of high income families are able to invest relatively more and this can increase the chance of their children to be also high income people. On the other hand, low income families lack similar resources and their children might stuck in the low income status. This means lower mobility and less opportunity equality for future generations.

How increasing dispersion in inputs of the skill formation, i.e. parental time and expenditure, would result in the output, i.e. skills and hence income in adulthood, depends on the shape of childhood skill formation function. Therefore how does rising inequality will affect the intergenerational mobility depends on the functional form of the childhood skill formation. In fact, Blanden, Doepke and Stuhler (2022) highlights that the dispersion in test scores seems fixed over time.

In the remaining of the section, I provide an estimate of intergenerational mobility trend to see if the mobility has been declining over time as the inequality rises. In the second subsection, I explain the role of childhood skill formation in the relationship between inequality and the intergenerational mobility with a simple simulation and a theoretical model. Lastly, I comment on the potential implications of the estimation results in the previous section.

3.1 Intergenerational Elasticity of Earnings Trend

I focus on trends in intergenerational elasticity in log household earnings. The ideal way to measure this elasticity would be using life-time earnings of parents and children. However, this requires a panel data covers at least entire two generations. Typically, we can observe earnings at certain ages for each individual with intergenerational connection. Panel Study of Income Dynamics (PSID) started to follow a representative sample of households starting from 1968 and kept following their descendant to this date.

There are several different approaches in the literature to address this problem. Lee and Solon (2009) uses income at different ages for each individual as an observation and controls for a polynomial of both children's and parents' age at the time of measurement to account for the life cycle bias. They use PSID and find no trend in intergenerational elasticity of income.

Justman and Krush (2013); Justman, Krush and Millo (2017) take a two step approach. First, predict the life-time income of fathers and sons and using predicted income they estimate international elasticity of income for different cohorts. They conclude that there is an upward trend in the intergenerational elasticity of income alongside the rise in the inequality.

The last approach uses average income of children and parents on certain ages and compare elasticity for different cohorts. Chetty et al. (2014) measure parents' income average of ages when children were between 15-19 years old and their adult income is measured around age 30. They conclude that there is no trend for cohorts born years between 1970-1985. In a recent paper, Davis and Mazumder (2020) also follows a similar approach and compare earlier cohorts who were born in 1950 and 1960 using NLSY dataset and they find increase in the elasticity.

I follow the last approach and measure both parents' and children's household earnings around age 40 in PSID. In particular, I run the following regression,

$$y_{ic}^{child} = \alpha + \beta y_{ic}^{parent} + \gamma_c y_{ic}^{parent} + \epsilon_{ic}, \tag{3.1}$$

where y_{ic}^{child} is log household earnings of child i from cohort group c, and y_{ic}^{parent} is household earnings for parents of child i from cohort group c and γ_c is the intergenerational elasticity of earnings allowed to differ in cohorts. I measure the household income by taking the mean of income in three years around age 40. PSID is annual before 1997 and biennial afterwards. I use ages 39, 40, 41 for the annual part and the closest ages available to age 40 for the biennial part, e.g. 38, 40 ,42 or 39, 41, 43.

I use only biological parents who lived with their children in their childhood (until age 18) at least one year in observable years in the data. If the head of house changes during childhood, I assign whoever was the head longer as the head of household between father and mother. This only matters whose age, father or mother, is to used for calculation of average household earnings.

This approach requires to observe both patents and children at ages around age 40. I grouped cohorts in ten years groups and the coefficient for cohort born in between 1950 and 1960 is omitted.

Table 1 provides results. In column (1), earnings measured around age 40, surpris-

Table 1: Intergenerational Elasticity of Earnings for Ten Tears Cohort Groups

	Dependent variable:			
	Children's Log Earnings			
	Age 40	(Drop %1)	Age 30	(Drop %1)
	(1)	(2)	(3)	(4)
Parents' Log Earnings	0.500***	0.507***	0.264**	0.612*
	(0.079)	(0.078)	(0.126)	(0.363)
	p = 0.000	p = 0.000	p = 0.037	p = 0.093
Parents' Log Earnings x (61-70)	−0.026	−0.010	0.037	−0.230
	(0.094)	(0.097)	(0.143)	(0.369)
	p = 0.784	p = 0.918	p = 0.798	p = 0.534
Parents' Log Earnings x (71-80)	−0.213**	−0.111	0.146	−0.169
	(0.101)	(0.104)	(0.137)	(0.367)
	p = 0.036	p = 0.286	p = 0.285	p = 0.645
Parents' Log Earnings x (81-90)			0.082	−0.222
			(0.136)	(0.367)
			p = 0.546	p = 0.545
Observations	1,416	1,388	1,707	1,672
Note:		*p<0.1; **p<0.05; ***p<0.01		

ingly the elasticity drops for 1961-1970 cohort almost by half and it is significant at 5% level. In the column (2), I drop the observations in top and bottom top 1% of parents' household earnings. The coefficient for the cohort 1961-1970 drops by half and loses its significance. It is hard to say if this surprising result is driven by a fundamental change on social mobility on tails or driven by a few outliers by chance. Columns (3) and (4) repeat the same exercise but uses average earnings around age 30. The cohort group dummies are not significant but the signs are different between two specification but with very large p-values.

I found no evidence for increasing trend in intergenerational elasticity of earnings, if anything there is decrease for cohort born after 1970 driven by the top and bottom tails of earnings distribution. This result is in line with many papers in the literature which alse estimates the trends in intergenerational mobility using different datasets and methodologies (Chetty et al., 2014; Lee and Solon, 2009; Song et al., 2020). In the next subsection, I will claim that it is possible to explain a flat mobility trend despite rising inequality with the features of the childhood skill formation function.

3.2 Role of Childhood Skill Formation

Figure 6 illustrates how intergenerational correlations in earnings are formed on a very stylized way. There is a distribution of parents' earnings and in the family they decide how much to invest for their children. The investment is proportional to parents' earnings. The childhood skill formation function takes this inputs and translates to skills. In the labor market children are paid proportional to their skill level. We have scatter plot on the right side of graphs showing the intergenerational correlation in earnings.

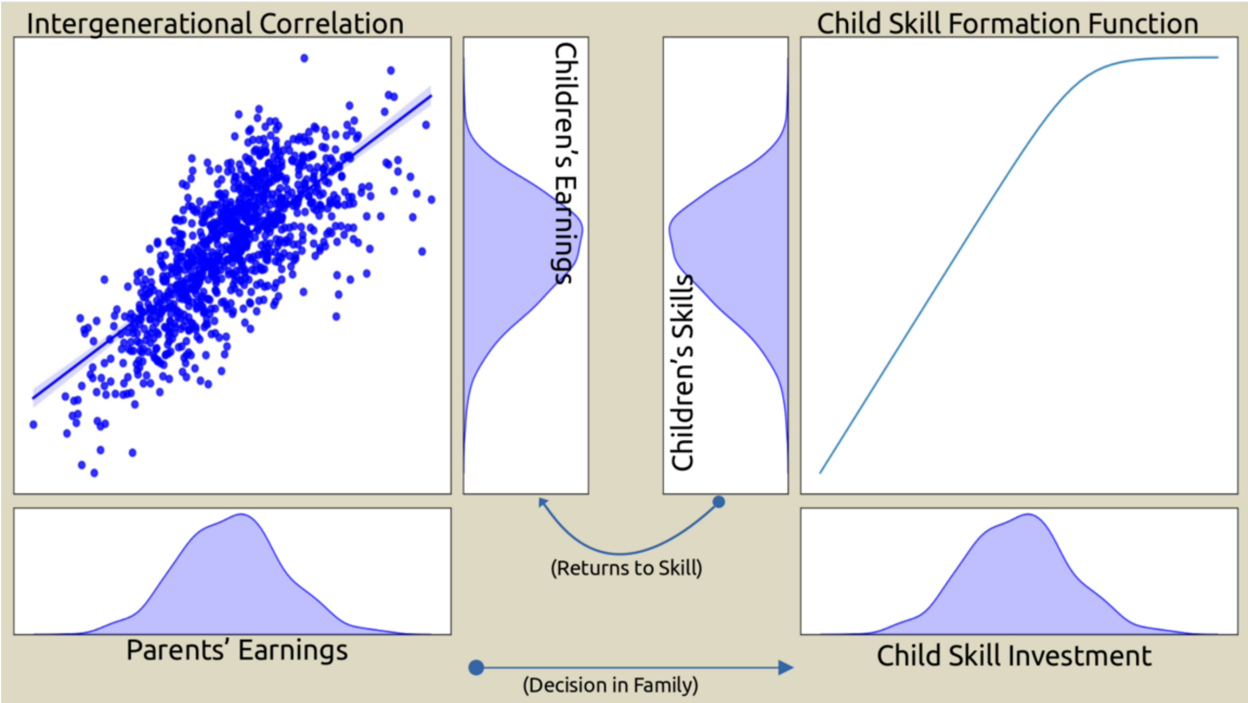
I choose the childhood skill function such the return is really high and for low and middle levels of investment but after some level of investment the returns decrease very fast and the function becomes flat.

The thought experiment is assume that the returns to skills in the labor market is higher because of some exogenous technological change. Figure 6b illustrates this thought experiment. Parents care their children and increase in return to skill would make them to invest more for childhood skills. Since the returns are small for the high levels of investment, the parents who are already investing more will benefit very little from the increase in the skill investment. However the children of the low income parents will enjoy high returns and their skills will increase more. This is going to be reflected in the earnings distribution and the intergenerational correlation would be smaller than before. On the other hand the higher returns to skills would create higher dispersion in earnings of children. Therefore, in this stylized example, we can get a better mobility even if the inequality rises.

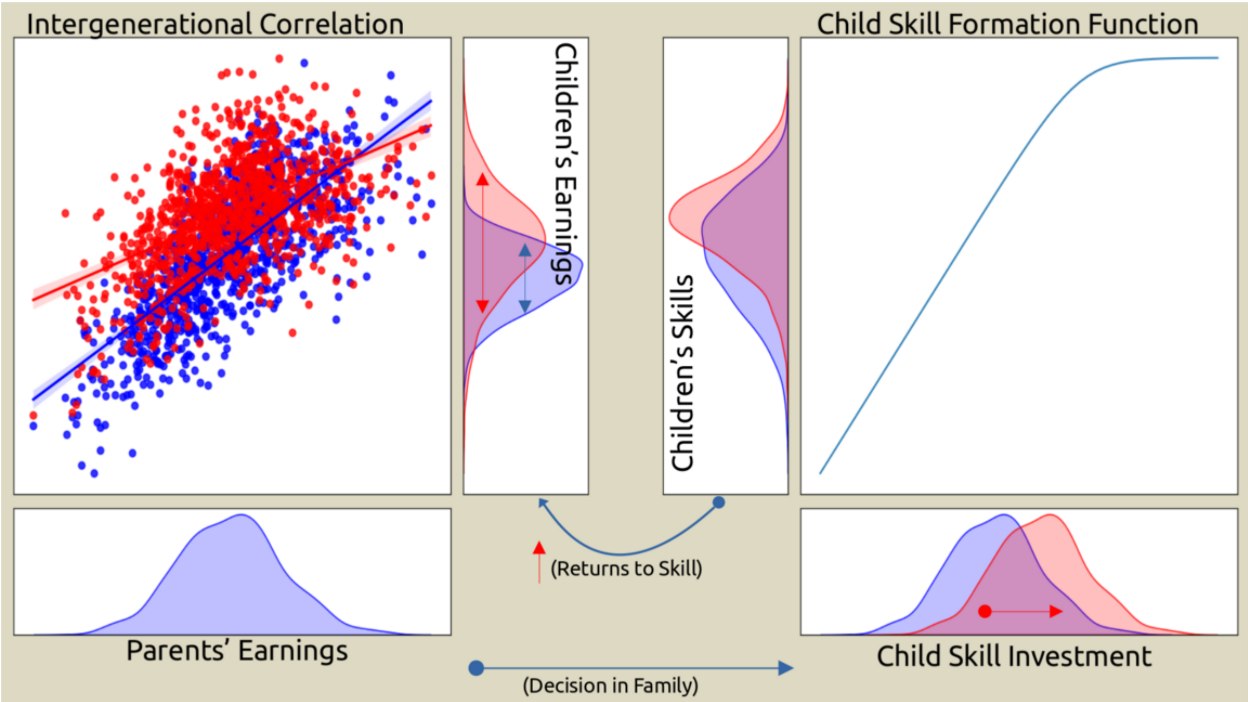
3.2.1 A Simple Model of Intergenerational Mobility

This section provides a simple two period model to make the point that the shape of childhood skill function matters for the implications of higher inequality for intergenerational elasticity of earnings in a more formal way.

I use a simple two period model with general childhood skill formation function similar to the one in Becker et al. (2018) with Cobb-Douglas assumption. I keep both



(a) Baseline



(b) Effect of rise in returns to skill

Figure 6: Illustration of how skill formation function determines intergenerational mobility.

childhood skill production function and return to skills function unspecified to highlight the effect of functional forms and I do not include any uncertainty in the model to focus of the effect of functional forms.

There are two periods: childhood and adulthood. There two agents the parent and the child. The parent is alturistic and care about the child and has two ways contribute their future, bequest and skill investment. The parent maximizes the following utility function,

$$V(y_p) = \max_{c,b_c,I} u(c) + \delta y_c, \tag{3.2}$$

$$\text{s.t.} \quad c + \frac{b_c}{R} + I = y_p, \quad b_c \geq 0, \tag{3.3}$$

where c consumption, b_c bequest for the child, R intergenerational return for savings, I skill investment for the child. y_p and y_c are income of the parent and the child which is sum of earnings and bequest,

$$y_j = E_j + b_j, \quad \text{for} \quad j \in \{p, c\}, \tag{3.4}$$

where E_j is earnings. I assume the parent only case about level of the child's future income for simplicity, all results follows if I assume that they care the utility.

Earnings are a function of skill level, $E_j = g(\theta_j)$ where θ_j is skills for the parent or the child, i.e. $j \in \{p, c\}$. Lastly, the skills are produces by a production function depending on the investment and parent's skills,

$$\theta_c = F(I, \theta_p). \tag{3.5}$$

The solution for an unconstrained parent is where the return of two channels, skill investment and bequest are equal,

$$\frac{\partial y_c}{\partial I} = R \quad \Longrightarrow \quad g'(\theta_c) F_I(I^*, \theta_p) = R, \tag{3.6}$$

and for the constrained parent the return of skill investment should be equal to the marginal utility becasue they cannot borrow against their child's earnings,

$$\delta \frac{\partial y_c}{\partial I} = u'(c) \Longrightarrow g'(I_c) F_I(I^*, \theta_p) = \delta^{-1} u'(c). \tag{3.7}$$

By taking derivative of solution equations 3.6 and 3.7, we can derive expressions for intergenerational elasticity of earnings. In the unconstrained case, it is given by,

$$\frac{\partial \log E_c}{\partial \log E_p} = \frac{\frac{g'(\theta_c)}{g(\theta_c)}}{\frac{g'(\theta_p)}{g(\theta_p)}} \left[-F_I \frac{\frac{g''(\theta_c)}{g'(\theta_c)} F_\theta + \frac{F_{I\theta}}{F_I}}{\frac{g''(\theta_c)}{g'(\theta_c)} F_I + \frac{F_{I\theta}}{F_I}} + F_\theta \right] \tag{3.8}$$

Therefore, the elasticities depend on curvatures and cross derivatives of functions, e.g. $\frac{g''}{g'}$, $\frac{F_{yy}}{F_y}$ and $\frac{F_{yH}}{F_y}$. This shows the importance of estimating childhood skill function in a flexible way so that these features are not restricted like in the case of CES.

3.3 Implications of Estimation Results

In this section, I comment on the potential implications of my estimation results for the flexible childhood skill formation function using the framework developed in the previous subsections. I claim that the features of flexible skill formation function can create mechanisms for a flat mobility in a rising inequality environment.

The strong concavity and flat region at the top of flexible skill function illustrated in the Figure 1, can lead to a constant or higher mobility in a rising inequality environment. The mechanism is the same as it is illustrated in the Figure 6, the high income families are already investing a lot for their children but because of strong concavity they cannot increase children's skill. However this is not true for low income families.

The return to investment decreases faster for more educated parents, as illustrated in the Figure 3. This means that the more educated parents would hit the flat region of the skill formation function quicker than low educated parents. Similar to previous argument, this will limit how much the additional skill can be obtained by children of high educated parents. On the other hands, children of low educated parents may be able to catch up thanks to high and slowly decreasing returns.

The estimation results for the flexible case shows that the current level of skill and the investment are substitutes, hence the investments is more productive for low skilled children as Figure 2 shows. This can create an additional catch up effect for disadvantaged children can improve the mobility when the parents would increase the skill investment even in small amount.

Lastly, the uncertainty around the childhood skill function has non normal features, in particular there is more negative skewness for more educated parents and their children are more subject to negative risk. The uncertainty is more symmetric for the children of low educated parents. This asymmetry can contribute to mean reversion because the children of top part of income distribution would be more likely to end up relatively lower levels of skills and this will improve over all mobility.

4 Conclusion

Human capital production starts in the family just after birth and continues throughout the childhood. Understanding how the childhood skills are produced is crucial to learn about the roots of inequalities and to improve the lives of future generations through better policy and better parenting.

I provide a flexible estimation of childhood skill formation function without making restrictive functional or distributional form assumptions building on the identification results in the literature. My results show that childhood formation function has features that cannot be captured by restrictive functional form assumptions such as CES. For example, I found that the investment is more productive if the child currently has low levels of skill, in contrast to the current results in the literature. This is an optimistic result because it suggests that any missing investment at earlier ages can be substituted at later ages. Also, results show that more educated parents are more productive but return is decreasing at a faster rate compared to the low educated parents.

CES functional form cannot capture these features all together because a single parameter, elasticity of substitution, governs substitution patterns for all inputs and shape of skill production function. There is no reason for these features of the skill formation function to be connected. My flexible estimation shows that this is indeed the case.

I discuss the curial role of skill formation in intergenerational mobility. Skill formation in the family is one of the main channels on transmission of inequality across generations.

The rising inequality can be worrying because it may lead lower mobility and less opportunity equality for the future generations. The intuition is that while high income families are able to provide better education opportunities for their children, low income families may lack enough resources and this gap increases as the inequality increases. However there is good news in the data, the intergenerational earnings mobility seems constant despite substantially rising inequality. I show that the childhood skill formation function can be an explanation for this trends because it possible to get a flat mobility trend in a rising inequality environment if the skill formation function has certain features.

I claim that my flexible children skill formation function has some features can be an explanation for the flat mobility trend given rising inequality. In particular, the investment is more productive for disadvantaged children and they might be able to catch up children of high income families. Also high educated families reach the flat part of the skill function more quickly than low educated parents. Hence the children of low educated parents can enjoy higher returns over wide range of investment. Lastly, the more negative skewness for high educated parent can create a mean reversion effect and improve mobility.

My results have important policy implications. First, there might be no need to panic for the possibility of fading American dream because rising inequality may not lead worse intergenerational mobility at least through the childhood skill formation channel. However, it is known that the intergenerational mobility is already low in the US compared to other developed countries. To improve that an optimal policy should focus on the disadvantaged children even if at later ages in order to maximize the effect in terms of more skills. Also, increasing productivity of low educated parents through parental education can be good strategy.

References

Agostinelli, Francesco, and Matthew Wiswall. 2021. "Estimating the Technology of Children's Skill Formation."

Arellano, Manuel, and Stéphane Bonhomme. 2016. "Nonlinear Panel Data Estimation via Quantile Regressions." *The Econometrics Journal*, 19(3): C61–C94.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica.*

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

Becker, Gary S., Scott Duke Kominers, Kevin M. Murphy, and Jörg L. Spenkuch. 2018. "A Theory of Intergenerational Mobility." *Journal of Political Economy*, 126(S1): S7–S25.

Blanden, Jo, Matthias Doepke, and Jan Stuhler. 2022. "Educational Inequality." Working Paper.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review*, 104(5): 141–147.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives*, 27(3): 79–102.

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

Davis, Jonathan, and Bhashkar Mazumder. 2020. "The Decline in Intergenerational Mobility After 1980." *Working Paper.*

Justman, Moshe, and Anna Krush. 2013. "Less Equal and Less Mobile: Evidence of a Decline in Intergenerational Income Mobility in the United States." Social Science Research Network SSRN Scholarly Paper ID 2370189, Rochester, NY.

Justman, Moshe, Anna Krush, and Hadas Millo. 2017. "The intergenerational elasticity of income in the United States is rising in tandem with income inequality and returns to schooling." Working Paper. Available from http://www. ecineq. org/ecineq_nyc17 . . . .

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *The Review of Economics and Statistics*, 91(4): 766–772.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. 2020. "Long-term decline in intergenerational mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences*, 117(1): 251–258.