Interpreting Trends in Intergenerational Mobility

Martin Nybom[*], Jan Stuhler[‡]

December 21, 2023

Abstract

Studying a dynamic model of intergenerational transmission, we show that past events affect contemporaneous trends in intergenerational mobility. Structural changes may generate long-lasting mobility trends that can be non-monotonic, and declining mobility may reflect past gains rather than a recent deterioration of equality of opportunity. We provide two applications. We first show that changes in the parent generation have partially offset the effect of rising skill premia on income mobility in the US. We then show that a Swedish school reform reduced the transmission of inequalities in the directly affected generation, but increased their persistence in the next.

[*]IFAU Uppsala, SOFI and Uppsala Center for Labor Studies (martin.nybom@ifau.uu.se)

[†]Universidad Carlos III de Madrid, SOFI, HCEO, and IZA (jan.stuhler@.uc3m.es)

[‡]Financial support from the Swedish Council of Working Life (FAS), the Jan Wallander and Tom Hedelius Foundation, the German National Academic Foundation, and the Centre for Research and Analysis of Migration is gratefully acknowledged. We especially thank Anders Björklund, Christian Dustmann, Markus Jäntti and Uta Schönberg for comments and advice. We also received helpful input from four anonymous referees, Robert Erikson, David Green, Raquel Fernández, Raffaella Giacomini, Ricard Grebol, Stephen Jenkins, Mikael Lindahl, Matthew Lindquist, Steve Machin, Magne Mogstad, Lars Nesheim, Marieke Schnabel, Gary Solon, and seminar participants at University College London, RES at Royal Holloway, the Workshop on Intergenerational Mobility at the University of Copenhagen, the Swedish Institute for Social Research, EALE at the University of Torino, the Universita Cattolica in Milan, the AEA meetings in Boston, and the Federal Reserve Bank of Chicago. We are also grateful to Helena Holmlund for kindly sharing her programs for coding of the reform indicator used in our empirical application. The paper was edited by James J. Heckman.

Introduction

The evolution of economic inequality over time is a fundamental topic in the social sciences and public debate. Two central dimensions of interest are the extent of cross-sectional inequality between individuals and its persistence across generations, as advantages are transmitted from parents to their children. Both have important implications for individual welfare and the functioning of political and economic systems (see Erikson and Goldthorpe, 1992; Bénabou and Ok, 2001). The rise in income inequality starting from around 1980 in developed countries is well documented (Autor and Katz, 1999; Atkinson et al., 2011), but less is known about trends in *intergenerational mobility* (see Solon, 1999, Black and Devereux, 2011, and Mogstad and Torsvik, 2021 for reviews). Yet, we do know that income mobility differs substantially across countries, and the observation that those differences appear negatively correlated with cross-sectional inequality has received much attention (e.g. Corak, 2013). A central theme in the recent literature is thus if inequality has not only increased, but also become more persistent across generations.[1]

But how should trends in mobility be *interpreted* – do they reflect changes in the effectiveness of current policies and institutions in promoting equal opportunities? Our main contribution is to provide a dynamic perspective to this question. We show that contemporaneous shifts in income mobility can be caused by events in a more distant past, as structural changes generate transitional dynamics in mobility over *multiple* generations. Such dynamic responses are of particular importance in the study of intergenerational persistence, since even a single transmission step – one generation – corresponds to a long time period.

The interpretation of mobility trends therefore benefits from a dynamic perspective, but existing theoretical work focuses instead on the relation between transmission mechanisms and the *steady-state* level of mobility. In contrast, we examine the *dynamic* implications of a simultaneous equations model of intergenerational transmission (e.g., Conlisk, 1974a). Motivated by the observation that earnings are influenced by multiple dimensions of skill (Heckman, 1995), we deviate from previous work also by allowing income to depend on a vector of skills rather than a single factor.

We first show that the level of intergenerational mobility depends not only on contemporaneous transmission mechanisms, but also on the joint distribution of income and skills in the parent generation – and thus on past mechanisms. This result has a number of implications. First, a one-time policy or institutional change can generate long-lasting mobility trends. The

[1] In countries such as the US it is now frequently argued that the combination of rising inequality and low mobility threatens social cohesion and the notion of "American exceptionalism". Exemplary articles are *"Moving Up: Challenges to the American Dream"* in The *Wall Street Journal* (May 2005), *"The Mobility Myth"* in *The New Republic* (Feb. 2012), "The American Dream of Upward Mobility is Broken" in The Guardian (March 2021) or the *"Great Divide"* series on nytimes.com. Alan Krueger, former Chairman of the Council of Economic Advisers, warned that mobility should be expected to decline further as of the recent rise in income inequality (speech at the Center for American Progress, January 12th, 2012).

resulting shifts in mobility are not necessarily largest in the first affected generation, but can amplify in magnitude over later generations. As a consequence, contemporaneous shifts in mobility might stem not from recent structural changes but events in the more distant past. We focus on differences over time, but the argument extends: mobility differences across countries, or across groups within countries, may reflect the consequences of past instead of current policies or institutions.

Second, we find that a broad class of structural changes cause *non-monotonic transitions* between steady states: the response in mobility at some point switches sign, and mobility in the first affected generation and in steady state may shift in different directions. The dynamics in cross-sectional inequality may contribute to such non-monotonic transitions, and affect different mobility measures differently. The initial mobility response may thus be a poor indicator of the long-run consequences of a structural change, both qualitatively and quantitatively.

In particular, some changes in the economic environment alter the prospects of some families relative to others, generating transitional mobility. For example, a shift towards a more meritocratic society – a rise in the importance of own skill relative to parental status – is to the advantage of talented children from poor families. But while mobility increases in the first affected generation, it is bound to decline again in subsequent generations, if the more highly rewarded skills of the newly rich are passed on to their children. Even structural changes that are mobility-enhancing in the long run can therefore cause negative trends over some generations. Similar transitional mobility gains can occur in response to changing skill returns in a model with multiple skills.

Our main analysis relies on a generation-specific framework, abstracting from the fact that a theoretical generation consists of many cohorts which can be differentially affected both in current and previous generations by shifts in the transmission system. Empirical research instead studies mobility trends across calendar years or cohorts, with a within-family definition of generations. While we note this limitation of our analysis, we highlight the more gradual responses of cohorts both theoretically and in our empirical applications.

We illustrate our main arguments in two applications. First, we revisit the evidence on mobility trends in the US, and discuss their interaction with changes in income inequality and skill premia. Using data from the PSID, we demonstrate that mobility trends over recent birth cohorts also reflect important changes in the *parent* generation: While rising skill returns may have depressed income mobility, such effect was (at least partly) counteracted by the mobility-enhancing effects of decreasing educational inequality among the parents of those birth cohorts. Indeed, a simple quantification suggests that had the parental schooling distribution stayed constant, the intergenerational elasticity of income could have risen by 20-25 percent, all else equal.

Finally, we examine a Swedish compulsory school reform to provide *causal* evidence for our key theoretical argument – that shocks in the economic environment in the parent generation can still affect mobility trends in the next generation, and that those transitional dynamics can be pronounced and complex. Exploiting administrative data covering three generations, we first show that by reducing the transmission of income and educational inequalities the reform increased mobility in the first generation (as in Holmlund, 2008). But the same reform then *decreased* mobility in the next generation. Shifts in the variance of education and income are central to understand this pattern.

Our work contributes to both the theoretical and empirical literature on intergenerational mobility. Most theoretical studies examine only the *steady-state* relationship between transmission mechanisms and mobility. An early exception is Atkinson and Jenkins (1984). While they show that failure of the steady-state assumption impedes the identification of structural parameters, we instead consider the dynamic effects of changes in such parameters on mobility. There are a few papers using utility-maximizing frameworks to analyze the dynamics of intergenerational transmission. For example, Solon (2004) examines how structural changes affect mobility in the *first* affected generation, and Davies et al. (2005) note that the observation of mobility trends may help to distinguish between alternative causes of rising cross-sectional inequality. Our paper also relates to Becker and Tomes (1979) and the related literature on individual income processes. While they analyze the dynamics of *individual outcomes* within families, we study how such processes relate to the dynamics of aggregate measures of intergenerational mobility.

The empirical literature is broad. Many studies examine occupational and class mobility over time (see Breen, 2004, Hauser, 2010, Long and Ferrie, 2013, and Modalsli, 2017). A more recent literature studies mobility trends in *income* or *educational attainment*, how those trends differ between groups, or how they are affected by institutional aspects. Such studies face substantial data requirements, and the evidence is still debated.[2] A central concern in many of these papers and in public debate is that mobility may have declined in conjunction with the recent rise in income inequality. Various potential causes – such as educational expansion, rising returns to education, or immigration – have been proposed (e.g., Levine and Mazumder, 2007, and further articles in the same issue). Common to most is that they relate trends to *recent* events that directly affected the respective cohorts. We argue that their cause might also lie in the more distant past.

The paper proceeds as follows. In the next section we present our model of intergenerational transmission. We derive current and steady-state mobility levels in terms of its struc-

[2]For example, Hertz (2007), Lee and Solon (2009) and Chetty et al. (2014a) find no major trend in income mobility in the second half of the 20th century in the US, while Davis and Mazumder (2020) show that mobility has fallen over the 1950s cohorts (corroborating related evidence by Levine and Mazumder, 2007, and Aaronson and Mazumder, 2008). A decline has also been found for the UK (Nicoletti and Ermisch, 2007; Blanden et al., 2013), while mobility was stable or even increased in the Nordic countries (Pekkala and Lucas, 2007; Björklund et al., 2009; Pekkarinen et al., 2017; Markussen and Røed, 2020).

tural parameters and summarize our main propositions on transitional dynamics in Section 2. Section 3 presents a set of simple cases to illustrate our main results. We study the interrelation between cross-sectional inequality and mobility in Section 4, which also includes a discussion of mobility trends in the US. Section 5 presents our Swedish application, and Section 6 concludes.

1 A Model of Intergenerational Transmission

In this section we describe a simple dynamic model of intergenerational transmission based on a system of linear difference equations. We summarize the dynamic implications of the model in Section 2 before discussing specific cases and applications that illustrate our main arguments.

1.1 Measuring Intergenerational Mobility

In our main analysis we focus on the intergenerational elasticity of income (IGE), the most popular descriptive measure in the economic literature. Consider a simplified one-parent one-offspring family structure, with $y_{i,t}$ denoting the log lifetime income of the offspring in generation t of family i and $y_{i,t-1}$ the log lifetime income of the parent. For ease of exposition, we here emphasize generations and abstract from potential cohort differences within a generation. The IGE is given by the slope coefficient in the linear regression

$$y_{it} = \alpha_t + \beta_t y_{i,t-1} + \epsilon_{it}. \tag{1}$$

The IGE β_t captures a *statistical* relationship so the error ϵ_{it} is uncorrelated with the regressor by construction. Under stationarity of the variance of y_{it} it equals the intergenerational correlation, which adjusts the IGE for changes in cross-sectional inequality. The IGE captures to what degree percentage differences in parental income on average transmit to the next generation, with a low IGE indicating high mobility. We refer to mobility (or "persistence") primarily in terms of the IGE, but also illustrate how our core arguments extend to alternative measures, such as the intergenerational and sibling correlations.

1.2 Main Model

As motivated below, we model intergenerational transmission as a system of stochastic linear difference equations, in the tradition of the simultaneous equation approach developed by Conlisk (1969; 1974a) and Atkinson and Jenkins (1984). While Becker and Tomes (1979) and related models (e.g., Solon, 2004) explicitly consider the roles of preferences and constraints, we show in Appendix A.1 that the pathways represented by these equations can be derived from such utility-maximization frameworks (see also Goldberger, 1989). The structural equations of our model are

$$y_{it} = \gamma_{y,t}\, y_{it-1} + \boldsymbol{\delta}_t'\, \boldsymbol{h}_{it} + u_{y,it} \quad (2)$$

$$\boldsymbol{h}_{it} = \boldsymbol{\gamma}_{h,t}\, y_{it-1} + \boldsymbol{\Theta}_t\, \boldsymbol{e}_{it} + \boldsymbol{u}_{h,it} \quad (3)$$

$$\boldsymbol{e}_{it} = \boldsymbol{\Lambda}_t\, \boldsymbol{e}_{it-1} + \boldsymbol{\Phi}_t \boldsymbol{v}_{it}. \quad (4)$$

From equation (2), the income y_{it} of an individual of family i in generation t is determined by parental income y_{it-1}, own human capital $\boldsymbol{h}_{it}$, and market luck $u_{y,it}$. The parameter $\gamma_{y,t}$ captures a direct effect of parental income that is independent of offspring productivity. We model human capital as a $J \times 1$ vector $\boldsymbol{h}_{it}$, reflecting distinct skill dimensions such as formal schooling, health, and cognitive and non-cognitive skills, which are valued on the labor market according to the $J \times 1$ price vector $\boldsymbol{\delta}_t$. The random shock term $u_{y,it}$ captures factors that do not relate to parental background. For our analysis it makes no difference if these are interpreted as pure market luck or as the impact of other characteristics that are not transmitted within families.

From equation (3), human capital is determined by parental income y_{it-1}, own *endowments* $\boldsymbol{e}_{it}$, and chance $\boldsymbol{u}_{h,it}$. A role for parental income through the $J \times 1$ vector $\boldsymbol{\gamma}_{h,t}$ may stem from parental investment into offspring human capital, and the elements of $\boldsymbol{\gamma}_{h,t}$ may differ if parental investments are more consequential for some types of human capital than others. Parental income may thus affect offspring income directly (through $\gamma_{y,t}$) or indirectly (through $\boldsymbol{\gamma}_{h,t}$).[3] The $J \times K$ matrix $\boldsymbol{\Theta}_t$ governs how endowments such as abilities or preferences, represented by the $K \times 1$ vector $\boldsymbol{e}_{it}$, affect the accumulation of human capital. Endowments are partly inherited from parental endowments $\boldsymbol{e}_{it-1}$ through the $K \times K$ heritability matrix $\boldsymbol{\Lambda}_t$, and partly due to chance $\boldsymbol{v}_{it}$. We use the term heritability in a broad sense, potentially reflecting both genetic inheritance and family environment. Market luck $u_{y,it}$ and the elements of $\boldsymbol{u}_{h,it}$ and $\boldsymbol{v}_{it}$ are assumed to be uncorrelated with each other and past values of all variables.

For convenience we omit the individual subscript i and make a few simplifying assumptions. As we focus on relative mobility, assume that all variables are measured as trendless indices with constant mean zero (as in Conlisk, 1974a). To avoid case distinctions, assume further that those indices measure positive characteristics ($\gamma_{y,t}$ and elements of $\boldsymbol{\gamma}_{h,t}$ and $\boldsymbol{\delta}_t'\boldsymbol{\Theta}_t$ are non-negative) and that parent and offspring endowments are not negatively correlated (elements of $\boldsymbol{\Lambda}_t$ are non-negative), for all t.

[3] The direct effect may arise as of nepotism, statistical discrimination, credit constraints, parental information and networks, or (if total income is considered) returns to bequests. The distinction between a direct and indirect effect may not be sharp in practice; for example, parental credit constraints might affect educational attainment and human capital acquisition of offspring, but might also affect their career choices for a given level of human capital.

Using equation (3) to substitute out $\boldsymbol{h}_{i,t}$ we can simplify the model as

$$y_t = \gamma_t\, y_{t-1} + \boldsymbol{\rho}_t'\, \boldsymbol{e}_t + \sigma_t\, u_t \tag{5}$$

$$\boldsymbol{e}_t = \boldsymbol{\Lambda}_t\, \boldsymbol{e}_{t-1} + \boldsymbol{\Phi}_t\, \boldsymbol{v}_t, \tag{6}$$

where the parameter $\gamma_t = \gamma_{y,t} + \boldsymbol{\delta}_t'\boldsymbol{\gamma}_{h,t}$ aggregates the direct and indirect effects of parental income, the $1 \times K$ vector $\boldsymbol{\rho}_t' = \boldsymbol{\delta}_t'\boldsymbol{\Theta}_t$ captures the returns to inheritable endowments and acquired skills (affected both by the importance of endowments in the accumulation of and the returns to human capital), and $\sigma_t u_t = u_{y,t} + \boldsymbol{\delta}_t'\boldsymbol{u}_{h,t}$ aggregates the luck terms related to income and human capital. As $\boldsymbol{\rho}_t$ captures returns to both endowments and skills we use these terms interchangeably below.[4] Note that our model allows for strong intergenerational persistence in these underlying skills even if the parent-child mobility in income is high, in line with the pattern observed for some Scandinavian countries (Landersø and Heckman, 2017). We normalize the variance of u_t to one in all periods, such that changes in the importance of market luck are captured by σ_t.

Our model has a similar structure as the model in Conlisk (1974a), which in turn is similar to the statistical framework implied by Becker and Tomes' economic model (Goldberger, 1989). But in contrast to the previous literature we allow for income to depend on human capital through a vector of skill dimensions. This addition is central for some of our findings but for some arguments it suffices to consider a simpler scalar model with a single skill. Similarity to the existing literature in other dimensions is advantageous since it suggests that our findings do not arise due to non-standard assumptions. The second deviation from previous work is simply the addition of t-subscripts to all parameters, reflecting our focus on the dynamic response to changes in the transmission framework.

Each parameter is a reduced-form representation of multiple underlying mechanisms, and an underlying change may affect multiple parameters at once. For example, an expansion of public childcare may affect the transmission and supply of skills and in turn their returns on the labor market. A behavioral model would endogenize some of these linkages. However, to trace how a shift in one parameter may lead to subsequent shift in others, while interesting, is not needed to illustrate our main arguments. We therefore only provide examples of such links and assume instead that the economic environment is exogenous.

Definition 1. The economic environment $\boldsymbol{\xi}_t$ consists of the set of transmission mechanisms

[4] We recognize that the multidimensionality and the different layers of the model – with multiple underlying endowments potentially influencing the different types of market-valued human capital – make the concept of returns to human capital rather complex. However, we focus our analysis on the simplified two-equations model in equations (5) and (6), and treat the underlying endowments as the main dimension of analysis, abstracting from the implicit human-capital channels through which endowments affect income. Further, we for simplicity often impose that all off-diagonal elements of $\boldsymbol{\Lambda}_t$ are zero, such that endowments only transmit to the along the same endowment types across generations.

that generation t is subject to, represented by the parameters $\boldsymbol{\xi}_t = \{\gamma_t, \boldsymbol{\rho}_t, \sigma_t, \boldsymbol{\Lambda}_t, \boldsymbol{\Phi}_t\}$. A structural change is a permanent change in any of the features of the environment in generation $t = T$, such that $\boldsymbol{\xi}_{t<T} = \boldsymbol{\xi}_1 \neq \boldsymbol{\xi}_{t\geq T} = \boldsymbol{\xi}_2$.

For simplicity, we assume that the moments of all variables were in steady-state equilibrium before the structural change occurs, and that the system is stable (implicitly restricting the parameter space, see Appendix A.2).[5] We also normalize the variances of y_t and elements of $\boldsymbol{e}_t$ in the initial steady state to one.

2 The Importance of Past Transmission Mechanisms

We express intergenerational mobility as a function of our model to illustrate some central implications. The IGE is derived by plugging equations (5) and (6) into (1), such that

$$\beta_t = \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} = \gamma_t + \frac{\boldsymbol{\rho}_t' \boldsymbol{\Lambda}_t Cov(\boldsymbol{e}_{t-1}, y_{t-1})}{Var(y_{t-1})}. \tag{7}$$

Thus, β_t depends on current transmission mechanisms (parameters γ_t, $\boldsymbol{\rho}_t$ and $\boldsymbol{\Lambda}_t$), but also on the variance and cross-covariance between income and endowments in the parent generation. The intuition is simple: if income and favorable endowments are concentrated in the *same* families, then intergenerational mobility will be low (the IGE will be high). Two populations currently subject to the same transmission mechanisms can therefore still differ in their levels of mobility, since current mobility also depends on the joint distribution of income and endowments in the parent generation.

The cross-covariance between income and endowments in the parent generation is in turn determined by past transmission mechanisms, and thus past values of $\{\gamma_t, \boldsymbol{\rho}_t, \boldsymbol{\Lambda}_t\}$. We can iterate equation (7) backwards to express β_t in terms of parameter values,

$$\begin{aligned}
\beta_t &= \gamma_t + \frac{\boldsymbol{\rho}_t' \boldsymbol{\Lambda}_t \left(\boldsymbol{\Lambda}_{t-1} Cov(\boldsymbol{e}_{t-2}, y_{t-2})\gamma_{t-1} + Var(\boldsymbol{e}_{t-1})\boldsymbol{\rho}_{t-1}\right)}{Var(y_{t-1})} \\
&= \ldots \\
&= \gamma_t + \boldsymbol{\rho}_t' \boldsymbol{\Lambda}_t \boldsymbol{\rho}_{t-1} + \boldsymbol{\rho}_t' \boldsymbol{\Lambda}_t \left(\sum_{r=1}^{\infty} \left(\prod_{s=1}^{r} \gamma_{t-s} \boldsymbol{\Lambda}_{t-s} \right) \boldsymbol{\rho}_{t-r-1} \right),
\end{aligned} \tag{8}$$

where for simplicity we assumed that all off-diagonal elements of $\boldsymbol{\Lambda}_t$ are zero, that the variances remain constant and normalized to $Var(y_t) = Var(e_{j,t}) = 1 \; \forall j, t$, and that the process is infinite.[6] The current level of intergenerational mobility thus depends on current and past transmission mechanisms.

[5] Jenkins (1982) discusses stability conditions for systems of stochastic linear difference equations.

[6] For a finite process, β_t will also depend on the initial condition $Cov(\boldsymbol{e}_0, y_0)$. If cross-sectional inequality

If no structural changes occur, $\boldsymbol{\xi}_t = \boldsymbol{\xi}\ \forall t$, equation (7) implies that

Proposition 1. STEADY STATE. *The steady-state intergenerational elasticity equals*

$$\beta = \gamma + \frac{(1-\gamma^2)\rho^2\lambda\Phi^2}{\rho^2\Phi^2(1+\gamma\lambda) + \sigma^2(1-\lambda^2)(1-\gamma\lambda)} \tag{9}$$

in the scalar model with a single skill, and

$$\beta = \gamma + \frac{(1-\gamma^2)\,\boldsymbol{\rho}'\boldsymbol{\Lambda}\,(\boldsymbol{I}-\gamma\boldsymbol{\Lambda})^{-1}\,(\boldsymbol{I}-\boldsymbol{\Lambda}\boldsymbol{\Lambda}')^{-1}\,\boldsymbol{\Phi}^2\boldsymbol{\rho}}{\boldsymbol{\rho}'\,(\boldsymbol{I}-\boldsymbol{\Lambda}\boldsymbol{\Lambda}')^{-1}\,\boldsymbol{\rho}\boldsymbol{\Phi}^2 + 2\gamma\boldsymbol{\rho}'\boldsymbol{\Lambda}\,(\boldsymbol{I}-\gamma\boldsymbol{\Lambda})^{-1}\,(\boldsymbol{I}-\boldsymbol{\Lambda}\boldsymbol{\Lambda}')^{-1}\,\boldsymbol{\rho}\boldsymbol{\Phi}^2 + \sigma^2}, \tag{10}$$

in the multi-skill model. It decreases in the importance of market luck σ^2, and increases in the effect of parental income γ and the variance of endowment luck Φ^2. It increases in the returns to endowments ρ and the heritability of endowments λ in the single-skill model, but may decrease in returns ρ_k for some skill k in the model with multiple skills.

Proof. See Appendix A.4.1.

□

The steady-state implications of the scalar model are comparable to those from standard models in the literature, such as Becker and Tomes (1986) or Solon (2004). An increase in the returns to parental income or endowments, or in the heritability or variance of endowments increases the IGE, while market luck diminishes it. However, in the model with multiple skills, an increase in the return to one skill has an ambiguous effect on the steady-state IGE and may lower it if the heritability of that skill is low (see Section 4).

2.1 Dynamic Properties of the System

The literature has almost exclusively focused on how changes in structural parameters affect the IGE in steady state, as given by equations (9) or (10). We will instead analyze its *transition path* as determined by equations (7) and (8). Following a structural change of the economic environment, what can we say about the transition path of the IGE towards the new steady state? We first focus on the single-skill model, before illustrating some further implications of the multi-skill model in Sections 3 and 4. Throughout we assume that all variables were in steady state in $t = T-1$ when a structural change occurs in generation T, such that $\boldsymbol{\xi}_{t<T} \neq \boldsymbol{\xi}_{t\geq T}$ (see Definition 1). We use the normalization that $Var(e_t) = Var(y_t) = 1$ for $t < T$, and occasionally consider a constant-variances case in which the variances remain constant for all t. Since we consider one-time, permanent shifts we use notation such

varies over generations, or if $\boldsymbol{\Lambda}_t$ is not diagonal, the derivation of equation (8) requires backward iteration of $Var(y_t)$ and the variance-covariance matrix of $\boldsymbol{e}_t$.

as $\rho_1 = \rho_{t<T}$ and $\rho_2 = \rho_{t\geq T}$ for model parameters. We further use abbreviations such as $\triangle Cov(e_T, y_T) = Cov(e_T, y_T) - Cov(e_{T-1}, y_{T-1})$ for changes in statistical moments. While $\triangle$ generally denotes first-differences, we use $\triangle_\infty$ for the steady-state shift between $T-1$ and the new steady state (see Appendix A.3 for details).[7] We relegate most derivations to Appendix A.4.

From equation (7), it follows that in the aftermath of a structural change the IGE may not immediately shift to its new steady state. Our next proposition characterizes the conditions for such prolonged transition over multiple generations.

Proposition 2. **TRANSITIONAL DYNAMICS**. *Following a permanent structural change in the economic environment $\boldsymbol{\xi}_t$ at $t = T$, the intergenerational elasticity β_t may shift over multiple generations to its new steady state. Specifically:*

(a) A structural change triggers convergence over more than one generation iff $\rho_2 > 0$, $\lambda_2 > 0$, and $\frac{\Delta Cov(e_T, y_T)}{Cov(e_{T-1}, y_{T-1})} \neq \frac{\Delta Var(y_T)}{Var(y_{T-1})}$. This inequality always holds for changes in σ^2 or Φ^2, and holds for other parameter changes except in special cases. Moreover, if either $\gamma_2 > 0$ or $\lambda_2^2 \neq 1 - \Phi_2^2$, convergence is in infinite time.

(b) The convergence steps can increase in absolute size ("amplification"). Amplification in period $T+1$ ($|\Delta\beta_{T+1}| > |\Delta\beta_T|$) always occurs after parameter changes in σ^2 or Φ^2, is possible after a change in ρ or λ, and never occurs after a change in only γ. Amplification in later periods ($|\Delta\beta_{T+k+1}| > |\Delta\beta_{T+k}|$ for some $k \geq 1$) is possible for changes in any parameter.

Proof. See Appendix A.4.2 for derivations and Case 1 in Section 3 for illustrations. □

Proposition 2 has important implications for the interpretation of mobility trends. First, mobility tends to shift over more than one generation towards its new steady state, even if no other changes in the economic environment occur. An observed shift in the IGE today can therefore be due to a one-time structural change that occurred in a previous generation. Indeed, for changes in σ^2 and Φ^2, the IGE shifts only from the second generation onwards (see Table A.1). For changes in other parameters, the convergence process lasts over at least two generations if the IGE reflects the transmission of skills ($\lambda_2 > 0$) and their effect on income ($\rho_2 > 0$), except in knife-edge cases that shift the covariance between income and endowments and the variance of income in generation T by the exact same proportion ($\frac{\Delta Cov(e_T, y_T)}{Cov(e_{T-1}, y_{T-1})} = \frac{\Delta Var(y_T)}{Var(y_{T-1})}$). Second, IGE trends may fail to reflect the impact of a contemporaneous structural change if they are dominated by the ongoing response to another change that occurred in past generations.

The implication of prolonged mobility trends is more than a theoretical curiosity. Even adjustments that fully materialize within two generations can generate long-lasting transitional dynamics over cohorts (see Proposition 5). Moreover, the size of the convergence

[7] For example, $\triangle Var(y_\infty) = Var(y_\infty) - Var(y_{T-1})$ and so on.

steps can increase after the initial generation T. In particular, Proposition 2b considers conditions under which the IGE might shift more strongly in generation $T + 1$ than in generation T. *For example, an increase in the returns to endowments ($\rho_2 > \rho_1$) is likely to satisfy $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ for low values of ρ_1.* In principle, amplification can also occur in later periods ($|\Delta\beta_{T+k+1}| > |\Delta\beta_{T+k}|$ for some $k \geq 1$), but the parameter values that trigger such delayed amplification appear less plausible (see Appendix A.4).

The literature often relates observed shifts in mobility to recent policy changes. However, mobility may fail to respond to an apparent change in the economic environment, or shift in response to previous structural changes (which can also affect consecutive generations of a given family, as illustrated in Section IV of Becker and Tomes, 1979). An important challenge in applications is therefore to determine if mobility trends reflect the response to contemporaneous changes in the economic environment or the ripple effect of structural changes in the past. It follows from equation (7) that the key statistics to distinguish the two are the variances and cross-covariance between income and skills or endowments in the parent generation. We return to this implication below, in the context of US mobility trends in Section 4.

Proposition 2 implies that transitional dynamics can obscure the quantitative effects of structural changes on mobility. Another interesting observation is that structural changes can trigger non-monotonic transitions of the IGE, also complicating the analysis of mobility trends in a *qualitative* sense:

Proposition 3. NON-MONOTONICITY. *Following a permanent structural change in the economic environment $\boldsymbol{\xi}_t$ at $t = T$, the transition path of the elasticity β_t between the old steady state and new steady state $\beta_{t\to\infty} = \beta_\infty$ can be non-monotonic:*

(a) The transition path switches sign in generation T+1 iff $\frac{\Delta Cov(e_T, y_T)}{Cov(e_{T-1}, y_{T-1})} < \frac{\Delta Var(y_T)}{Var(y_{T-1})}$ for an initial shift $\Delta\beta_T > 0$ and the reverse inequality for an initial shift $\Delta\beta_T < 0$. These conditions can hold for changes in γ, ρ or λ, but not for changes in only σ^2 or Φ^2.

(b) The initial shift can be larger than the steady-state shift ("weak non-monotonicity"), such that $|\Delta\beta_T| > |\Delta\beta_\infty|$. For a single parameter change, this condition can hold for a change in γ, ρ or λ, but not for changes in σ^2 or Φ^2.

(c) The initial shift and the steady-state shift can have opposite signs ("strong non-monotonicity"), $sign(\Delta\beta_T) \neq sign(\Delta\beta_\infty)$, for a single parameter change in the multi-skill model or if two parameters shift in the single-skill model.

Proof. See Appendix A.4.3 for derivations and Cases 2 and 3 in Section 3 for illustrations. □

The proposition distinguishes different forms of non-monotonicity. The first two parts consider a weak form of non-monotonicity, in which the mobility trend at some point changes sign, but in which the initial response $\Delta\beta_T = \beta_T - \beta_{T-1}$ still has the same sign as the steady-state response $\Delta\beta_\infty = \beta_\infty - \beta_{T-1}$. Part (a) considers the first two generations after a structural change, while part (b) considers sign changes in later generations along the transition path. Part (a) implies that a structural change can increase the IGE initially ($\Delta\beta_T > 0$), but subsequently decrease it ($\Delta\beta_{T+1} < 0$), or vice versa, if the initial shift in the variance of income, $\Delta Var(y_T)$, is large in relative terms. Non-monotonicity in later generations ($sign(\Delta\beta_{T+k+1}) \neq sign(\Delta\beta_{T+k})$ for some $k \geq 1$) is also possible, but less likely. Part (b) generalizes this result to the cumulative response along the transition path.

Under weak non-monotonicity, mobility trends may be misleading in that over some generations the IGE shifts in one direction while the steady-state shift in fact goes in the other direction. Case 2 in Section 3 provides an illustration. Numerical analyses show that the scenarios outlined by parts (a) and (b) of Proposition 2 are likely for shifts in γ but less common (though possible) for shifts in ρ or λ.

The final part of the proposition distinguishes a strong form of non-monotonicity in which the initial response $\Delta\beta_T$ and the steady-state response $\Delta\beta_\infty$ have opposite signs. The condition for strong non-monotonicity cannot hold for single parameter shifts in the single-skill model, but it can hold if two parameters shift or for a single parameter shift in the multi-skill model. Under strong non-monotonicity, the cumulative effect of all shifts after the first affected generation dominate the initial shift, such that only considering $\Delta\beta_T$ provides a qualitatively false picture of the long-run effect on mobility. Case 4 in Section 4 shows strong non-monotonicity for a single-parameter shift in the multi-skill model and illustrates that the proposition extends to other mobility measures, such as the intergenerational correlation.

In particular, non-monotonic transitions are commonplace for changes in the relative strength of different transmission mechanisms in a multi-skill model that only imply small steady-state shifts in the IGE:

Remark. **TRANSITIONAL MOBILITY GAINS.** In a model with multiple transmission mechanisms, a change in the strength of one mechanism relative to another tends to temporarily increase mobility (relative to its old and new steady-state level). Accordingly, the transition path is non-monotonic if the difference between the old and new steady-state IGE is sufficiently small.

This result is derived formally in Section 3 (Case 6). Intuitively, changes in the economic environment alter the prospects of some families relative to others, such that mobility is particularly high in the generation in which this reshuffling of prospects takes place. For example, when skills are differently distributed across families, then a change in the relative importance of one skill has a stronger effect on some families than others. Specifically, if the return to a particular skill rises, then the income prospects of families in which this skill is comparatively abundant will rise. If this skill was a relatively unimportant determinant of incomes prior to the change, then intergenerational mobility will be high. However, as time elapses the newly rich will pass on their advantages to their children and mobility will return to lower levels. Thus, mobility will tend to be temporarily high in times of changes in the economic environment.

Together, Propositions 2 and 3 have important implications for the interpretation of mobility trends. The effect of a structural change on mobility in the first affected generation may not be representative of its long-term impact, neither quantitatively nor qualitatively.

2.2 Other Mobility Measures and Cohort Dynamics

While we focus on the IGE, our arguments also apply to other measures of the importance of family background, such as intergenerational correlations (e.g.,Hertz et al., 2008), rank correlations (Chetty et al., 2014b), or sibling correlations (Björklund et al., 2009). We further consider how arguments on the transitional dynamics over generations apply to dynamics over cohorts.

Other Mobility Measures. Different measures of intergenerational mobility can exhibit different transitional dynamics, even when their steady-state responses are similar. Comparing the elasticity β with the intergenerational correlation $r_t = Corr(y_t, y_{t-1})$, the result follows trivially from the observation that $r_t{=}\beta_t\sqrt{Var(y_{t-1})/Var(y_t)}$, such that $r_t = \beta_t$ in steady state but $r_t \neq \beta_t$ when $Var(y_t) \neq Var(y_{t-1})$ along the transition path. Moreover, the initial responses can have opposite signs:

Proposition 4. **THE INTERGENERATIONAL CORRELATION VS THE ELASTICITY.** *Following a permanent structural change in the economic environment $\boldsymbol{\xi}_t$ at $t = T$, the initial responses of the intergenerational elasticity β_t and correlation r_t differ if $\Delta Var(y_T) \neq 0$ and can have different signs if $\Delta Var(y_T)$ is sufficiently large. Specifically, changes in market luck σ or endowment luck Φ always yield $\Delta r_T \neq 0$ and $\Delta\beta_T = 0$, while changes in the direct effect of parental income γ yield $\Delta r_T \neq \Delta\beta_T$ but $sign(\Delta r_T) = sign(\Delta\beta_T)$. For changes in returns ρ or heritability λ, $sign(\Delta r_T) \neq sign(\Delta\beta_T)$ is possible, depending on parameter values.*

Proof. See Appendix A.4.4. □

The intergenerational correlation tends to respond more immediately to structural changes because it depends on the variance of income in the current period. In particular, structural changes with a large influence on the variance but a small influence on intergenerational transmission in the first affected generation will have qualitatively different effects on the correlation and the IGE. For example, if the returns to a weakly inheritable skill increases, then the correlation decreases, while the IGE tends to increase marginally. We illustrate these results in Section 4. Opposing patterns may also occur if we allow two parameters to change simultaneously. For example, if the effect of market luck and skill returns on income increase at the same time, the initial responses of the correlation and the IGE can have opposite signs. Similar considerations hold for other mobility measures. Sibling correlations depend less directly on conditions in the parent generation and thus respond even more rapidly to changes in the economic environment. For an illustration, see Appendix A.5. It is more tedious to analyze the dynamic response of the rank correlation, as it depends on additional distributional assumptions. However, in simulations based on normal distributions, its dynamic pattern closely tracks the dynamics of the intergenerational correlation.

Non-steady State Dynamics over Cohorts. While the theoretical literature models transmission between *generations*, empirical studies estimate mobility trends over *cohorts*. This distinction is not relevant for steady-state analysis, and has thus received less attention in the theoretical literature. But it does affect the transitional dynamics and thus the interpretation of mobility trends. Most importantly, the dynamic effect of structural changes on mobility trends will be smoothed out by variation in the timing of fertility around the mean age at which parents give birth. Mobility may therefore shift over multiple decades even when the system converges within a single generation. In contrast, *sudden* shifts in mobility across child cohorts must be due to contemporaneous events. We summarize these arguments in the following proposition and illustrate them further in Section 5:

Proposition 5. MOBILITY TRENDS OVER COHORTS. *While changes in the economic environment can have a sudden impact on mobility in the first affected generation, their effect on mobility trends over cohorts in subsequent generations will be gradual.*

Proof. See Appendix A.4.5. □

The IGE for a given cohort depends on the cohort-specific economic environment, and the variance and covariance of income and endowments among parents. However, as parents have children at different ages, parents of a given child cohort will belong to different cohorts and may thus be subject to different economic environments. Mobility levels and trends therefore depend on the current economic environment and a weighted average of the cross-covariances of income and endowments in previous cohorts, where the weights depend on the distribution of parental age at birth.[8] The effect of past structural changes on mobility trends in the current generation will therefore be gradual, as earlier generations are subject to different economic environments depending on the timing of fertility. Parental moments may vary by parental age also because of the selective nature of fertility.

The distribution of parental age is thus a key determinant of mobility trends, and its explicit consideration may help to isolate the impact of past structural changes on current

[8] A number of other implications follow. For example, mobility may adjust more quickly to structural changes in populations in which individuals become parents at younger ages, and mobility differentials across groups or countries may be partly driven by different *weights* on past economic environments due to differences in fertility pattern.

trends. For tractability, we however abstract from staggered fertility and life cycle dynamics in our theoretical discussion below. While this is an important limitation, the cohort-level dynamics – which effectively "smear" across the generational dynamics studied in our theoretical discussion – are not needed to understand some key main arguments (as summarized in Propositions 1-4). Moreover, we consider a cohort-level perspective in Section 5, showing how variation in the exposure to a school reform among parents shifts mobility across child cohorts, and provide a numerical example of cohort-level mobility trends in Appendix A.4.5. These examples illustrate how an explicit consideration of parental age-at-birth can help researchers to detect the impact of past structural shocks on current mobility trends.

3 Transitional Dynamics

In this section, we illustrate how the IGE shifts in response to different structural changes in the economic environment ξ. Our objective is to illustrate and to provide intuition for our analytical results above. We assume again that the structural change occurs in generation $t = T$, such that $\xi_{t<T} \neq \xi_{t\geq T}$, and that all moments were in steady-state equilibrium in generation $T-1$. We start with simplified versions of our baseline model, considering scalar versions of equations (5) and (6) with a single endowment e, and normalize the pre-shock variances of y and e to one. We consider separate shifts in each parameter, but also present cases in which two parameter change at once to provide some additional insights. We focus on the response of the IGE and study the joint *dynamics* of inequality and mobility instead in the next section.

Case 1. **A CHANGE IN THE EFFECT OF PARENTAL INCOME (γ).**

Figure 1a illustrates the response to an increase in the effect of parental income from $\gamma_{t<T} = \gamma_1$ to $\gamma_{t\geq T} = \gamma_2$. As is intuitive, an increase in the direct effect of parental income unequivocally increases the steady-state IGE (see Proposition 1). However, the shift to the new steady-state IGE is in general not immediate (Proposition 2a) and, depending on initial conditions, can be non-monotonic (Proposition 3), as in the parametrization considered here.

Case 2. **A CHANGE IN THE RETURNS TO SKILLS (ρ).**

Figure 1b illustrates the response to an increase in the returns to endowments or skills from $\rho_{t<T} = \rho_1$ to $\rho_{t\geq T} = \rho_2$. In the scalar version of our model with a single skill and $\lambda > 0$, increasing returns result in a higher steady-state IGE.[9] Again, the shift to the new steady-state IGE is not immediate. Indeed, in our chosen example, the second-generation shift is greater than the first-generation shift ("amplification", Proposition 2b). While we focus on

[9]This is not generally true in a model with multiple skills, as shown in the next section.

transition paths across generations here, Figure A.1 plots the "cohort-level" counterpart to the "generation-level" transition path shown in Figure 1b. As the timing of fertility varies across parents, the second-generation effect is spread out over many cohorts, providing an illustration of Proposition 5.

To understand why amplification is possible, consider a simplified case in which $\gamma = 0$ and the returns to transmittable skills (e) increase relative to other factors that are not transmitted from parents to children (u), such that the variance of y remains constant at their normalized pre-shock value ($Var(y_t) = 1$ for all t). The IGE in the first affected generation then shifts according to

$$\Delta\beta_T = \beta_T - \beta_{T-1} = (\rho_2 - \rho_1)\lambda\rho_1, \tag{11}$$

induced by the change in returns for generation in T. The second-generation shift,

$$\Delta\beta_{T+1} = \rho_2\lambda\Delta Cov(e_T, y_T) = \rho_2\lambda(\rho_2 - \rho_1), \tag{12}$$

is induced by the change in the covariance between income and endowments among the *parents* of generation $T+1$, caused by changing returns to those endowments in generation T. This second-generation shift $\Delta\beta_{T+1}$ is larger than the first-generation shift $\Delta\beta_{T+1}$, as the correlation between income and endowments is now strong for both generations.

Case 3. **A CHANGE IN THE HERITABILITY OF ENDOWMENTS (λ).**

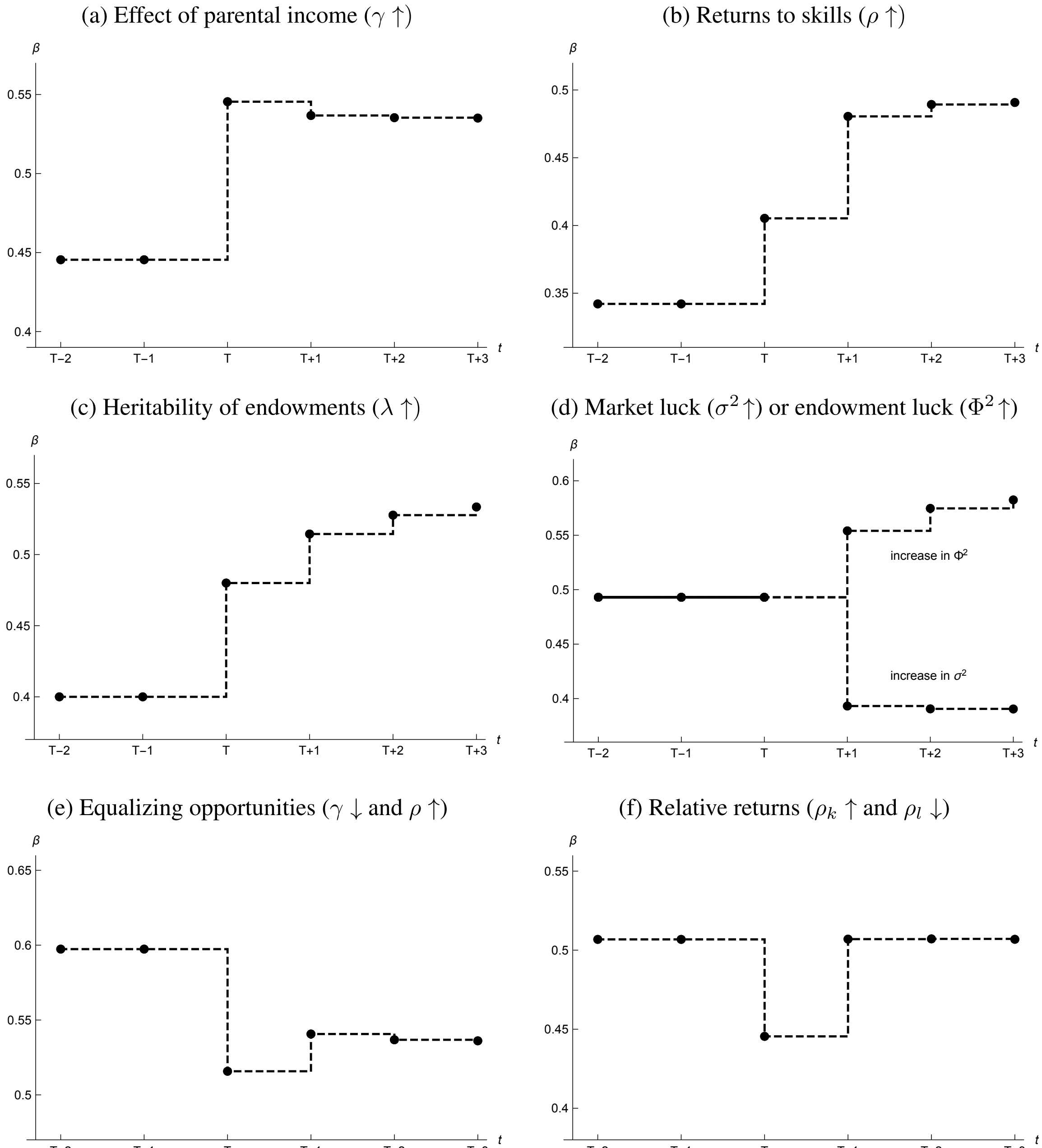
Figure 1c illustrates the response to an increase in the heritability of endowments, which always increases the steady-state IGE (see Proposition 1). However, the shift towards the new steady-state is comparatively slow, as increases in the variance of e and its covariance with y propagate further in subsequent generations.

Case 4. **A CHANGE IN THE VARIANCE OF MARKET LUCK (σ^2) OR ENDOWMENT LUCK (Φ^2).**

Figure 1d illustrates the response of the IGE to an increase in the variance of market luck (σ^2) or endowment luck (Φ^2). As is intuitive, an increase in market luck decreases the steady-state IGE (see Proposition 1). However, the shift towards the new steady-state is delayed, starting only in generation $T+1$, while the IGE in generation T remains unchanged. An increase in endowment luck increases the steady-state IGE instead (Proposition 1), and the IGE starts shifting only in generation $T+1$.

Which structural parameters should be shifted if the goal is to increase mobility both in the short-run and long-run? A comparison of panels a-d in Figure 1 illustrates that changes in the effect of parental income γ have the most immediate effect on the IGE, while changes in other parameters tend to have more delayed impacts. Of course, it remains difficult to map

Figure 1: Comparative Transitional Dynamics: Numerical Examples



Note: Numerical examples of trends in the intergenerational elasticity: (a) in generation T the parental income effect γ increases from $\gamma_1 = 0.2$ to $\gamma_2 = 0.3$ (assuming $\rho = \lambda = 0.6$); (b) in generation T the returns to skill increase from $\rho_1 = 0.2$ to $\rho_2 = 0.5$ (assuming $\gamma = 0.3$ and $\lambda = 0.8$); (c) in generation T the heritability of endowments λ increases from $\lambda_1 = 0.5$ to $\lambda_2 = 0.7$ (assuming $\gamma = 0.2$ and $\rho = 0.6$); (d) in generation T the variance of market luck or endowment luck doubles (assuming $\gamma = 0.2$, $\rho = 0.6$ and $\lambda = 0.7$); (e) in generation T the impact of parental income γ declines from $\gamma_1 = 0.4$ to $\gamma_2 = 0.2$ while the returns to skills increase from $\rho_1 = 0.5$ to $\rho_2 = 0.8$ (assuming $\lambda = 0.6$); and (f) in generation T the returns to skills k and l increase from $\rho_{k,1} = 0.3$ to $\rho_{k,2} = 0.6$ and decrease from $\rho_{l,1} = 0.6$ to $\rho_{l,2} = 0.3$ (assuming $\gamma = 0.2$ and $\lambda_k = \lambda_l = 0.6$).

this particular result on structural parameters into specific policies. For example, a reduction in tuition fees for college could alleviate credit constraints, and reduce the direct impact of parental income on college attendance – which we would interpret as a downward shift in γ. But if the role of parental income weakens, other characteristics might in turn become more important predictors of college attendance, which may have additional implications for intergenerational mobility.

While panels a-d of Figure 1 illustrate the consequences of single parameter changes, the remaining two panels illustrate the effect of changes in the relative importance of different transmission mechanisms. Consider first an example of "equalizing opportunities":[10]

> Case 5. **EQUALIZING OPPORTUNITIES.** Assume that the direct effect of parental income diminishes ($\gamma_1 > \gamma_2$), while skills are instead more strongly rewarded ($\rho_1 < \rho_2$).

In other words, assume that in generation T the economy becomes less *plutocratic* and more *meritocratic*. For example, parental status may become less and own merits more important for allocations into colleges, firms or occupations. Figure 1e provides an illustration, in which the IGE first decreases in generation T and then increases again in generation $T+1$. To understand why, consider a simplified case in which the parameters shift such that the variance of y remains constant ($Var(y_T) = Var(y_{T-1}) = 1$). Mobility then shifts in the first affected generation according to

$$\Delta\beta_T = (\gamma_2 - \gamma_1) + (\rho_2 - \rho_1)\lambda Cov(e_{T-1}, y_{T-1}), \tag{13}$$

due to both the declining importance of parental income, $\gamma_2 < \gamma_1$, and the increasing returns to endowments, $\rho_2 > \rho_1$. However, the latter effect is attenuated, for two reasons: endowments are imperfectly transmitted within families ($\lambda < 1$) and explain only part of the variation in parental income, such that $Cov(e_{T-1}, y_{T-1}) < 1$. Mobility thus tends to initially increase (for similarly-sized shifts in ρ and γ). Mobility also shifts in the second generation,

$$\Delta\beta_{T+1} = \rho_2\lambda\Delta Cov(e_T, y_T) = \rho_2\lambda\left((\rho_2 - \rho_1) + (\gamma_2 - \gamma_1)\lambda Cov(e_{T-1}, y_{T-1})\right), \tag{14}$$

due to shifts in the covariance between parental income and endowments. The relative impact of each parameter change is now reversed, with the change in γ rather than ρ being attenuated by $\lambda Cov(e_{T-1}, y_{T-1})$. Intuitively, a change towards a more meritocratic society increases the correlation between endowments and income, thereby *decreasing* mobility from the second affected generation and onwards.

[10] As noted by Conlisk (1974a), "opportunity equalization" is an ambiguous term that may relate to different types of structural changes in models of intergenerational transmission.

The example illustrates that the dynamic response of the IGE can be *non-monotonic*.[11] Whether the response is weakly or strongly non-monotonic as defined in Proposition 2 depends on parameter values; strong non-monotonicity with declining mobility in steady state is more likely when λ is high. The pattern stems from the relative gains and losses that the structural change generates. A rise in the returns to own skills relative to parental income is detrimental for children with high-income, low-skill parents. In contrast, it benefits talented children from poor families, providing opportunities for upward mobility that were not available to their parents. Mobility is high when these relative gains and losses occur. But the children of those who thrive under the new meritocratic setting will also do relatively well, due to the inheritance of endowments, so that mobility then decreases.[12]

The example also illustrates how changes that are mobility-enhancing in the long run may nevertheless cause a *decreasing* trend over several generations. A decline in mobility may then not necessarily reflect a recent deterioration of meritocratic principles, but rather major gains made in the past. From this perspective, if a country became more meritocratic in the early or mid 20th century, mobility should perhaps be *expected* to decline in more recent cohorts. Of course, the transitional dynamics to such change also depend on behavioral responses, that we do not model here (see for example Comerford et al., 2022).

As we discussed a quite specific structural change, one may expect that non-monotonic responses are more of an exception than a rule. We next illustrate that in a model with multiple skills, as in equations (5) and (6), such responses are instead typical:

> Case 6. **CHANGING RETURNS TO SKILLS.** Assume that the returns to different types of skills or endowments change on the labor market ($\boldsymbol{\rho}_1 \neq \boldsymbol{\rho}_2$).

Changes in the returns to different skills could stem from changes in relative supplies or demand: for example, demand may shift from physical to cognitive skills as a labor market transitions from agricultural to white-collar employment, or shift because of automation (e.g., Autor et al., 2003). Figure 1f provides a simple numerical example with two endowments k and l that are equally transmitted within families, but their returns swap in generation T ($p_{2,k} = \rho_{1,l} \neq p_{1,k} = \rho_{2,l}$). Mobility first increases, but decreases in subsequent generations.[13] Intuitively, mobility initially increases because the endowment for which returns

[11] Specifically, it will be non-monotonic if $(\gamma_1-\gamma_2)/(\rho_2-\rho_1) > \lambda Cov(e_{T-1}, y_{T-1}) < (\rho_2-\rho_1)/(\gamma_1-\gamma_2)$, which holds if $(\gamma_1-\gamma_2)$ and $(\rho_2-\rho_1)$ are sufficiently similar in absolute size. While non-monotonicity here requires changes in two parameters, it can also arise from a change in a single parameter if we allow for dynamic responses in the variances (see Proposition 2).

[12] That a shift towards "meritocratic" principles can also have depressing effects on mobility was already noted by the sociologist Michael Young, who coined the term in the book *The Rise of the Meritocracy* (1958). In contrast to its usage today, Young intended the term to have a derogatory connotation.

[13] We have $\Delta\beta_T = -\left(\rho_{k,2}-\rho_{k,1}\right)^2 \lambda/(1-\gamma\lambda)$, which is negative, and $\Delta\beta_{T+1} = \lambda(\rho_{k,2}-\rho_{k,1})^2 + \lambda(\rho_{k,2}^2 + \rho_{k,1}^2 + (2\rho_{k,1}\rho_{k,2}\lambda\gamma)/(1-\gamma\lambda))(1/Var(y_T)-1)$, which is positive since $Var(y_T) = 1-2\gamma\lambda(\rho_{k,2}-\rho_{k,1})^2/(1-\gamma\lambda) < 1$. These findings are not due to shifts in cross-sectional inequality; if instead $Var(y_T) = 1$ (i.e. changes in ρ_k and ρ_l are offset by changes in the variance of u_t) we still have that $\Delta\beta_T < 0$ and $\Delta\beta_{T+1} > 0$.

increase from low levels is less prevalent among high-income parents than the endowment for which returns decrease from high levels. But the endowment for which returns rise becomes increasingly associated with income in subsequent generations, causing a decreasing mobility trend. This result has implications for how we expect institutional or technological change to affect mobility. Previous work suggests that technological progress can lead to non-monotonic mobility trends through *repeated* changes in skill premia (Galor and Tsiddon, 1997). We find that even a *one-time* change can generate such trends if comparative advantages in skills or endowments are partially transmitted within families.

To better understand this non-monotonic response, consider the general case in which the returns to *any* number of skills change. We assume here a diagonal heritability matrix, while the derivation for non-diagonal $\mathbf{\Lambda}$ is given in Appendix A.7. The steady-state IGE before the structural change is then equal to

$$\beta_{T-1} = \gamma + \boldsymbol{\rho}_1' \mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1} \boldsymbol{\rho}_1, \tag{15}$$

while, if the income variance remains constant, its steady-state level after the change is

$$\beta_{\infty} = \gamma + \boldsymbol{\rho}_2' \mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1} \boldsymbol{\rho}_2. \tag{16}$$

The IGE in the first affected generation, $\beta_T = \gamma + \boldsymbol{\rho}_1' \mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1} \boldsymbol{\rho}_2$, can therefore be expressed as

$$\beta_T = \frac{1}{2}\left(\beta_{T-1} + \beta_{\infty}\right) - \frac{1}{2}\left(\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'\right) \mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1} \left(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1\right), \tag{17}$$

where the quadratic form in the last term is greater than zero for $\boldsymbol{\rho}_2 \neq \boldsymbol{\rho}_1$ since $\mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1}$ is positive definite. It can therefore be decomposed into two parts, the average of the old and the new steady-state IGE minus a *transitional* drop. Changes in returns thus cause a temporary spike in mobility (β_T is below β_{T-1} and $\beta_{t\to\infty}$) as long as the steady-state IGE does not shift too strongly, specifically if

$$|\beta_{\infty} - \beta_{T-1}| < (\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1') \mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1} \left(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1\right). \tag{18}$$

This argument also holds if cross-sectional inequality is lower in the new than in the old steady state.[14]

Based on our last two cases we can formulate a more general conclusion that extends on Proposition 3. A change in the strength of one channel of intergenerational transmission *relative to* another affects the prospects of families differently. For example, a decline in the

[14]Eq. (17) then includes the additional term $\boldsymbol{\rho}_2' \mathbf{\Lambda} \left(\boldsymbol{I} - \gamma \mathbf{\Lambda}\right)^{-1} \boldsymbol{\rho}_2 \, (1 - \frac{1}{Var(y_{t\to\infty})})$, which is negative if $Var(y_{t\to\infty}) < Var(y_{T-1}) = 1$.

importance of parental income relative to own skills diminishes the prospects of children with high-income parents. Similarly, a decline in returns to a particular skill hurts those families in which that skill is more abundant. Economic and social changes that generate such relative gains and losses will tend to generate *transitional* mobility in the generation in which they occur – times of change tend to be times of high mobility.[15] Many developed countries have experienced greater societal transformations in the first than in the second half of the 20th century, and those transformations may have increased mobility in those generations that were directly affected but decreased it subsequently. Our analysis suggests that such transitional gains diminish as the economic environment stabilizes.

4 Joint Dynamics of Mobility and Inequality

We already noted that the transitional dynamics of the IGE and other mobility measures depend also on shifts in the variance of income across generations. We now consider such shifts in cross-sectional inequality, their interrelation with intergenerational mobility, and how evidence on recent mobility trends in the US can be interpreted in light of our findings.

4.1 Transitional Dynamics in Cross-Sectional Inequality

The steady-state relationship between cross-sectional inequality and intergenerational persistence was emphasized already by Becker and Tomes (1979), has been studied further in Solon (2002), Davies et al. (2005) and Hassler et al. (2007), and recently reviewed by Durlauf et al. (2021). The transitional *dynamics* of inequality and mobility are also intertwined. Note first that due to intergenerational mechanisms, changes in cross-sectional inequality tend to propagate across generations. For example, in response to a shift in the variance of market luck from σ_1^2 to σ_2^2 in generation T, the variance of income initially shifts by the same amount ($\Delta Var(y_T) = \sigma_2^2 - \sigma_1^2$), but then continues to shift in future generations according to (see Appendix A.3):

$$\Delta Var(y_t) = \quad \gamma^2 \Delta Var(y_{t-1}) \quad \forall t > T. \tag{19}$$

In this example, the variance of income will transition in infinite time towards its new steady state if $\gamma > 0$. Such transitional dynamics in cross-sectional inequality also affect aggregate measures of mobility, and affect the transition paths of different mobility measures differently

[15] This argument extends to other contexts. For example, assume that the vector $\boldsymbol{e}_t$ includes the location of individuals, "inherited" with some probability from their parents. We can then relate our argument to Long and Ferrie (2013), who argue that US occupational mobility has been high in the 19th century as of exceptional *geographic* mobility. Our result illustrates that not only internal migration itself but also its underlying causes may increase intergenerational mobility, if local shocks affect parents and their (non-migrating) children differently.

(see Proposition 4). In particular, the variance of market luck has no effect on the covariance between endowments and income in eq. (7), and thus no initial effect on the IGE. However, the IGE shifts in generation $T+1$, according to (see Appendix A.3)

$$\Delta\beta_{T+1} = \beta_{T+1} - \beta_T = -\frac{\rho^2\lambda}{1-\gamma\lambda}\frac{\sigma_2^2 - \sigma_1^2}{1+\sigma_2^2-\sigma_1^2}. \tag{20}$$

Other mobility measures, such as the intergenerational or sibling correlation, shift already in generation T.[16] We next illustrate how these insights affect the interpretation of mobility trends.

4.2 Rising Skill Premia and Income Mobility

Interest in the relationship between inequality and mobility has been spurred by two observations. First, the US and other rich countries have experienced rising skill premia and an increase in income inequality since around 1980. Second, many studies find a negative correlation between cross-sectional inequality and intergenerational mobility across (Corak, 2013) and within countries (Chetty et al., 2014a, Güell et al., 2018, Connolly et al., 2019), a relation now popularly known as the "Great Gatsby Curve".[17] But despite this association, and the prediction from standard models that rising skill premia decrease intergenerational mobility (Solon, 2004), it remains debated whether and to what extent mobility actually deteriorated in recent decades. Lee and Solon (2009) reject large changes in PSID data up until around 2000 and Chetty et al. (2014b) find stable mobility in tax data over cohorts born in the 1970s and early 1980s, while Justman and Stiassnie (2021) find a mobility decrease in more recent waves of the PSID. Davis and Mazumder (2020) incorporate data on earlier cohorts, finding a drop in mobility that occurred just prior to the cohorts observed in the PSID.[18]

Why was there not a more pronounced decrease of income mobility in more recent decades, despite rising returns to skill and inequality? Our framework points to three potential explanations. First, in a multi-skill model, rising skill returns do not necessarily decrease intergenerational mobility in steady state (see also Proposition 1). Second, the transitional dynamics of inequality and mobility can deviate strongly from their steady-state relationship. In particular, the effect of rising skill returns can be very different in the first and second af-

[16] See Proposition 4 for a discussion of the intergenerational correlation. Appendix A.5 analyzes the transitional dynamics of the sibling correlation.

[17] The name was coined in a speech by Krueger (2012), who notes that "[...] based on the rise in inequality that the United States has seen from 1985 to 2010 and the empirical evidence of a Great Gatsby Curve relationship, I calculated that intergenerational mobility will slow by about a quarter for the next generation of children." See also footnote 1 and Durlauf et al. (2021) for a more theory-focused discussion of the Great Gatsby Curve.

[18] To identify this shift, Davis and Mazumder (2020) compare cohorts who entered the labor market before and after the sharp increase in inequality around 1980, which is not possible to do well in the PSID but can be done in datasets like the National Longitudinal Surveys. Other recent papers on US mobility trends include Jácome et al. (2021) and Palomino et al. (2017).

fected generation. Third, contemporaneous mobility trends in the US might be also affected by structural changes that pre-date and offset the recent increase in skill returns. Of course, it could also be offset by other contemporaneous structural changes, such as a decrease in the direct influence of parental income. For example, Lee and Solon (2009) note that the mobility-depressing effect of increasing skill returns could have been offset by more progressive public investment in children's human capital. We illustrate the first two arguments using our theoretical model, and then provide evidence from the Panel Study of Income Dynamics supporting our third argument in the next section.

Rising Skill Returns and Steady-State Mobility. In a model with multiple skills or endowments, an increase in skill returns has ambiguous effects on steady-state mobility. For illustration, consider the following example:

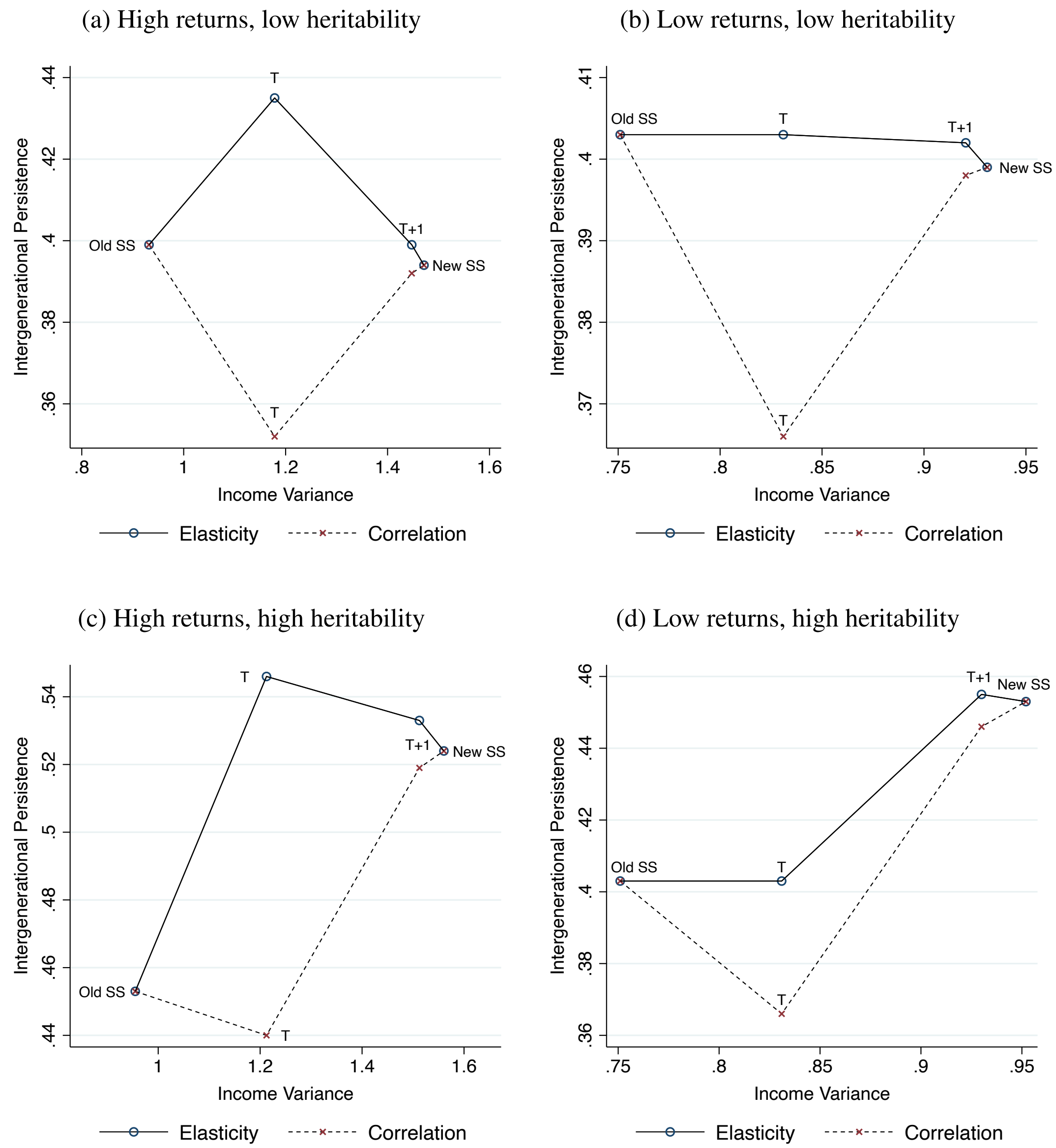
> Case 7. **TRANSITIONAL DYNAMICS IN THE "GREAT GATSBY CURVE".** Assume th*at children inherit two endowments k and l from their parents. In* generation T, the return to endowment k increases, while the return to endowment l remains unchanged.

Figure 2 plots, for four different environments (or "countries"), the transition paths of the intergenerational elasticity, the correlation, and the variance of income.[19] While otherwise characterized by the same environment, the initial returns to endowment k ($\rho_{k,1}$) and thus inequality are higher in countries (a) and (c). Moreover, its heritability λ_k is lower in (a) and (b) than in the other two countries. We find that the same absolute change in the returns to endowment k decreases *steady-state* persistence in (a) and (b) but increases it in (c) and (d). A sufficient condition for a non-increasing steady-state IGE after an increase in returns ρ_k is $\beta_{T-1} \geq \frac{\gamma+\lambda_k}{1+\gamma\lambda_k}$, *which* holds if the heritability of k is sufficiently low relative to other determinants of the IGE.[20] This observation contrasts with the prediction from standard models with a single skill, in which increasing returns unambiguously increase the IGE (Solon, 2004). It further suggests that the same structural change can have opposing effects on steady-state mobility if countries differ in initial conditions. For example, steady-state persistence increases the most in (c), where endowment k is both strongly inheritable and yields high returns. This is intuitive, since it speaks to a case where rich families in the past were rich due to the same endowment that drives increasing inequality in current times.

[19]Inequality measures may mix information from multiple generations, and therefore lead to a temporal aggregation problem as illustrated in Working (1960). We consider the average of the variance in the parent and child generation here. To measure inequality in a cross-section with overlapping generations may lead to stronger transitional dynamics, in particular if average incomes change across generations.

[20]This condition is derived in the proof to Proposition 1 in Appendix A.4.1. The result extends to settings with more than two skills: an increase in the return to a single endowment decreases steady-state mobility only if *its heritability is high relative to the combined importance of other determinants of income.* The arguments can be easily understood by noting that a non-heritable skill is akin to market luck, which increases mobility.

Figure 2: An Increase in the Returns to a Single Skill



Note: Transitional dynamics of the intergenerational elasticity (solid line) or correlation (dashed line) and the average of the variance of income in the parent and child generation. Parameters are $\gamma = 0.2$, $\lambda_l = 0.8$, $Var(u_t) = 0.5$, and $\rho_l = 0.4$. $\lambda_k = 0.2$ in sub-figures a and b, and $\lambda_k = 0.5$ in sub-figures c and d. $\rho_{k,1} = 0.4$ in sub-figures a and c, and $\rho_{k,1} = 0$ in sub-figures b and d. In generation T, the returns to skill k increase by $\rho_{k,2} - \rho_{k,1} = 0.4$ in all cases. See Appendix A.5 for a corresponding numerical illustration of the dynamics of the sibling correlation.

Rising Skill Returns and Transitional Dynamics. Most relevant for the recent debate, however, is the effect of rising skill returns on mobility in the first affected generation. Figure 2 illustrates that the initial response in the first affected generation can be small if the pre-shock return to that endowment is small (cases (b) and (d)). More generally, the transition path that countries take through the Gatsby diagram can be complex: while the path of *inequality* is monotonic, the paths of the two mobility measures can be non-monotonic. In some cases, such as in subfigure (a), the first-generation and steady-state shifts have different signs ("strong non-monotonicity", see Proposition 3). Even if the steady-state response is in line with the static Gatsby diagram, the first-generation effect may not be (case (d)). An understanding of transitional dynamics is thus useful not only for the interpretation of mobility trends, but also for their relationship to cross-sectional inequality.

Can these observations help us understand mobility changes in recent decades? In particular, does rising inequality primarily reflect a rising importance of skills that are not strongly transmitted within families, or that were not very important in the parent generation? Some recent evidence would be in line with this interpretation: For example, Deming (2017) and Edin et al. (2021) find substantial increases in the earnings returns to socio-emotional skills over the last decades, which appear to be less strongly transmitted within families than cognitive skills (e.g. Loehlin, 2005). Yet, the gap in the transmission of cognitive and socio-emotional skills would need to be very large to explain why a rise in returns to the latter would have no initial effect on the IGE, leading us to believe that this cannot be the main explanation.[21]

Figure 2 also plots the intergenerational correlation (IGC), which tends to react quite differently from the IGE during the transition. In all examples, the IGC decreases in the first affected generation. Because the IGC decreases in the contemporaneous variance of income, an increase in returns tends to decrease the IGC, unless the affected skill was the dominant determinant of the intergenerational income correlation prior to the change. Different measures of mobility follow therefore different transitional dynamics, especially in the first affected generations (illustrating Proposition 4).[22]

4.3 Evidence on US Mobility Trends

We finally study whether US mobility trends might be influenced by structural changes that pre-date and offset the recent increase in income inequality. As reflected in equation (7), the IGE is not only a function of the current economic environment, but also of the covariance of income and endowments in the parent generation. A decrease in this covariance, for example, might counteract a mobility-depressing effect of rising skill prices.

4.3.1 Data

To explore this hypothesis, we analyze trends in income mobility in the Panel Study of Income Dynamics. Our sampling choices are guided by Lee and Solon (2009) and further de-

[21]Moreover, Grönqvist et al. (2017) show that the heritabilities of cognitive and non-cognitive/socio-emotional skills are quite similar once measurement error is taken into account.

[22]The latter argument may also help to explain why Levine and Mazumder (2007) find a sharp increase in *sibling* correlations since 1980, while there is less evidence of an increase in *intergenerational* persistence. The steady-state response to growing skill returns in our model is similar in both measures. But sibling correlations respond more immediately (if $\gamma = 0$ they respond fully in generation T) because they depend less directly on returns in the parent generation (see Appendix A.5).

scribed in Appendix A.8. We first identify parent-child pairs and construct income and skill measures for both generations. Specifically, we measure the household income of parents as an average over the years when the child was age 15-17. We measure annual household incomes in the child generation when the son or daughter was age 30-35, which allows us to include the 1980s birth cohorts in our analysis and facilitates comparability with both Lee and Solon (2009) and Chetty et al. (2014b).[23] While the IGE (in lifetime income) will be understated in this age range, our objective is to measure its trend rather than its level.

4.3.2 Results and interpretation

Table 1, Panel A reports estimates from a regression of log child income on log parent income, year and age controls, and an interaction between child age and log parental income to control for lifecycle effects (see Appendix A.8). We estimate this regression separately for four groups of birth cohorts born in the 1950s, 1960s, 1970s and 1980s. We find no dramatic changes in the IGE over this period (with $\hat{\beta}$ at or slightly above 0.4).[24]

In Panel B we report the corresponding trend in income inequality in the parent and child generations. Consistent with prior evidence, the variance of income increases substantially over cohorts. Perhaps more surprisingly, this trend affects the parent and child measures similarly. This observation is primarily due the fact that – as other studies – we measure parents' income at a later age than the income of their children. This asymmetry reduces the gap in calendar time between the measurement of parent and child income, and amplifies measures of income inequality in the parent generation (as age-income profiles tend to diverge over age). For comparability with previous research, we retain this asymmetry here.[25]

In Panel C we report trends in the skill premium, as approximated by a regression of log incomes on years of schooling (again controlling for year and age). We consider schooling as a proxy measure of e in our single-skill model.[26] Consistent with prior evidence, we find that the premium increased over the child cohorts in our sample. Other things equal, we would expect this rising skill price to increase the IGE (see Appendix A.3), in particular since schooling is relatively persistent between generations (Hertz et al., 2008). However, the variance of schooling in the parent generation drops strongly for the 1960s cohorts and again for the 1970s cohorts, while remaining more constant in the child generation. As noted by Hilger (2015), this evolution is driven by rising high school attainment among the parents of

[23]For simplicity and data reasons, we use years of schooling as our measure of skill, and do not consider multiple skills.

[24]As mentioned earlier, our estimates here do not capture a potential decline in mobility relative to earlier cohorts that are not well captured by the PSID (Davis and Mazumder, 2020).

[25]Transitional dynamics in income inequality therefore have a less mechanical effect on standard estimates of the IGE than one might otherwise expect. The same argument could explain why *both* the IGE and measures based on adjusted distributions, such as Pearson or rank correlations, can remain stable over time.

[26]The estimated skill premia are lower but exhibit a similar trend when controlling for parental income, a specification that corresponds more closely to our structural equation (5).

the cohorts born in the mid-twentieth century.

Moreover, Panel D shows that the correlation between parental education and parental income also decreased over cohorts. As such, the ratio of the covariance between parent education and income and the variance of income, corresponding to $\frac{Cov(\boldsymbol{e}_{t-1}, y_{t-1})}{Var(y_{t-1})}$ as featured in our decomposition of the IGE in Section 2, falls substantially – it is less than half as large for the 1980s as compared to the 1950s cohort. Finally, the last row in Table 1 illustrates how the IGE *would have* evolved over cohorts, had the ratio $\frac{Cov(e_{t-1}, y_{t-1})}{Var(y_{t-1})}$ stayed constant. Instead of the observed marginal decrease from 0.42 to 0.40, the IGE would have increased by 20 percent to 0.50.[27]

This evidence therefore suggests that important changes in the *parent* generation offset the effect of the rise in skill prices: rising skill premia did depress mobility, but this effect was counteracted by the mobility-enhancing effects of an increasingly compressed distribution of schooling in the parent generation. Of course, our analysis here ignores general equilibrium effects, such as the effect of changes in skill supplies on skill returns (Katz and Murphy, 1992), but it does illustrate that observing a largely stable IGE does not necessarily imply that the transmission system itself has remained stable.

Researchers should therefore consider both current and more distant events when interpreting contemporaneous trends in intergenerational mobility. The variances and covariance of income and education in the parent generation are key statistics to consult in this regard. With richer data on parents, researchers may extend this analysis to other parental characteristics, such as cognitive and non-cognitive skills (for an example, see Markussen and Røed, 2020).

5 The Dynamic Effects of a Compulsory Schooling Reform

Our theoretical analysis suggests that a single structural change can generate trends in intergenerational mobility across multiple generations. We now aim to provide causal evidence on such long-lasting dynamics, a task that is demanding in terms of both data coverage and identification. We consider the Swedish compulsory schooling reform, first studied by Meghir and Palme (2005) and outlined in Holmlund (2007). Gradually implemented across municipal-

[27] Separately for each cohort group (decade) c, we estimate ρ and λ using the single-skill counterparts of equations (5) and (6). Combining these estimates with our estimates of $\frac{Cov(e_{t-1}, y_{t-1})}{Var(y_{t-1})}$ (Panel D of Table 1) we compute the counterfactual IGE as

$$\tilde{\beta}_c = \hat{\beta}_c + \hat{\rho}_c \hat{\lambda}_c \left(\frac{\widehat{Cov(e_{1950,t-1}, y_{1950,t-1})}}{Var(y_{1950,t-1})} - \frac{\widehat{Cov(e_{c,t-1}, y_{c,t-1})}}{Var(y_{c,t-1})} \right),$$

where $\hat{\beta}_c$ is the original decade-of-birth specific IGE estimate reported in Panel A of Table 1. All estimates are conditional on calendar year, child and parental age (where applicable), and normalized to child age 33 (see also Appendix A.8).

Table 1: Income Mobility in the US over Four Decades

Birth cohort	1950s	1960s	1970s	1980s
Panel A: IGE				
IGE	0.421	0.419	0.427	0.400
	(0.040)	(0.039)	(0.036)	(0.042)
No. individuals	1,094	1,011	1,111	902
No. individual x year observations	5,829	3,874	2,975	1,788
Panel B: Income inequality				
std. dev. INC_{it}	0.67	0.62	0.70	0.83
std. dev. $PINC_i$	0.58	0.58	0.68	0.80
std. dev. INC_{it} / std. dev. $PINC_i$	1.15	1.07	1.03	1.04
Panel C: Returns to schooling				
$\hat{\rho}$ (child generation)	0.122	0.121	0.160	0.178
	(0.010)	(0.010)	(0.011)	(0.016)
$\hat{\rho}$ (parent generation)	0.088	0.098	0.122	0.148
	(0.005)	(0.006)	(0.008)	(0.010)
std. dev. EDU_i	2.21	2.08	2.07	2.05
std. dev. $PEDU_i$	3.36	2.89	2.36	2.35
std. dev. EDU_i / std. dev. $PEDU_i$	0.66	0.72	0.88	0.87
Panel D: Covariances among parents				
$Corr(PEDU_i, PINC_i)$	0.513	0.486	0.425	0.436
	(0.028)	(0.027)	(0.029)	(0.030)
$Cov(PEDU_i, PINC_i)$	1.03	0.83	0.68	0.82
$Cov(PEDU_i, PINC_i)/Var(PINC_i)$	3.08	2.46	1.49	1.29
Panel E: Counterfactual IGE with constant covariance-variance ratio				
Counterfactual IGE	0.421	0.439	0.503	0.492

Note: See Appendix A.8 for sample and variable definitions. Panel A reports IGE estimates from a regression of log child family income (INC_{it}) on log parent family income ($PINC_i$), year and age controls, and an interaction between child age and $PINC_i$, estimated separately for cohorts born in the 1950s, 1960s, 1970s and 1980s. Panel B estimates income inequality in the parent and child generations, with INC_{it} measured when the child was age 30-35 and $PINC_i$ measured as an average when the child was age 15-17. Panel C reports skill premia in the parents and child generations based on regressions of log incomes on years of schooling (EDU_i or $PEDU_i$), controlling for year and age. Panel D shows correlations and covariances between parental education and parental income, and the ratio of the covariance between parent education and income and the variance of income, as a proxy $Cov(e_{t-1}, y_{t-1})/Var(y_{t-1})$ as featured in Section 2. Panel E reports a counterfactual IGE, had the ratio $Cov(PEDU_i, PINC_i)/Var(PINC_i)$ stayed constant at the level of the 1950s cohort group (see footnote 27). Standard errors in parentheses are clustered at the individual (child). Source: Panel Study of Income Dynamics.

ities from the late 1940s, the reform raised compulsory schooling from seven (or eight) to nine years and postponed tracking decisions (see Appendix A.10 for details).

This application is interesting for three reasons. First, education is a key mechanism for the transmission of income (Becker and Tomes, 1979) and educational reforms are thus potential determinants of mobility trends (e.g. Machin, 2007). Reforms similar to the Swedish one were enacted in many Western countries during this period, and did indeed raise mobility in the directly affected generation (Holmlund, 2008; Pekkarinen et al., 2009; Karlson and

Landersø, 2021). Second, we have access to an unusually rich data set. While we lack useful data to analyze the role of skill multiplicity, they do cover long-run outcomes and parent-child linkages of three generations. Third, the reform's gradual implementation across areas allows separation of the reform from regional or time-specific effects.[28]

5.1 Data and Descriptive Evidence

Our sample is based on a random 35 percent draw of the Swedish population born 1943-1955 (the directly affected cohorts), their parents, and their children. We add income data from tax declaration files and years of schooling from an education register. For further data details, see Appendix A.11.

Figure 3 illustrates the timing of the reform. The share of children subject to the reform increases sharply in cohorts 1943-1955 (grey area). These individuals become parents themselves from the 1960s, but their share among all fathers (black area) increases only slowly over child cohorts, due to variation in the timing of fertility. As summarized in Proposition 5, the dynamic effect of structural changes on mobility trends should thus be *gradual* from the second affected generation and onwards.[29] Figure 3 also shows that the roll out of the reform coincides with a large drop in the slope coefficient in a regression of child's years on father's years of schooling. The degree to which differences in schooling are transmitted to the next generation declines by more than a third, consistent with our theoretical expectation.[30] However, after its large decline, the coefficient starts to gradually rise again among cohorts born in the late 1960s. The trend line is similar to the one based on the sibling correlation in Björklund et al. (2009), although the initial drop is larger and the subsequent rebound occurs somewhat earlier in their study (in line with our expectations from Appendix A.9).

5.2 The Reform Effect on Intergenerational Mobility

We exploit the roll out of the reform to estimate its causal impact, adapting a difference-in-differences approach as in Holmlund (2008). The specification is easier to describe as a two-step procedure.[31] In a first step consider, for each cohort c and municipality m, the

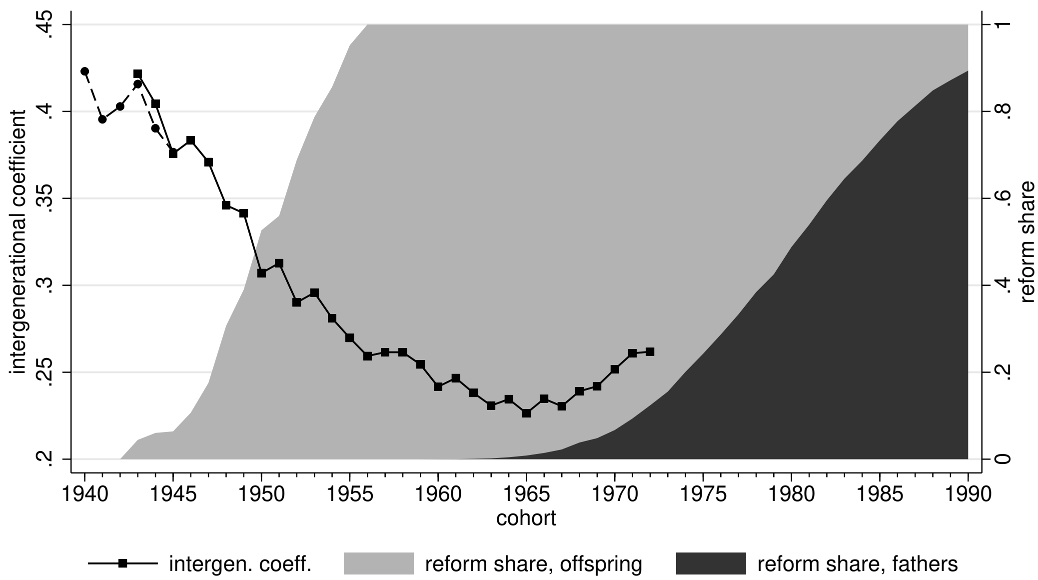
[28]A number of studies exploit this characteristic to assess the reform impact on *individual* outcomes in directly affected or in subsequent generations (see e.g. Meghir and Palme, 2005; Holmlund et al., 2011; Meghir et al., 2011). We examine instead its effect on summary measures of intergenerational mobility.

[29]Since we observe schooling only for those born 1911 and later we restrict our estimation sample to fathers who were 33 years or younger at the birth of their child. Our results will therefore understate the longevity of the reform's effect on mobility measures.

[30]The impact of a compulsory schooling policy on educational and income mobility can be predicted from a variant of our theoretical framework (see Appendix A.9). Our model predicts a drop in the intergenerational coefficient in education and income in the first affected generation, and a gradual increase in the next.

[31]We thank an anonymous referee for this suggestion.

Figure 3: Reform Coverage and Trends in the Intergenerational Educational Coefficient



Note: The figure shows intergenerational educational coefficients (see left-side y axis), i.e. coefficients from regressions of years of schooling of offspring in the respective birth cohort on years of schooling of their fathers, based on intergenerational sample (fathers aged below 33 at birth, solid line) and subsample (fathers aged below 30, dashed line). It also shows the shares of offspring (grey area) and fathers (black area) subject to school reform over offspring cohorts in source data (see right-side y axis).

regression model

$$y_{cmt} = \alpha_{cm} + \beta_{cm} y_{cmt-1} + u_{cmt}, \tag{21}$$

where y_{cmt} is a measure of socio-economic status of the child in generation t of family i (subscript suppressed), y_{cmt-1} the corresponding measure of the father, and β_{cm} a measure of intergenerational persistence (e.g. the IGE). Our interest centers on the second-step model

$$\beta_{cm} = \boldsymbol{\alpha}_1' \boldsymbol{D}_c + \boldsymbol{\alpha}_2' \boldsymbol{D}_m + \gamma R_{cm} + v_{cm}, \tag{22}$$

which allows for mobility differences across cohorts and municipalities (captured by indicator vectors $\boldsymbol{D}_c$ and $\boldsymbol{D}_m$). The indicator R_{cm} equals one if the reform was in place for cohort c in municipality m, and γ captures the reform effect.

We estimate this reform effect in both the first affected and the subsequent generation.[32] In the former (cohorts born 1943-1955 and their fathers), subscript c refers to the child's cohort, while in the latter (cohorts born 1966-1972) it refers to the father's cohort and treatment status – while all children of this generation attended reformed schools, only some of their fathers

[32] As the dependent variable in equation (22) is estimated, its sampling distribution needs to be taken into account to obtain standard errors and efficient estimates of γ (see Hanushek, 1974). In practice, we estimate both steps at once, pooling across cohorts and municipalities, and interacting the intercept and regressor of equation (21) with each of the regressors in the second-stage equation.

Table 2: Reform Effect on Educational and Income Mobility

	Generation 1		Generation 2	
	Education	Income	Education	Income
A. Regression slope	*(# years)*	*(log)*	*(# years)*	*(log)*
baseline	0.422***	0.139***	0.294***	0.244***
	(0.0075)	(0.0162)	(0.0041)	(0.0093)
reform effect	-0.037***	-0.020**	0.066***	0.041*
	(0.0072)	(0.0100)	(0.0128)	(0.0216)
B. Standardized slope	*(# years)*	*(log)*	*(# years)*	*(log)*
baseline	0.362***	0.106***	0.402***	0.191***
	(0.0063)	(0.0126)	(0.0059)	(0.0073)
reform effect	-0.040***	-0.015*	0.080***	0.033*
	(0.0074)	(0.0087)	(0.0163)	(0.0173)
C. Rank-rank slope	*(rank)*	*(rank)*	*(rank)*	*(rank)*
baseline	0.420***	0.117***	0.410***	0.213***
	(0.0078)	(0.0115)	(0.0061)	(0.0062)
reform effect	-0.023***	-0.009	0.053***	0.014
	(0.0088)	(0.0087)	(0.0147)	(0.0155)
N	220,335	199,340	111,173	110,317

Note: The table reports estimates of γ in equation (22) based on child cohorts 1943-1955 (first generation) or 1966-1972 (second generation) and their fathers, using years of schooling or log income as status measure (Panel A), standardized (Panel B) or percentile ranked (Panel C) within each child and father cohort. Clustered (municipality level) standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

did (see Figure 3). The identifying variation is local changes in mobility after introduction of the reform. While controlling for fixed cohort and area effects, a common concern with this type of strategy is differences in area-specific trends. Moreover, the reform indicator is potentially measured with some error, which may introduce attenuation bias. Appendix A.12 provides sensitivity analyses, showing that our main results are robust to both these as well as a set of other potential concerns.

Panel A of Table 2 reports estimates of the reform effect γ on the intergenerational coefficient in years of schooling and log income (the IGE).[33] Upon introduction of the reform, persistence in both schooling and income decreased by about ten percent. In line with Holmlund (2008), we thus find that the reform raised mobility in the first affected generation. But our main question is if the reform caused prolonged dynamics in later cohorts. Figure 3 shows that after its long decline, the intergenerational coefficient starts rising again among

[33] As we measure average incomes when the children are young (age 30-35) but the fathers older (age 53-59), our baseline estimate understates the IGE in lifetime income (Nybom and Stuhler, 2016). Moreover, our estimates capture mobility within areas, which do not aggregate immediately to mobility at the national level (see Hertz, 2008).

cohorts born in the late 1960s, the first cohorts in which some fathers had attended reformed schools. Indeed, Appendix Figure A.2 shows that the coefficient increases only for fathers who were sufficiently young to be exposed to the reform. The estimates in Table 2 confirm that the persistence in both schooling and income indeed increased in response to the reform in the *previous* generation.

The estimates of γ are larger for the second than for the first generation, for two reasons. First, the timing of fertility (see Section 2.2): among cohorts born in the 1960s, only young parents can themselves have been subject to the reform. As young parents tend to have less schooling, the reform's impact on this group was large. Second, these parents are more likely to have been born in the early 1940s than later. As of the secular rise in average schooling over time, the minimum schooling restriction was more binding in these earlier cohorts – the reform effect is heterogeneous across first-generation cohorts.

The reform compresses the distribution of schooling and income, and the IGE is particularly sensitive to such variance changes. However, this sensitivity can extend to other mobility measures for which the link to cross-sectional inequality is less obvious. To show this, we standardize the variance of our status variables before estimation or transform them into percentile ranks within the national distribution of each cohort (as in Chetty et al., 2014a). The sign and magnitude of the estimated reform effect on the standardized (i.e. correlation) coefficient (Panel B of Table 2) is similar to the effect on the regression coefficient. Intuitively, by standardizing variables within the national distribution of a cohort we abstract from broad changes in inequality, but not from changes in inequality that occur within areas or subgroups. The magnitude of the reform effect on the rank-rank relationship (Panel C) is smaller, and statistically significant only for education.

These findings support and illustrate some of our key theoretical results. First, the existence of transitional dynamics: as also illustrated for the US, recent mobility trends can indeed be caused by events that occurred in previous generations (Proposition 1). Second, our findings confirm that transitions can be non-monotonic (Proposition 2), and illustrate the close relationship between the dynamics of cross-sectional and intergenerational inequality.

6 Conclusions

We examined the dynamic relationship between intergenerational mobility and its underlying structural factors, leading to four key theoretical results. First, changes in the economic environment affect mobility not only in the directly affected but also in *subsequent* generations; policy or institutional changes may therefore generate long-lasting mobility trends. Second, these transitional dynamics can be non-monotonic. Mobility shifts in the first affected generation may therefore give a misleading picture of the long-run consequences of structural changes. Third, such changes can lead to relative gains and losses that generate *transitional* mobility; times of change therefore tend to be times of high mobility and negative mobility trends may stem from *gains* in equality of opportunity in the past. Fourth, mobility measures interact with the transition path of cross-sectional inequality, and different mobility measures can exhibit quite dissimilar transitional dynamics.

We illustrated the first two results empirically, by studying US mobility trends as well as the effects of a Swedish compulsory schooling reform on parent-child mobility across multiple generations. We first showed that changes in the *parent* generation may be key to understand why mobility seems to have remained fairly stable across recent decades in the US. While rising skill returns did put downward pressure on mobility, a substantial compression of the parental schooling distribution counteracted this effect, resulting in a roughly constant IGE for cohorts born between the 1950s and 1980s. We then showed that the Swedish schooling reform increased income and educational mobility in directly affected cohorts (see also Holmlund, 2008). But the reform's impact in the subsequent generation went in the *opposite* direction, suggesting that its long-run effect on mobility may have been small.

Our model is of course stylized, and its application to other settings may require careful treatment of issues that we did not address. These include the timing of intergenerational transmission mechanisms over an individual's life cycle and their potential endogeneity to changes in the economic environment (see Heckman and Mosso, 2014), as well as the difficulties that hinder reliable estimation of mobility trends. In general, it is a difficult task to track how events in past generations affect mobility across multiple generations. Still, we illustrated how a consideration of transitional dynamics may be fruitful in the interpretation of mobility trends related to both specific events (such as the Swedish schooling reform) or broader structural change (such as the constancy of income mobility vis-a-vis rising skill returns in the US).

In addition, our results point to a number of specific implications that we discussed only briefly. We noted that rising skill premia shift intergenerational measures over at least two generations, suggesting that the overall effect may not yet be fully visible in current estimates. Other measures of the importance of family background respond more quickly, potentially explaining why sibling correlations in earnings did increase (Levine and Mazumder, 2007). We further noted that causes of geographic mobility may also generate transitional gains in intergenerational mobility, as is possibly relevant in settings in which both forms of mobility are high (as in Long and Ferrie, 2013).

Promising avenues for future research include the observation that different causes of mobility shifts or different transmission models could be distinguished by their divergent dynamic implications, or that the effect of past events on current mobility trends could be detected by conditioning mobility measures on parental age at birth. Perhaps the most immediate implication of our work is that the covariance between income, skills and endowments in the *parent* generation should be a key object of interest in mobility studies, as it plays a central role for the evolution of income mobility over cohorts and generations.

Data Availability

Codes replicating the tables and figures in this article can be found in the Harvard Dataverse (https://doi.org/10.7910/DVN/OQROKM; Nybom and Stuhler, 2023). The project pulls from two data sources: survey data from the Panel Study of Income Dynamics (PSID) and restricted access data from Swedish administrative registers. The replication package provides all codes used to generate the results and instructions for how to obtain the source PSID data and the restricted access data from Swedish administrative registers.

References

AARONSON, D., AND B. MAZUMDER (2008): "Intergenerational Economic Mobility in the United States, 1940 to 2000," *Journal of Human Resources*, 43(1).

ATKINSON, A. B., AND S. P. JENKINS (1984): "The Steady-State Assumption and the Estimation of Distributional and Related Models," *Journal of Human Resources*, 19(3), 358–376.

ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49(1), 3–71.

AUTOR, D., AND L. KATZ (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3, pp. 1463–1555. North-Holland.

AUTOR, D., F. LEVY, AND R. MURNANE (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118(4), 1279–1333.

BECKER, G. S., AND N. TOMES (1979): "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87(6), 1153–1189.

——— (1986): "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4(3), 1–39.

BÉNABOU, R., AND E. A. OK (2001): "Social Mobility And The Demand For Redistribution: The Poum Hypothesis," *The Quarterly Journal of Economics*, 116(2), 447–487.

BJÖRKLUND, A., M. JÄNTTI, AND M. LINDQUIST (2009): "Family Background and Income During the Rise of the Welfare State: Brother Correlations in Income for Swedish Men born 1932-1968," *Journal of Public Economics*, 93(5-6), 671–680.

BLACK, S. E., AND P. DEVEREUX (2011): "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4A. Elsevier.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005): "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*, 95(1), 437–449.

BLANDEN, J., P. GREGG, AND L. MACMILLAN (2013): "Intergenerational persistence in income and social class: the effect of within-group inequality," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 541–563.

BOWLES, S., AND H. GINTIS (2002): “The Inheritance of Inequality,” *Journal of Economic Perspectives*, 16(3), 3–30.

BREEN, R. (2004): “Social Mobility in Europe between 1970 and 2000,” in *Social Mobility in Europe*, ed. by R. Breen, pp. 37–75. Oxford University Press.

CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014a): “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *Quarterly Journal of Economics*, 129(4), 1553–1623.

CHETTY, R., N. HENDREN, P. KLINE, E. SAEZ, AND N. TURNER (2014b): “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility,” *American Economics Review: Papers and Proceedings*, 104(5), 141–47.

COMERFORD, D., J. RODRÍGUEZ MORA, AND M. J. WATTS (2022): “Meritocracy and the Inheritance of Advantage,” *Journal of Economic Growth*, 27(2), 235–272.

CONLISK, J. (1969): “An Approach to the Theory of Inequality in the Size Distribution of Income,” *Economic Inquiry*, 7(2), 180–186.

——— (1974a): “Can Equalization of Opportunity Reduce Social Mobility?,” *The American Economic Review*, 64(1), pp. 80–90.

——— (1974b): “Stability in a Random Coefficient Model,” *International Economic Review*, 15(2), pp. 529–533.

CONNOLLY, M., M. CORAK, AND C. HAECK (2019): “Intergenerational Mobility Between and Within Canada and the United States,” *Journal of Labor Economics*, 37(S2), S595–S641.

CORAK, M. (2013): “Income Inequality, Equality of Opportunity, and Intergenerational Mobility,” *Journal of Economic Perspectives*, 27(3), 79–102.

DAVIES, J. B., J. ZHANG, AND J. ZENG (2005): “Intergenerational Mobility under Private vs. Public Education,” *Scandinavian Journal of Economics*, 107(3), 399–417.

DAVIS, J. M., AND B. MAZUMDER (2020): “The Decline in Intergenerational Mobility After 1980,” Working Paper Series 11, Stone Center on Socio-Economic Inequality.

DEMING, D. J. (2017): “The Growing Importance of Social Skills in the Labor Market,” *The Quarterly Journal of Economics*, 132(4), 1593–1640.

DURLAUF, S. N., A. KOURTELLOS, AND C. M. TAN (2021): “The Great Gatsby Curve,” Working Paper Series 43, Stone Center on Socio-Economic Inequality.

EDIN, P.-A., P. FREDRIKSSON, M. NYBOM, AND B. ÖCKERT (2021): "The Rising Return to Non-cognitive Skill," *American Economic Journal: Applied Economics*, Advance online publication.

ERIKSON, R., AND J. GOLDTHORPE (1992): *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford University Press, USA.

ERIKSON, R., AND J. O. JONSSON (1996): "Introduction: Explaining Class Inequality in Education: The Swedish Test Case," in *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, ed. by R. Erikson, and J. Jonsson, pp. 1–64. Boulder, Colorado: Westview Press.

GALOR, O., AND D. TSIDDON (1997): "Technological Progress, Mobility, and Economic Growth," *American Economic Review*, 87(3), 363–82.

GOLDBERGER, A. S. (1989): "Economic and Mechanical Models of Intergenerational Transmission," *The American Economic Review*, 79(3), pp. 504–513.

GRÖNQVIST, E., B. ÖCKERT, AND J. VLACHOS (2017): "The Intergenerational Transmission of Cognitive and Noncognitive Abilities," *Journal of Human Resources*, 52(4), 887–918.

GÜELL, M., M. PELLIZZARI, G. PICA, AND J. V. R. MORA (2018): "Correlating Social Mobility and Economic Outcomes," *The Economic Journal*, 128(612), F353–F403.

HAIDER, S., AND G. SOLON (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 96(4), 1308–1320.

HANUSHEK, E. A. (1974): "Efficient Estimators for Regressing Regression Coefficients," *The American Statistician*, 28(2), 66–67.

HASSLER, J., J. R. MORA, AND J. ZEIRA (2007): "Inequality and Mobility," *Journal of Economic Growth*, 12(3), 235–259.

HAUSER, R. M. (2010): "Intergenerational Economic Mobility in the United States Measures, Differentials and Trends," Cde working paper no. 98-12, CDE.

HECKMAN, J. J. (1995): "Lessons from the Bell Curve," *Journal of Political Economy*, 103(5), 1091–1120.

HECKMAN, J. J., AND S. MOSSO (2014): "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 6, 689–733.

HERTZ, T. (2007): "Trends in the Intergenerational Elasticity of Family Income in the United States," *Industrial Relations*, 46(1), 22–50.

——— (2008): "A Group-specific Measure of Intergenerational Persistence," *Economics Letters*, 100(3), 415–417.

HERTZ, T., T. JAYASUNDERA, P. PIRAINO, S. SELCUK, N. SMITH, AND A. VERASHCHAGINA (2008): "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends," *The B.E. Journal of Economic Analysis & Policy*, 7(2), 1–48.

HILGER, N. G. (2015): "The Great Escape: Intergenerational Mobility Since 1940," Working Paper 21217, National Bureau of Economic Research.

HOLMLUND, H. (2007): "A Researcher's Guide to the Swedish Compulsory School Reform," Discussion paper, SOFI.

——— (2008): "Intergenerational Mobility and Assortative Mating: Effects of an Educational Reform," Cee discussion paper 91, Centre for the Economics of Education.

HOLMLUND, H., M. LINDAHL, AND E. PLUG (2011): "The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods," *Journal of Economic Literature*, 49(3), 615–51.

JÁCOME, E., I. KUZIEMKO, AND S. NAIDU (2021): "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century," Discussion paper, National Bureau of Economic Research, Working Paper No. 29289.

JENKINS, S. (1987): "Snapshots versus Movies: 'Lifecycle biases' and the Estimation of Intergenerational Earnings Inheritance," *European Economic Review*, 31(5), 1149–1158.

JENKINS, S. P. (1982): "Tools for the Analysis of Distributional Models," *The Manchester School of Economic & Social Studies*, 50(2), 139–50.

JUSTMAN, M., AND H. STIASSNIE (2021): "Intergenerational Mobility in Lifetime Income," *Review of Income and Wealth*, n/a(n/a).

KARLSON, K. B., AND R. LANDERSØ (2021): "The Making and Unmaking of Opportunity: Educational Mobility in 20th Century-Denmark," IZA Discussion Papers 14135, Institute for the Study of Labor (IZA).

KATZ, L. F., AND K. M. MURPHY (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

LANDERSØ, R., AND J. J. HECKMAN (2017): "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US," *The Scandinavian Journal of Economics*, 119(1), 178–230.

LEE, C.-I., AND G. SOLON (2009): "Trends in Intergenerational Income Mobility," *The Review of Economics and Statistics*, 91(4), 766–772.

LEFGREN, L., M. J. LINDQUIST, AND D. SIMS (2012): "Rich Dad, Smart Dad: Decomposing the Intergenerational Transmission of Income," *Journal of Political Economy*, 120(2), 268 – 303.

LEVINE, D. I., AND B. MAZUMDER (2007): "The Growing Importance of Family: Evidence from Brothers' Earnings," *Industrial Relations: A Journal of Economy and Society*, 46(1), 7–21.

LINDQVIST, E., AND R. VESTMAN (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3(1), 101–128.

LOEHLIN, J. C. (2005): "Resemblance in Personality and Attitudes Between Parents and their Children," in *Unequal chances: Family background and economic success*, ed. by S. Bowles, H. Gintis, and M. O. Groves, chap. 6, pp. 192–207. Princeton University Press.

LONG, J., AND J. FERRIE (2013): "Intergenerational Occupational Mobility in Great Britain and the United States since 1850," *American Economic Review*, 103(4), 1109–37.

MACHIN, S. (2007): "Education Expansion and Intergenerational Mobility in Britain," in *Schools and the Equal Opportunity Problem*, ed. by L. Woessman, and P. Peterson. MIT Press.

MARKUSSEN, S., AND K. RØED (2020): "Economic mobility under pressure," *Journal of the European Economic Association*, 18(4), 1844–1885.

MEGHIR, C., AND M. PALME (2005): "Educational Reform, Ability, and Family Background," *American Economic Review*, 95(1), 414–424.

MEGHIR, C., M. PALME, AND M. SCHNABEL (2011): "The Effect of Education Policy on Crime: An Intergenerational Perspective," IZA Discussion Papers 6142, Institute for the Study of Labor (IZA).

MODALSLI, J. (2017): "Intergenerational Mobility in Norway, 1865–2011," *The Scandinavian Journal of Economics*, 119(1), 34–71.

MOGSTAD, M., AND G. TORSVIK (2021): "Family Background, Neighborhoods and Intergenerational Mobility," Discussion Paper w28874, NBER Working Paper.

NICOLETTI, C., AND J. ERMISCH (2007): "Intergenerational Earnings Mobility: Changes across Cohorts in Britain," *The B.E. Journal of Economic Analysis & Policy*, 7(2), Article 9.

NYBOM, M., AND J. STUHLER (2014): "Interpreting Trends in Intergenerational Mobility," Working Paper Series 3/2014, Swedish Institute for Social Research.

——— (2016): "Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation," *Journal of Human Resources*, 51(1).

——— (2023): "Replication Data for: "Interpreting trends in intergenerational mobility"," .

PALOMINO, J. C., G. A. MARRERO, AND J. G. RODRÍGUEZ (2017): "One size doesn't fit all: a quantile analysis of intergenerational income mobility in the U.S. (1980–2010)," *The Journal of Economic Inequality*.

PEKKALA, S., AND R. E. B. LUCAS (2007): "Differences across Cohorts in Finnish Intergenerational Income Mobility," *Industrial Relations: A Journal of Economy and Society*, 46(1), 81–111.

PEKKARINEN, T., K. G. SALVANES, AND M. SARVIMÄKI (2017): "The Evolution of Social Mobility: Norway during the Twentieth Century," *The Scandinavian Journal of Economics*, 119(1), 5–33.

PEKKARINEN, T., R. UUSITALO, AND S. KERR (2009): "School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform," *Journal of Public Economics*, 93(7-8), 965–973.

SOLON, G. (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3A, chap. 29, pp. 1761–1800. Elsevier.

SOLON, G. (2002): "Cross-Country Differences in Intergenerational Earnings Mobility," *Journal of Economic Perspectives*, 16(3), 59–66.

SOLON, G. (2004): "A Model of Intergenerational Mobility Variation over Time and Place," *Generational Income Mobility in North America and Europe*, pp. 38–47.

WORKING, H. (1960): "Note on the Correlation of First Differences of Averages in a Random Chain," *Econometrica*, 28(4), 916–918.

A Online Appendix

A.1 An Economic Model of Intergenerational Transmission

We model the optimizing behavior of parents to derive the "mechanical" transmission equations presented in Section 1. For this purpose we extend the model in Solon (2004), considering parental investments in multiple distinct types of human capital and statistical discrimination on the labor market.

Assume that parents allocate their lifetime *after tax earnings* $(1-\tau)Y_{t-1}$ between own *consumption* C_{t-1} and *investments* $I_{1,t-1},...,I_{J,t-1}$ in J distinctive types of human capital of their children. Parents do not bequeath financial assets and face the budget constraint

$$(1-\tau)\,Y_{t-1} = C_{t-1} + \sum_{j=1}^{J} I_{j,t-1}. \tag{A.1}$$

Accumulation of *human capital* h of type j in offspring generation t depends on parental investment, a $Kx1$ vector of inherited *endowments* $\boldsymbol{e}_t$, and chance $u_{j,t}$,

$$h_{j,t} = \gamma_j logI_{j,t-1} + \boldsymbol{\theta}_j'\boldsymbol{e}_t + u_{j,t} \;\; \forall j \in 1,...,J, \tag{A.2}$$

where γ_j and elements of the vector $\boldsymbol{\theta}_j$ measure the marginal product of parental investment and each endowment. Endowments represent early child attributes that may be influenced by nature (genetic inheritance) or nurture (e.g. parental upbringing). We assume that they are positively correlated between parents and their children, as implied by the autoregressive process

$$e_{k,t} = \lambda_k e_{k,t-1} + v_{k,t} \;\; \forall k \in 1,...,K, \tag{A.3}$$

where $v_{k,t}$ is a white-noise error term and the heritability coefficient λ_k lies between 0 and 1. We may allow endowments to be correlated within individuals, leading to the more general transmission equation (4). Finally, assume that income of offspring equals

$$logY_t = \begin{cases} \boldsymbol{\delta}'\boldsymbol{h}_t + u_{y,t} & \text{with probability p} \\ \boldsymbol{\delta}'E\left[\boldsymbol{h}_t|Y_{t-1}\right] + u_{y,t} & \text{with probability } 1-p \end{cases}. \tag{A.4}$$

With probability p employers observe human capital of workers and pay them their marginal product $\boldsymbol{\delta}'\boldsymbol{h}_t$ plus a white-noise error term $u_{y,t}$, which reflects market luck. With probability $1-p$ employers cannot uncover true productivity, and remunerate workers instead for their *expected* productivity given observed parental background. In particular, employers observe that on average parents invest income share s_j in offspring human capital of type j, such that $E[I_{j,t-1}|Y_{t-1}] = s_j Y_{t-1}$, and that the offspring of high-income parents tend to have more favorable endowments, such that $E[e_{k,t}|Y_{t-1}] = \gamma_k Y_{t-1}$ (with $\gamma_k \geq 0$) for all $k \in 1, ..., K$.

Parents choose investment in the child's human capital as to maximize the utility function

$$U_{t-1} = (1-\alpha)\, logC_{t-1} + \alpha E\left[logY_t | Y_{t-1}, \boldsymbol{I}_{t-1}, \boldsymbol{e}_t\right], \tag{A.5}$$

where the altruism parameter $\alpha \in [0,1]$ measures the parent's taste for own consumption relative to the child's expected income. Given equations (A.1) to (A.5), the Lagrangian for parent's investment decision is

$$\begin{aligned} \mathcal{L}(C_{t-1}, \boldsymbol{I}_{t-1}, \mu) \quad = \quad & (1-\alpha) logC_{t-1} + \alpha \boldsymbol{\delta}' \left(pE\left[\boldsymbol{h}_t | Y_{t-1}, \boldsymbol{I}_{t-1}, \boldsymbol{e}_t\right] + (1-p) E\left[\boldsymbol{h}_t | Y_{t-1}\right]\right) \\ & + \mu \left((1-\tau) Y_{t-1} - C_{t-1} - \mathbf{1}' \boldsymbol{I}_{t-1}\right) \end{aligned}$$

The first-order conditions require that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial C_{t-1}} &= \tfrac{1-\alpha}{C_{t-1}} - \mu \quad = 0, \\ \frac{\partial \mathcal{L}}{\partial I_{j,t-1}} &= \tfrac{\alpha(1-p)\delta_j \gamma_j}{I_{j,t-1}} - \mu \quad = 0 \ \ \forall j \in 1, ..., J, \\ \frac{\partial \mathcal{L}}{\partial \mu} &= (1-\tau) Y_{t-1} - C_{t-1} - \mathbf{1}' \boldsymbol{I}_{t-1} \quad = 0. \end{aligned}$$

Optimal investments,

$$I_{j,t-1} = \frac{\alpha p \delta_j \gamma_j}{(1-\alpha) + \sum_{l=1}^{J} \alpha p \delta_l \gamma_l} (1-\tau) Y_{t-1} \quad \forall j \in 1, ..., J, \tag{A.6}$$

increase in parental altruism and income, and in the probability that offspring human capital is observed and acted on by employers. Parents invest more into those skills in which the marginal product of investment or the return on the labor market are large. Plugging optimal investment into equation (A.2) yields (ignoring constants, which are irrelevant for our analysis) equation (3), which if plugged in turn into equation (A.4) motivates equation (2).

A.2 Reduced Form and Stability

The reduced form of equations (5) and (6) is

$$\begin{pmatrix} y_t \\ \boldsymbol{e}_t \end{pmatrix} = \begin{pmatrix} \gamma_t & \boldsymbol{\rho}_t' \boldsymbol{\Lambda}_t \\ \mathbf{0} & \boldsymbol{\Lambda}_t \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \boldsymbol{e}_{t-1} \end{pmatrix} + \begin{pmatrix} \sigma_t u_t + \boldsymbol{\rho}_t' \boldsymbol{\Phi}_t \boldsymbol{v}_t \\ \boldsymbol{v}_t \end{pmatrix}, \tag{A.7}$$

which we may shorten to

$$\boldsymbol{x}_t = \boldsymbol{A}_t \boldsymbol{x}_{t-1} + \boldsymbol{w}_t. \tag{A.8}$$

Let subscripts $1, 2$ index parameter values before and after a structural shock occurs in generation T.[34] The stability condition $\lim_{s\to\infty} \boldsymbol{A}_2^s = 0$ is then satisfied by assuming that γ_t and all eigenvalues of $\boldsymbol{\Lambda}_2$ are non-negative and below one. For example, if $\boldsymbol{\Lambda}_2$ is diagonal and elements of the endowment vector $\boldsymbol{e}_t$ are uncorrelated then the diagonal elements of $\boldsymbol{\Lambda}_2$ are required to be strictly between zero and one. Normalization of the variances of y_t and elements of $\boldsymbol{e}_t$ in the initial steady state leads to additional parameter restrictions. Take the covariance of (A.8) and denote the covariance matrices of x_t and w_t by $\boldsymbol{S}_t$ and $\boldsymbol{W}_t$, such that

$$\boldsymbol{S}_t = \boldsymbol{A}_t \boldsymbol{S}_{t-1} \boldsymbol{A}_t' + \boldsymbol{W}_t.$$

Denote by γ, $\boldsymbol{\rho}$, and $\boldsymbol{\Lambda}$ the steady-state parameter values before a structural change occurs in generation $t = T$. Note that in steady state $\boldsymbol{S}_t = \boldsymbol{S}_{t-1} = \boldsymbol{S}$, normalize all diagonal elements of $\boldsymbol{S}$ to one, and solve for the elements of $\boldsymbol{W}_t$. For example, if $\boldsymbol{\Lambda}_t$ is diagonal then $Var(e_{j,t}) = 1\ \forall j$ iff $Var(v_{j,t}) = 1 - \lambda_j^2\ \forall j$; the variances are non-negative iff $\lambda_{jj} \leq 1\ \forall j$, as is also required for stability of the system.

A.3 Overview of Comparative Transitional Dynamics

The change in the intergenerational elasticity can be expressed as

$$\begin{aligned}\Delta\beta_t &= \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} - \frac{Cov(y_{t-1}, y_{t-2})}{Var(y_{t-2})} \\ &= \frac{\Delta Cov(y_t, y_{t-1})Var(y_{t-1}) - \Delta Var(y_{t-1})Cov(y_t, y_{t-1})}{Var(y_{t-1})Var(y_{t-2})}\end{aligned} \tag{A.9}$$

To track the dynamics of the IGE, we thus need to track the transition paths of its constituent moments. Plugging in equations (5) and (6) and using notation such as $\rho_1 = \rho_{t<T}$ and $\rho_2 = \rho_{t\geq T}$ to denote model parameters before and after the structural change, the shifts in the IGE in the first two affected generations can be expressed as

$$\triangle\beta_T = (\gamma_2 - \gamma_1) + (\rho_2\lambda_2 - \rho_1\lambda_1)Cov(e_{T-1}, y_{T-1}) \tag{A.10}$$

$$\triangle\beta_{T+1} = \rho_2\lambda_2 \left(\frac{Cov(e_T, y_T)}{Var(y_T)} - \frac{Cov(e_{T-1}, y_{T-1})}{Var(y_{T-1})} \right) \tag{A.11}$$

where $Var(y_{T-1})$ is standardized to one and $Cov(e_{T-1}, y_{t-1}) = \frac{\rho_1}{1-\gamma_1\lambda_1}$. Table A.1 provides case-specific expressions for $\Delta Cov(y_T, y_{T-1})$, $\Delta Var(y_T)$, $\Delta Cov(\boldsymbol{e}_T, y_T)$, and $\Delta Var(\boldsymbol{e}_T)$.

[34] Conlisk (1974b) derives stability conditions in a random coefficients model with repeated shocks.

Table A.1: Transitional Dynamics in Components of the IGE

	$\Delta Cov(y_T, y_{T-1})$	$\Delta Var(y_T)$	$\Delta Cov(e_T, y_T)$	$\Delta Var(e_T)$
Scalar Model:				
$\sigma_1 \to \sigma_2$	0	$\sigma_2^2 - \sigma_1^2$	0	0
$\Phi_1 \to \Phi_2$	0	$\rho^2 \Delta Var(e_T)$	$\rho \Delta Var(e_T)$	$\Phi_2^2 - \Phi_1^2$
$\rho_1 \to \rho_2$	$\frac{(\rho_2 - \rho_1)\lambda\rho_1}{1-\gamma\lambda}$	$\left(\rho_2^2 - \rho_1^2\right) + \frac{2\gamma(\rho_2 - \rho_1)\lambda\rho_1}{1-\gamma\lambda}$	$\rho_2 - \rho_1$	0
$\lambda_1 \to \lambda_2$	$\frac{(\lambda_2 - \lambda_1)\rho^2}{1-\gamma\lambda_1}$	$\rho^2 \Delta Var(e_T)$ $+ 2\gamma(\lambda_2 - \lambda_1)\rho^2/(1-\gamma\lambda_1)$	$\gamma(\lambda_2 - \lambda_1)\rho/(1-\gamma\lambda_1)$ $+ \rho\Delta Var(e_T)$	$\lambda_2^2 - \lambda_1^2$
$\gamma_1 \to \gamma_2$	$\gamma_2 - \gamma_1$	$\left(\gamma_2^2 - \gamma_1^2\right) + \frac{2(\gamma_2 - \gamma_1)\lambda\rho^2}{1-\gamma_1\lambda}$	$\frac{(\gamma_2 - \gamma_1)\lambda\rho}{1-\gamma_1\lambda}$	0
$\gamma_1 \to \gamma_2$; $\rho_1 \to \rho_2$	$(\gamma_2 - \gamma_1) + \frac{(\rho_2 - \rho_1)\lambda\rho_1}{1-\gamma_1\lambda}$	$\left(\gamma_2^2 - \gamma_1^2\right) + \left(\rho_2^2 - \rho_1^2\right)$ $+ \frac{2\left(\gamma_2\rho_2 - \gamma_1\rho_1\right)\lambda\rho_1}{1-\gamma_1\lambda}$	$(\rho_2 - \rho_1) + \frac{(\gamma_2 - \gamma_1)\lambda\rho_1}{1-\gamma_1\lambda}$	0
Multi-skill Model:				
$\boldsymbol{\rho}_1 \to \boldsymbol{\rho}_2$	$(\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')\boldsymbol{\Lambda}\left(\boldsymbol{I} - \gamma\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\rho}_1$	$(\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')\left(\boldsymbol{\rho}_2 + \boldsymbol{\rho}_1\right)$ $+ 2\gamma\left(\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1'\right)\boldsymbol{\Lambda}\left(\boldsymbol{I} - \gamma\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\rho}_1$	$(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1)Var(\boldsymbol{e}_{T-1})$	0
$\boldsymbol{\Lambda}_1 \to \boldsymbol{\Lambda}_2$	$\boldsymbol{\rho}'(\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1)\left(\boldsymbol{I} - \gamma\boldsymbol{\Lambda}_1\right)^{-1}\boldsymbol{\rho}_1$	$\boldsymbol{\rho}'\Delta Var(\boldsymbol{e}_T)\boldsymbol{\rho}$ $+ 2\gamma\boldsymbol{\rho}'(\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1)\left(\boldsymbol{I} - \gamma\boldsymbol{\Lambda}_1\right)^{-1}\boldsymbol{\rho}_1$	$\gamma(\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1)\left(\boldsymbol{I} - \gamma\boldsymbol{\Lambda}_1\right)^{-1}\boldsymbol{\rho}_1$ $+\boldsymbol{\rho}\Delta Var(\boldsymbol{e}_T)$	$(\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1)(\boldsymbol{\Lambda}_2 + \boldsymbol{\Lambda}_1)$

Note: The table reports the change in intergenerational and cross-sectional moments in the first two generations after a specific structural change (left column) occurs in generation T.

Assuming that the environment remains stable after generation T, the transition paths for the periods from $t = T + 1$ onwards can be defined for all cases by the recursive processes

$$\Delta Cov(y_t, y_{t-1}) = \gamma_2 \Delta Var(y_{t-1}) + \boldsymbol{\rho}_2' \boldsymbol{\Lambda}_2 \Delta Cov(\boldsymbol{e}_{t-1}, y_{t-1}) \tag{A.12}$$

$$\Delta Cov(\boldsymbol{e}_t, y_t) = \gamma_2 \boldsymbol{\Lambda}_2 \Delta Cov(\boldsymbol{e}_{t-1}, y_{t-1}) + \Delta Var(\boldsymbol{e}_t) \boldsymbol{\rho}_2 \tag{A.13}$$

$$\Delta Var(y_t) = \gamma_2^2 \Delta Var(y_{t-1}) + \boldsymbol{\rho}_2' \Delta Var(\boldsymbol{e}_t) \boldsymbol{\rho}_2 + 2\gamma_2 \boldsymbol{\rho}_2' \boldsymbol{\Lambda}_2 \Delta Cov(\boldsymbol{e}_{t-1}, y_{t-1}) \tag{A.14}$$

$$\Delta Var(\boldsymbol{e}_t) = \boldsymbol{\Lambda}_2 \Delta Var(\boldsymbol{e}_{t-1}) \boldsymbol{\Lambda}_2. \tag{A.15}$$

By plugging in the moments for period T from Table A.1 below into equation (A.9) and equations (A.12) to (A.15) we can therefore derive the complete transition path of the IGE after a permanent change in the economic environment from $\boldsymbol{\xi}_{t<T} = \boldsymbol{\xi}_1$ to $\boldsymbol{\xi}_{t \geq T} = \boldsymbol{\xi}_2$.

A.4 Proofs of Propositions

Below we provide proofs to the propositions in Section 2. When possible, we provide general conditions; for others, general conditions become too abstract to be useful, given the array of structural changes, generations, and sets of initial conditions that are necessary to consider. In these cases, we provide case distinctions instead. In addition, we perform numerical analyses that computes the impulse responses across all *feasible* combinations of initial conditions and parameter changes in a discretized parameter space. Thus, we consider the dynamics of the IGE for elements of $\boldsymbol{\xi}_{t<T} = \boldsymbol{\xi}_1 = \{\gamma_1, \rho_1, \sigma_1, \lambda_1, \Phi_1\}$ following one-time parameter changes in each of the five parameters (e.g. $\rho_2 > \rho_1$) on a case-by-case basis. The five-dimensional joint distribution in $\boldsymbol{\xi}_1$ is uniform in increments of 0.01 between 0 and 1. The "feasible" subspace of $\boldsymbol{\xi}_1$ is defined by stability conditions (see Appendix A.2), which implies $0 \leq \beta_{T-1} \leq 1$. We further impose the normalization $Var(y_{T-1}) = Var(e_{T-1}) = 1$ (which implicitly determines σ_1 and Φ_1 as a function of the other parameters). These simulations serve two purposes. First, they provide a useful verification of our analytical results. Second, they enable us to gauge whether a condition is "likely" to hold (assuming a uniform prior regarding the likely distribution of initial conditions).

A.4.1 Proposition 1

Proof. Starting from equation (7), the steady-state elasticity in the scalar model with a single skill can be expressed as

$$\begin{aligned}
\beta &= \gamma + \frac{\rho^2 \lambda}{1 - \gamma\lambda} \frac{Var(e)}{Var(y)} \\
&= \gamma + \frac{\rho^2 \lambda \Phi^2 (1 - \gamma^2)}{\rho^2 \Phi^2 (1 + \gamma\lambda) + \sigma^2 (1 - \lambda^2)(1 - \gamma\lambda)}
\end{aligned}$$

where in the first row we substituted the steady-state expression for the covariance of endowment and income, $Cov(e,y) = \frac{\rho Var(e)}{(1-\gamma\lambda)}$, and in the second row we substituted the steady-state variances of endowment and income, $Var(e) = \frac{\Phi^2}{1-\lambda^2}$ and $Var(y) = \gamma^2 Var(y) + \rho^2 Var(e) + 2\gamma\rho\lambda Cov(e,y) + \sigma^2 = \frac{\rho^2\Phi^2(1+\gamma\lambda)+\sigma^2(1-\lambda^2)(1-\gamma\lambda)}{(1-\gamma^2)(1-\lambda^2)(1-\gamma\lambda)}$, and thus $\frac{Var(e)}{Var(y)} = \frac{\Phi^2(1-\gamma^2)(1-\gamma\lambda)}{\rho^2\Phi^2(1+\gamma\lambda)+\sigma^2(1-\lambda^2)(1-\gamma\lambda)}$. Re-expressing the steady-state elasticity as

$$\beta = \frac{\rho^2\Phi^2(\gamma+\lambda) + \gamma\sigma^2(1-\lambda^2)(1-\gamma\lambda)}{\rho^2\Phi^2(1+\gamma\lambda) + \sigma^2(1-\lambda^2)(1-\gamma\lambda)} = \frac{g(x)}{h(x)},$$

its derivative with respect to the variance of market luck is

$$\frac{\partial\beta}{\partial\sigma^2} = -\frac{\rho^2\lambda\Phi^2(1-\gamma^2)(1-\lambda^2)(1-\gamma\lambda)}{h(x)^2} \leq 0,$$

which is strictly negative if $\rho > 0$, $\lambda > 0$, and $\Phi > 0$. The derivative with respect to endowment luck is

$$\frac{\partial\beta}{\partial\Phi^2} = \frac{\sigma^2\rho^2\lambda(1-\gamma^2)(1-\lambda^2)(1-\gamma\lambda)}{h(x)^2} \geq 0,$$

which is strictly positive if $\rho > 0$, $\lambda > 0$, and $\sigma > 0$. The derivative with respect to γ is

$$\frac{\partial\beta}{\partial\gamma} = \frac{(1-\lambda^2)\left(p^2\Phi^2\sigma^2\left(2(1-\gamma\lambda)+\lambda^2(1-\gamma^2)\right) + (1-\lambda^2)\,\sigma^2(\gamma\lambda-1)^2 + p^4\Phi^2\right)}{h(x)^2} \geq 0,$$

which is strictly positive if either $\sigma > 0$ or if $\Phi > 0$ and $\rho > 0$. The derivative with respect to ρ is

$$\frac{\partial\beta}{\partial\rho} = \frac{2\rho\lambda\sigma^2\Phi^2\left(1-\gamma^2\right)\left(1-\lambda^2\right)\left(1-\gamma\lambda\right)}{h(x)^2} \geq 0,$$

which is strictly positive if $\rho > 0$, $\lambda > 0$, $\Phi > 0$ and $\sigma > 0$. The derivative with respect to λ is

$$\frac{\partial\beta}{\partial\lambda} = \frac{\rho^2\Phi^2(1-\gamma^2)\left(\rho^2\Phi^2 + \sigma^2(1+\lambda^2-2\gamma\lambda^3)\right)}{h(x)^2} \geq 0,$$

which is strictly positive if $\rho > 0$ and $\Phi > 0$.

However, in the model with multiple skills, an increase in the return to one skill can decrease the steady-state elasticity. To see this, consider a model with two uncorrelated skills k and l, such that $\rho_{k,t\geq T} = \rho_{k,2} > \rho_{k,t<T} = \rho_{k,1}$, $\rho_{l,1} = \rho_{l,2} = \rho_l$ and $Var(e_{kt}) = Var(e_{lt}) = 1$ $\forall t$. In this model the old steady state β_{T-1} and new steady state β_∞ are given by

$$\beta_{T-1} = \gamma + \frac{\rho_{k,1}^2\lambda_k}{1-\gamma\lambda_k} + \frac{\rho_l^2\lambda_l}{1-\gamma\lambda_l}$$

$$\beta_\infty = \gamma + \left(\frac{\rho_{k,2}^2\lambda_k}{1-\gamma\lambda_k} + \frac{\rho_l^2\lambda_l}{1-\gamma\lambda_l}\right)\frac{1}{1+\triangle Var(y_\infty)},$$

where $\triangle Var(y_\infty) = Var(y_\infty) - Var(y_{T-1}) = \frac{(\rho_{k,2}^2 - \rho_{k,1}^2)(1+\gamma\lambda_k)}{(1-\gamma\lambda_k)(1-\gamma^2)}$. The condition for $\beta_{T-1} > \beta_\infty$ is therefore

$$\begin{aligned}\beta_{T-1} &> \beta_\infty \\ \gamma + \frac{\rho_{k,1}^2\lambda_k}{1-\gamma\lambda_k} + \frac{\rho_l^2\lambda_l}{1-\gamma\lambda_l} &> \gamma + \left(\frac{\rho_{k,2}^2\lambda_k}{1-\gamma\lambda_k} + \frac{\rho_l^2\lambda_l}{1-\gamma\lambda_l}\right)\frac{1}{1+\triangle Var(y_\infty)} \\ \left(\frac{\rho_{k,1}^2\lambda_k}{1-\gamma\lambda_k} + \frac{\rho_l^2\lambda_l}{1-\gamma\lambda_l}\right)\frac{(\rho_{k,2}^2 - \rho_{k,1}^2)(1+\gamma\lambda_k)}{(1-\gamma\lambda_k)(1-\gamma^2)} &> \frac{\left(\rho_{k,2}^2 - \rho_{k,1}^2\right)\lambda_k}{1-\gamma\lambda_k} \\ \frac{\rho_{k,1}^2\lambda_k}{1-\gamma\lambda_k} + \frac{\rho_l^2\lambda_l}{1-\gamma\lambda_l} &> \frac{(1-\gamma^2)\lambda_k}{1+\gamma\lambda_k} \end{aligned} \tag{A.16}$$

where we substituted in the expression for $\triangle Var(y_\infty)$ in the third line. It is straightforward to see that this condition holds if the heritability of endowment k is sufficiently low compared to other determinants of the IGE. Figure 2 provides an illustration. For example, if $\gamma = 0$ condition (A.16) reduces to $\rho_l^2\lambda_l > (1-\rho_{k,1}^2)\lambda_k$, which can hold if λ_k is sufficiently low compared to λ_l. Note further that this condition cannot hold if endowment l is not rewarded ($\rho_l = 0$) or not inherited ($\lambda_l = 0$), i.e. the condition can hold in a multi-skill but not the single skill model. □

A.4.2 Proposition 2

Proof. (a) We first derive the conditions under which the IGE shifts in generation $T+1$, i.e. $\triangle\beta_{T+1} \neq 0$. Note that from equation (A.11) in Appendix A.3 it follows that $\triangle\beta_{T+1}$ is non-zero iff $\rho_2 > 0$, $\lambda_2 > 0$ and $\frac{Cov(e_T,y_T)}{Cov(e_{t-1},y_{T-1})} \neq \frac{Var(y_T)}{Var(y_{T-1})}$ or equivalently,

$$\frac{\Delta Cov(e_T, y_T)}{Cov(e_{t-1}, y_{T-1})} \neq \frac{\Delta Var(y_T)}{Var(y_{T-1})}, \tag{A.17}$$

i.e. the percentage changes in the covariance between income and endowments and the variance of income are not exactly equal.[35] This condition always holds for changes in σ^2 or Φ^2, and holds for changes in γ, ρ and λ except for knife-edge cases. Specifically, from Table A.1 it follows that

Case 1. For a change in σ^2, $\Delta Cov(e_T, y_T) = 0$, $\Delta Var(y_T) = \sigma_2^2 - \sigma_1^2$ and $\Delta\beta_{T+1} = -\frac{\lambda\rho^2}{1-\gamma\lambda}\frac{\sigma_2^2-\sigma_1^2}{1+\sigma_2^2-\sigma_1^2} < 0$.

[35] After some covariance iterations and using that $Var(e_T) = \lambda_2^2 Var(e_{T-1}) + \Phi_2^2 = \lambda_2^2 + \Phi_2^2$ and $Cov(e_{T-1}, y_{T-1}) = \rho_1/(1-\lambda_1\gamma_1)$ this inequality condition can be expressed as

$$\frac{\rho_2}{\rho_1}\left(\lambda_2^2 + \Phi_2^2\right)(1-\lambda_1\gamma_1) + \lambda_2\gamma_2 \neq \gamma_2^2 + \rho_2^2\left(\lambda_2^2 + \Phi_2^2\right) + \sigma_2^2 + 2\rho_2\lambda_2\gamma_2\rho_1/(1-\lambda_1\gamma_1).$$

Case 2. For a change in Φ^2, $\Delta Cov(e_T, y_T) = \rho(\Phi_2^2 - \Phi_1^2)$, $\Delta Var(y_T) = \rho^2(\Phi_2^2 - \Phi_1^2)$ and $\Delta\beta_{T+1} = \frac{\lambda\rho^2}{1-\gamma\lambda}\frac{(\Phi_2^2-\Phi_1^2)(1-\gamma\lambda-\rho^2)}{1+\rho^2(\Phi_2^2-\Phi_1^2)} > 0$.

Case 3. For a change in γ, condition (A.17) simplifies to $\lambda \neq (\gamma_2 + \gamma_1)\frac{1-\gamma_1\lambda}{1-\gamma_1\lambda-2\rho^2}$.

Case 4. For a change in ρ, condition (A.17) simplifies to $1 - \gamma\lambda \neq \rho_1\rho_2 + \rho_1^2\frac{1+\gamma\lambda}{1-\gamma\lambda}$.

Case 5. For a change in λ, condition (A.17) simplifies to $(\lambda_2 + \lambda_1) \neq \rho(\lambda_2 + \lambda_1) + \frac{(2\rho-1)\gamma}{1-\gamma\lambda_1}$.

Following the same covariance iterations as for equation (A.11), we have for all $k > 1$

$$\Delta\beta_{T+k} = \rho_2\lambda_2\left(\frac{Cov(e_{T+k-1}, y_{T+k-1})}{Var(y_{T+k-1})} - \frac{Cov(e_{T+k-2}, y_{T+k-2})}{Var(y_{T+k-2})}\right) \quad \text{(A.18)}$$

Convergence in infinite time therefore requires $\rho_2 > 0$, $\lambda_2 > 0$ and that either $Cov(e_t, y_t)$ or $Var(y_t)$ converge in infinite time, which in turn requires either $\gamma_2 > 0$ or $\Delta Var(e_T) = \Phi_2^2 - \Phi_1^2 + \lambda_2^2 - \lambda_1^2 \neq 0$ (see Section A.3). □

Proof. (b) By using the expressions (A.10) and (A.11) for $\Delta\beta_T$ and $\Delta\beta_{T+1}$ and substituting the expressions for the underlying moments $Cov(e_T, y_T)$ and $Var(y_T)$, we can show that $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ if

$$\frac{\rho_2\lambda_2\left(\rho_2(\lambda_2^2 + \Phi_2^2)(1 - \gamma_1\lambda_1) + \rho_1\gamma_2\lambda_2\right)}{Var(y_T)} > \rho_1(2\rho_2\lambda_2 - \rho_1\lambda_1) + (1 - \gamma_1\lambda_1)(\gamma_2 - \gamma_1) \quad \text{(A.19)}$$

and $sign(\Delta\beta_{T+1}) = sign(\Delta\beta_T)$, or

$$\frac{\rho_2\lambda_2\left(\rho_2(\lambda_2^2 + \Phi_2^2)(1 - \gamma_1\lambda_1) + \rho_1\gamma_2\lambda_2\right)}{Var(y_T)} < \rho_1^2\lambda_1 - (1 - \gamma_1\lambda_1)(\gamma_2 - \gamma_1) \quad \text{(A.20)}$$

and $sign(\Delta\beta_{T+1}) \neq sign(\Delta\beta_T)$, which is satisfied for some parameter values. Specifically,

Case 1. For a change in σ^2, the condition $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ is trivially satisfied because $\Delta\beta_T = 0$ and β shifts only from generation $T + 1$ onwards, as

$$\Delta\beta_{T+1} = -\frac{\lambda\rho^2}{1 - \gamma\lambda}\frac{\sigma_2^2 - \sigma_1^2}{1 + \sigma_2^2 - \sigma_1^2}.$$

We further have $\Delta\beta_{T+2} = \frac{\gamma^2}{1+(1+\gamma^2)(\sigma_2^2-\sigma_1^2)}\Delta\beta_{T+1}$. Because the multiplying ratio can be larger than one for $\sigma_2^2 < \sigma_1^2$ (e.g., if $\gamma = \frac{1}{2}, \rho = \frac{1}{4}, \lambda = \frac{1}{2}, \sigma_1 = \sqrt{\frac{31}{48}}, \sigma_2 = \frac{1}{10}$), the condition $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ can also be satisfied. For $\sigma_2^2 > \sigma_1^2$, the condition cannot be satisfied.

Case 2. For a change in Φ^2, the condition $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ is trivially satisfied because $\Delta\beta_T = 0$ and

$$\Delta\beta_{T+1} = \frac{\lambda\rho^2}{1-\gamma\lambda}\frac{\left(\Phi_2^2-\Phi_1^2\right)\left(1-\gamma\lambda-\rho^2\right)}{1+\rho^2\left(\Phi_2^2-\Phi_1^2\right)}.$$

We further have $\Delta\beta_{T+2} = -\frac{(\gamma+\lambda)\left(\gamma\rho^2\left(\Phi_2^2-\Phi_1^2\right)(1-\gamma\lambda)+\rho^2(\gamma+\lambda)-\lambda(1-\gamma\lambda)\right)}{(1-\gamma\lambda-\rho^2)\left(\rho^2\left(\Phi_2^2-\Phi_1^2\right)((\gamma+\lambda)^2+1)-1\right)}\Delta\beta_{T+1}$. Because the absolute value of the multiplying ratio can be larger than one (e.g., if $\gamma = \frac{1}{2}, \rho = \frac{1}{8}, \lambda = \frac{7}{8}, \Phi_1 = \sqrt{\frac{15}{64}}, \Phi_2 = 1$), the condition $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ can also be satisfied.

Case 3. For a change in γ, $\Delta\beta_T = \gamma_2 - \gamma_1$ and

$$\Delta\beta_{T+1} = \rho\lambda Cov(e_{T-1}, y_{T-1})\frac{\lambda - (\gamma_2+\gamma_1) - 2\lambda\rho Cov(e_{T-1}, y_{T-1})}{1+\gamma_2^2-\gamma_1^2+2(\gamma_2-\gamma_1)\lambda\rho Cov(e_{T-1}, y_{T-1})}\Delta\beta_T.$$

It follows that $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ cannot hold, as the multiplying ratio and the term preceding it are both smaller than one. Following similar steps, we can however show that $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ is possible, e.g. for $\gamma_1 = 0.2$, $\gamma_2 = 0.5$, $\rho = 0.3$, and $\lambda = 0.9$.[36] More generally, numerical analyses show that $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ holds for about 3 percent of all feasible parameter combinations.

Case 4. For a change in ρ, $\Delta\beta_T = (\rho_2 - \rho_1)\lambda Cov(e_{T-1}, y_{T-1})$ and

$$\Delta\beta_{T+1} = \rho_2\frac{\frac{1-\gamma\lambda}{\rho_1} - ((\rho_2+\rho_1)+2\gamma\lambda Cov(e_{T-1}, y_{T-1}))}{1+(\rho_2^2-\rho_1^2)+\frac{2\gamma(\rho_2-\rho_1)\lambda\rho_1}{1-\gamma\lambda}}\Delta\beta_T.$$

The condition $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ is likely to be satisfied for low values of ρ_1, e.g. for $\gamma = \frac{3}{8}$, $\rho_1 = \frac{1}{32}$, $\rho_2 = \frac{1}{4}$ and $\lambda = \frac{1}{2}$. Following similar steps, we can show that $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ can likewise hold, e.g. for $\gamma = 0.5, \lambda = 0.9, \rho_1 = 0.4$ and

[36] Specifically, we have $\Delta\beta_{T+2} = \frac{\gamma_2(\gamma_1\lambda-1)f(x)}{g(x)h(x)}\Delta\beta_{T+1}$, where

$$\begin{aligned}
f(x) =& \gamma_1^3\lambda^2\left(\gamma_2-\lambda\right)+\gamma_1\left(\lambda^3-\gamma_2\left(\gamma_2\lambda\left(\gamma_2\lambda-\lambda^2+1\right)+2\lambda^2\rho^2-1\right)\right)\\
&+\left(\gamma_2^2\lambda+\gamma_2+\lambda\right)\left(\gamma_2+\lambda\left(2\rho^2-1\right)\right)+\gamma_1^2\lambda\left(\lambda-2\gamma_2\right)\\
g(x) =& \gamma_1\left(-\left(\gamma_1+\gamma_2\right)\lambda+\lambda^2+1\right)+\gamma_2+\lambda\left(2\rho^2-1\right)\\
h(x) =& -\gamma_2\left(\gamma_2\left(2\gamma_2\lambda\rho^2+\gamma_2^2+2\lambda^2\rho^2+1\right)+2\lambda\rho^2\right)\\
&+\gamma_1\lambda\left(2\gamma_2\lambda\rho^2+\gamma_2^2\left(2\rho^2+1\right)+\gamma_2^4+2\rho^2+1\right)+\left(\gamma_2^2+1\right)\gamma_1^3(-\lambda)+\left(\gamma_2^2+1\right)\gamma_1^2-1
\end{aligned}$$

and it can be shown that the multiplying ratio $\frac{\gamma_2(\gamma_1\lambda-1)f(x)}{g(x)h(x)}$ can be larger than one in absolute value.

$\rho_2 = 0.5$.[37] More generally, numerical analyses show that $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ holds for about 48 percent and $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ for about 8 percent of all feasible parameter combinations.

Case 5. For a change in λ, $\Delta\beta_T = (\lambda_2 - \lambda_1)\rho Cov(e_{T-1}, y_{T-1})$ and

$$\Delta\beta_{T+1} = \lambda_2 \frac{\gamma + (\lambda_2 + \lambda_1)(1 - \gamma\lambda_1) - \rho^2(\lambda_2 + \lambda_1) - 2\gamma\rho Cov(e_{T-1}, y_{T-1})}{1 + \rho^2(\lambda_2^2 - \lambda_1^2) + 2\gamma(\lambda_2 - \lambda_1)\rho Cov(e_{T-1}, y_{T-1})} \Delta\beta_T,$$

and the condition $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ can be satisfied. Following similar steps, we can show that $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ can likewise hold (e.g., if $\gamma = 0.7, \rho = 0.9, \lambda_1 = 0.1, \lambda_2 = 0.9$). More generally, numerical analyses show that $|\Delta\beta_{T+1}| > |\Delta\beta_T|$ holds for about 18 percent and $|\Delta\beta_{T+2}| > |\Delta\beta_{T+1}|$ for about 7 percent of all feasible parameter combinations.

Amplification in later periods is therefore possible for any parameter change. In the constant-variances case with $Var(e_t) = Var(y_t) = 1$ for all t, we have $\Delta\beta_{T+k} = \gamma_2\lambda_2\Delta\beta_{T+k-1}$ for all $k \geq 2$ (see equation (A.12) and (A.13)), such that amplification may occur in period $T+1$ but cannot occur in later periods. □

A.4.3 Proposition 3

Proof. (a) We derive the conditions under which $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$. We assume that $\Delta\beta_{T+1} \neq 0$, i.e. that the conditions in Proposition 2a are satisfied. From the expressions (A.10) and (A.11) for $\Delta\beta_T$ and $\Delta\beta_{T+1}$, the conditions for a sign change $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ are

$$\frac{\Delta Cov(e_T, y_T)}{Cov(e_{T-1}, y_{T-1})} < \frac{\Delta Var(y_T)}{Var(y_{T-1})}$$

if the initial shift was positive ($\Delta\beta_T > 0$) and

$$\frac{\Delta Cov(e_T, y_T)}{Cov(e_{T-1}, y_{T-1})} > \frac{\Delta Var(y_T)}{Var(y_{T-1})}$$

if the initial shift was negative ($\Delta\beta_T < 0$). We plug in the case-specific expressions for $\Delta Cov(e_T, y_T)$ and $\Delta Var(y_T)$ from Table A.1 to evaluate whether these conditions can be

[37] Specifically, we have $\Delta\beta_{T+2} = \frac{\gamma(\gamma\lambda-1)f(x)}{g(x)h(x)}\Delta\beta_{T+1}$, where

$$\begin{aligned} f(x) &= \left((\gamma\lambda-1)^2\left(-\left(\rho_2^2(\gamma+\lambda)-\lambda\right)\right) + \gamma\rho_1\rho_2(2\gamma\lambda+1)(\gamma\lambda-1) + \lambda\rho_1^2(\gamma\lambda+1)(\gamma(\lambda-\gamma)-1)\right) \\ g(x) &= \rho_1\rho_2(\gamma\lambda-1) - \rho_1^2(\gamma\lambda+1) + (\gamma\lambda-1)^2 \\ h(x) &= -2\gamma^2\lambda\rho_1\rho_2(\gamma+\lambda) + \left(\gamma^2+1\right)\rho_1^2(\gamma\lambda+1) + (\gamma\lambda-1)\left(\rho_2^2\left(\gamma^2+2\gamma\lambda+1\right)+1\right) \end{aligned}$$

and the multiplying ratio $\frac{\gamma(\gamma\lambda-1)f(x)}{g(x)h(x)}$ can be larger than one in absolute value.

satisfied. We also use numerical analyses to assess the "likelihood" of the scenario outlined by the proposition. In our derivations we use that expression (A.11) can be rewritten as

$$\Delta\beta_{T+1} = \rho_2\lambda_2 \left(\frac{\Delta Cov(e_T, y_T) - Cov(e_{T-1}, y_{T-1})\Delta Var(y_T)}{Var(y_T)} \right).$$

Specifically,

Case 1. For changes in σ^2, the condition $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ cannot be satisfied because $\Delta\beta_T = 0$ and the IGE shifts only from generation $T+1$ onwards.

Case 2. For changes in Φ^2, the condition $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ cannot be satisfied because $\Delta\beta_T = 0$ and the IGE shifts only from generation $T+1$ onwards.

Case 3. For a single parameter change in γ, $\Delta\beta_T = (\gamma_2 - \gamma_1)$ and (see Prop. 2b, Case 3)

$$\Delta\beta_{T+1} = \lambda Cov(e_{T-1}, y_{T-1}) \frac{\lambda - (\gamma_2 + \gamma_1) - 2\lambda\rho Cov(e_{T-1}, y_{T-1})}{1 + \gamma_2^2 - \gamma_1^2 + 2(\gamma_2 - \gamma_1)\lambda\rho Cov(e_{T-1}, y_{T-1})} \Delta\beta_T.$$

We therefore have $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ if the numerator in the ratio is negative, i.e. if $\gamma_2 + \gamma_1 > \lambda - 2\lambda\rho Cov(e_{T-1}, y_{T-1})$ or $\gamma_2 + \gamma_1 > \frac{\lambda(1-\gamma_1\lambda - 2\rho^2)}{1-\gamma_1\lambda}$, which holds for plausible parameter values. For example, this condition necessarily holds if $\rho > 0.5$. More generally, numerical analysis shows that the condition holds for about 90 percent of all feasible parameter combinations. It is more likely to hold if γ_1 and ρ are high and if λ is low.

Case 4. For a change in ρ, $\Delta\beta_T = (\rho_2 - \rho_1)\lambda Cov(e_{T-1}, y_{T-1})$ and (see Prop. 2b, Case 4)

$$\Delta\beta_{T+1} = \rho_2 \frac{\frac{1-\gamma\lambda}{\rho_1} - ((\rho_2 + \rho_1) + 2\gamma\lambda Cov(e_{T-1}, y_{T-1}))}{1 + (\rho_2^2 - \rho_1^2) + \frac{2\gamma(\rho_2 - \rho_1)\lambda\rho_1}{1-\gamma\lambda}} \Delta\beta_T$$

We therefore have $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ if $(1-\gamma\lambda) < \rho_1(\rho_1 + \rho_2) + \frac{2\gamma\lambda\rho_1\rho_1}{(1-\gamma\lambda)}$. This condition holds for plausible parameter values, for example for $\rho_1(\rho_1 + \rho_2) > (1 - \gamma\lambda)$. More generally, numerical analysis shows that the condition holds for about 12 percent of feasible parameter combinations. The likelihood that the condition holds is strongly increasing in ρ_1.

Case 5. For a change in λ, $\Delta\beta_T = (\lambda_2 - \lambda_1)\rho Cov(e_{T-1}, y_{T-1})$ and (see Prob. 2b, Case 5)

$$\Delta\beta_{T+1} = \lambda_2 \frac{\gamma + (\lambda_2 + \lambda_1)(1 - \gamma\lambda) - \rho^2(\lambda_2 + \lambda_1) - 2\gamma\rho Cov(e_{T-1}, y_{T-1})}{Var(y_T)} \Delta\beta_T.$$

We therefore have $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ if $\gamma(1-\gamma\lambda_1-2\rho^2)+(\lambda_2+\lambda_1)(1-\gamma\lambda_1)(1-\gamma\lambda_1-\rho^2) < 0$, which can only be satisfied if ρ is large and λ_1 sufficiently small. More generally, numerical analysis shows that the condition holds for about 2 percent of feasible parameter combinations. The condition rarely holds for $\rho < 0.7$ or $\lambda_1 > 0.4$.

Non-monotonicity in later generations ($sign(\Delta\beta_{T+k+1}) \neq sign(\Delta\beta_{T+k})$ for some $k \geq 1$) is also possible, but less likely. For example, the conditions for $sign(\Delta\beta_{T+2}) \neq sign(\Delta\beta_{T+1})$ follow directly from the case-by-case expressions for $\Delta\beta_{T+2}$ as a function of $\Delta\beta_{T+1}$ in the proof to Proposition 2b in Appendix A.4.2. If the multiplying ratios in these equations are negative, then $sign(\Delta\beta_{T+2}) \neq sign(\Delta\beta_{T+1})$ holds. Using simulations, we confirm this possibility for shifts in Φ^2, γ, ρ or λ, but not for shifts in σ^2. In fact, we can have $sign(\Delta\beta_T) \neq sign(\Delta\beta_{T+1})$ and $sign(\Delta\beta_{T+1}) \neq sign(\Delta\beta_{T+2})$, i.e. two instances of sign change between T and $T+2$ (primarily for a shift in γ).

For the case with constant variances, the response is always monotonic. First, equation (A.11) simplifies to $\Delta\beta_{T+1} = \rho_2\lambda_2\Delta Cov(e_T, y_T) = \rho_2\lambda_2\left((\rho_2-\rho_1)+(\gamma_2\lambda_2-\gamma_1\lambda_1)\frac{\rho_1}{1-\gamma_1\lambda_1}\right)$. So for a single parameter change, $\Delta\beta_T$ and $\Delta\beta_{T+1}$ always have the same sign.[38] Second, the response in later generations can be expressed as $\Delta\beta_{T+k} = \gamma_2\lambda_2\Delta\beta_{T+k-1}\ \forall k \geq 2$ and is therefore likewise monotonic. □

Proof. (b) We compare the size of $\Delta\beta_T$ and $\Delta\beta_\infty$. From equations (A.10) and (9) it follows

$$\begin{aligned}\Delta\beta_T &= \gamma_2-\gamma_1+\frac{\rho_2\lambda_2\rho_1}{1-\gamma_1\lambda_1}-\frac{\rho_1\lambda_1\rho_1}{1-\gamma_1\lambda_1}\\ \Delta\beta_\infty &= \gamma_2-\gamma_1+\frac{\rho_2^2\lambda_2\Phi_2^2(1-\gamma_2^2)}{\rho_2^2\Phi_2^2(1+\gamma_2\lambda_2)+\sigma_2^2(1-\lambda_2^2)(1-\gamma_2\lambda_2)}-\frac{\rho_1\lambda_1\rho_1}{1-\gamma_1\lambda_1}\end{aligned} \tag{A.21}$$

and therefore

$$\Delta\beta_\infty-\Delta\beta_T = \rho_2\lambda_2\left(\frac{\rho_2\Phi_2^2(1-\gamma_2^2)}{\rho_2^2\Phi_2^2(1+\gamma_2\lambda_2)+\sigma_2^2(1-\lambda_2^2)(1-\gamma_2\lambda_2)}-\frac{\rho_1}{1-\gamma_1\lambda_1}\right) \tag{A.22}$$

For a positive initial shift ($\Delta\beta_T > 0$), we therefore have $\Delta\beta_T > \Delta\beta_\infty$ if the expression inside the parentheses is negative, and vice versa for a negative initial shift.

Case 1. For a shift in σ^2, $\Delta\beta_T = 0$ and $\Delta\beta_\infty = \left(\frac{1}{\rho^2\Phi^2(1+\gamma\lambda)+\sigma^2(1-\lambda^2)(1-\gamma\lambda)}-1\right)\frac{\rho_1\lambda_1\rho_1}{1-\gamma_1\lambda_1}$, so the condition $|\Delta\beta_T| > |\Delta\beta_\infty|$ cannot be satisfied. The initial shift understates the magnitude of the steady-state shift.

Case 2. For a shift in Φ^2, $\Delta\beta_T = 0$ and $\Delta\beta_\infty = \left(\frac{\Phi_2^2(1-\gamma_2^2)(1-\gamma_2\lambda_2)}{\rho_2^2\Phi_2^2(1+\gamma_2\lambda_2)+\sigma_2^2(1-\lambda_2^2)(1-\gamma_2\lambda_2)}-1\right)\frac{\rho_1\lambda_1\rho_1}{1-\gamma_1\lambda_1}$, so the condition $|\Delta\beta_T| > |\Delta\beta_\infty|$ cannot be satisfied. The initial shift understates the magnitude of the steady-state shift.

[38] For example, for a change in γ, $\Delta\beta_T = \gamma_2-\gamma_1$, $\Delta Cov(e_T, y_T) = \frac{(\gamma_2-\gamma_1)\lambda\rho}{1-\gamma\lambda}$, and $\Delta\beta_{T+1} = \frac{\lambda^2\rho^2}{1-\gamma\lambda}\Delta\beta_T$.

Case 3. For a shift in γ, equation (A.22) simplifies to

$$\Delta\beta_\infty - \Delta\beta_T = \rho^2\lambda\left(\frac{\Phi^2(1-\gamma_2^2)}{\rho^2\Phi^2(1+\gamma_2\lambda)+\sigma^2(1-\lambda^2)(1-\gamma_2\lambda)} - \frac{1}{1-\gamma_1\lambda}\right)$$

For a positive initial shift ($\Delta\beta_T = \gamma_2 - \gamma_1 > 0$), we therefore have $\Delta\beta_T > \Delta\beta_\infty$ if the expression inside the parentheses is negative. Using that because of our standardization $\Phi^2 = 1-\lambda^2$ and $\sigma^2 = (1-\gamma_1^2) - \rho^2\frac{1+\gamma_1\lambda}{1-\gamma_1\lambda}$, this condition can be simplified to

$$0 < \rho^2\left(\frac{1+\gamma_2\lambda}{1-\gamma_2\lambda} - \frac{1+\gamma_1\lambda}{1-\gamma_1\lambda}\right) + (\gamma_2-\gamma_1)\left(\frac{(\gamma_2+\gamma_1)-\lambda-\gamma_1\gamma_2\lambda}{1-\gamma_2\lambda}\right).$$

The first term on the right side is always positive. Thus, the inequality condition is more likely to hold if ρ is large. For $\gamma_2 > \gamma_1$, the second term is positive if $\gamma_2+\gamma_1 > \lambda+\gamma_1\gamma_2\lambda$, which will hold if $\gamma_1+\gamma_2$ is large compared to λ (e.g., $\gamma_1 = 0.2$, $\gamma_2 = 0.3$, $\lambda = 0.4$ and $\rho = 0.5$). According to numerical analyses, weak non-monotonicity holds for about 88 percent of all feasible parameter combinations. These analyses also confirm that weak non-monotonicity is more likely if γ_1 or ρ are large or if λ is small.

Case 4. For a shift in ρ, equation (A.22) simplifies to

$$\Delta\beta_\infty - \Delta\beta_T = \rho_2\lambda\left(\frac{\rho_2\Phi^2(1-\gamma^2)}{\rho_2^2\Phi^2(1+\gamma\lambda)+\sigma^2(1-\lambda^2)(1-\gamma\lambda)} - \frac{\rho_1}{1-\gamma\lambda}\right).$$

For a positive initial shift ($\Delta\beta_T = \frac{(\rho_2-\rho_1)\lambda\rho_1}{1-\gamma\lambda} > 0$), we therefore have $\Delta\beta_T > \Delta\beta_\infty$ if the expression inside the parentheses is negative. Using that $\Phi^2 = 1-\lambda^2$ and $\sigma^2 = (1-\gamma^2) - \rho_1^2\frac{1+\gamma\lambda}{1-\gamma\lambda}$ this condition simplifies to

$$(1-\gamma^2)\frac{(1-\gamma\lambda)}{(1+\gamma\lambda)} < \rho_1(\rho_2+\rho_1)$$

which will hold if λ and ρ_1 are sufficiently large (e.g., $\gamma = 0.4$, $\rho_1 = 0.5$, $\rho_2 = 0.6$ and $\lambda = 0.8$). According to numerical analyses, weak non-monotonicity holds for about 23 percent of all feasible parameter combinations. The likelihood that the condition holds is strongly increasing in ρ_1 but also increases in γ and λ.

Case 5. For a shift in λ, equation (A.22) simplifies to

$$\Delta\beta_\infty - \Delta\beta_T = \rho_2\lambda_2\left(\frac{\rho\Phi_2^2(1-\gamma^2)}{\rho^2\Phi_2^2(1+\gamma\lambda_2)+\sigma^2(1-\lambda_2^2)(1-\gamma\lambda_2)} - \frac{\rho}{1-\gamma\lambda_1}\right)$$

For a positive initial shift ($\Delta\beta_T = \frac{\rho^2}{1-\gamma\lambda}(\lambda_2 - \lambda_1) > 0$), we therefore have $\Delta\beta_T > \Delta\beta_\infty$ if the expression inside the parentheses is negative. Using that $\Phi^2 = 1 - \lambda_1^2$ and $\sigma^2 = (1-\gamma^2) - \rho^2\frac{1+\gamma\lambda_1}{1-\gamma\lambda_1}$ this condition simplifies to

$$(1-\gamma^2)\left(\frac{1-\gamma\lambda_1}{1-\gamma\lambda_2} - \frac{1-\lambda_2^2}{1-\lambda_1^2}\right) < \rho^2\left(\frac{1+\gamma\lambda_2}{1-\gamma\lambda_2} - \frac{1+\gamma\lambda_1}{1-\gamma\lambda_1}\frac{(1-\lambda_2^2)}{(1-\lambda_1^2)}\right)$$

Both differences inside the parentheses are positive if $\lambda_2 > \lambda_1$, and the condition holds if $\gamma > 0$ and both γ and ρ are large (e.g., $\gamma = 0.5$, $\rho = 0.6$, $\lambda_1 = 0.6$ and $\lambda_2 = 0.8$). According to our numerical analyses, weak non-monotonicity holds for about 9 percent of all feasible parameter combinations. It is more likely to hold the higher is γ and ρ and the lower is λ_1.

For the case with constant variances, we already showed that the response is monotonic for all periods $t \geq T$, such that $|\Delta\beta_T| > |\Delta\beta_\infty|$ cannot hold (see Proposition 3a). This can be also shown directly: we have $\Delta\beta_T = \gamma_2 - \gamma_1 + \frac{\rho_2\lambda_2\rho_1}{1-\gamma_1\lambda_1} - \frac{\rho_1\lambda_1\rho_1}{1-\gamma_1\lambda_1}$ and $\Delta\beta_\infty = \gamma_2 - \gamma_1 + \frac{\rho_2\lambda_2\rho_2}{1-\gamma_2\lambda_2} - \frac{\rho_1\lambda_1\rho_1}{1-\gamma_1\lambda_1}$, and therefore $\Delta\beta_\infty - \Delta\beta_T = \rho_2\lambda_2\left(\frac{\rho_2}{1-\gamma_2\lambda_2} - \frac{\rho_1}{1-\gamma_1\lambda_1}\right)$ which is always positive for positive changes in either γ, ρ or λ. □

Proof. (c) We compare the signs of $\Delta\beta_T$ and $\Delta\beta_\infty$. The sign of $\Delta\beta_\infty$ has already been derived in the proof of Proposition 1, so it remains to compare the sign of this shift to the sign of $\Delta\beta_T$. The condition is never satisfied for a single parameter change. For a change in σ^2, $\Delta\beta_T = 0$ while $\Delta\beta_\infty < 0$. For a change in Φ^2, $\Delta\beta_T = 0$ while $\Delta\beta_\infty > 0$. For a positive change in γ, ρ or λ, $\Delta\beta_T > 0$ and $\Delta\beta_\infty > 0$ (under the conditions given in the Proposition 1 proof). However, the condition can hold for multiple parameter changes in the single-skill model (see Case 2 in Section 3) or for a single parameter change in the multi-skill model (see Case 3 in Section 3). □

A.4.4 Proposition 4

Proof. We derive the conditions under which $sign(\Delta r_T) \neq sign(\Delta\beta_T)$. Assume first that $\Delta\beta_T > 0$ and find the condition such that $\Delta r_T < 0$. As $\Delta r_T = \frac{Cov(y_T, y_{T-1})}{\sigma_{y_T}\sigma_{y_{T-1}}} - \frac{Cov(y_{T-1}, y_{T-2})}{\sigma_{y_{T-1}}\sigma_{y_{T-2}}}$ and $\sigma_{y_{T-1}} = \sigma_{y_{T-2}}$, the condition $\Delta r_T < 0$ corresponds to $\frac{\beta_T}{\sigma_{y_T}} < \frac{\beta_{T-1}}{\sigma_{y_{T-1}}}$, which we rewrite as

$$\frac{\beta_T^2 - \beta_{T-1}^2}{\beta_{T-1}^2} < \frac{Var(y_T) - Var(y_{T-1})}{Var(y_{T-1})}. \tag{A.23}$$

Similarly, for the case $\Delta\beta_T < 0$ we have $\Delta r_T > 0$ if $\frac{\beta_T^2 - \beta_{T-1}^2}{\beta_{T-1}^2} > \frac{Var(y_T) - Var(y_{T-1})}{Var(y_{T-1})}$. We therefore have $sign(\Delta r_T) \neq sign(\Delta\beta_T)$ if the shifts in the IGE and the variance of y have the same sign, and the latter shifts more strongly than the square of the former (in percentage terms). This condition may hold because some parameter changes shift the variance but not the IGE in generation T. Specifically (see also Table A.1),

Case 1. For an increase in market luck, $\sigma_2 > \sigma_1$, we have $\Delta\beta_T = 0$ and $\Delta r_T < 0$. Since market luck is by definition uncorrelated with parental income, we have $\Delta\beta_T = Cov(y_T, y_{T-1}) - Cov(y_{T-1}, y_{T-2}) = 0$. However, $\Delta Var(y_T) = Var(y_T) - Var(y_{T-1}) = \sigma_2^2 - \sigma_1^2 > 0$ and condition (A.23) holds.

Case 2. For an increase in endowment luck, $\Phi_2 > \Phi_1$, we have $\Delta r_T < 0$ and $\Delta\beta_T = 0$. Since endowment luck is by definition uncorrelated with parental income, we have $\Delta\beta_T = 0$. However, $\Delta Var(y_T) = \rho^2(\Phi_2^2 - \Phi_1^2) > 0$ (assuming $\rho > 0$) and condition (A.23) holds.

Case 3. For an increase in the direct effect of parental income, $\gamma_2 > \gamma_1$, we always have $\Delta\beta_T > 0$ and $sign(\Delta r_T) = sign(\Delta\beta_T)$. Specifically, from Table A.1 we have $\Delta\beta_T = \gamma_2 - \gamma_1$ and $\Delta Var(y_T) = \gamma_2^2 - \gamma_1^2 + 2(\gamma_2 - \gamma_1)\rho\lambda Cov(y_{T-1}, e_{T-1})$, which can be rewritten as $\Delta Var(y_T) = (\gamma_2 - \gamma_1)^2 + 2(\gamma_2 - \gamma_1)\beta_{T-1}$. Moreover, we have $\beta_T^2 - \beta_{T-1}^2 = (\triangle\beta_T)^2 + 2\triangle\beta_T\beta_{T-1}$. We can therefore rewrite condition (A.23) as

$$\frac{\gamma_2 - \gamma_1 + 2\beta_{T-1}}{\beta_{T-1}^2} < \gamma_2 - \gamma_1 + 2\beta_{T-1}, \tag{A.24}$$

which cannot hold as the steady-state elasticity β_{T-1} satisfies $0 \leq \beta_{T-1} \leq 1$.

Case 4. For an increase in returns, $\rho_2 > \rho_1$, $sign(\Delta r_T) \neq sign(\Delta\beta_T)$ is possible, especially if λ or ρ_1 are small. From Table A.1 we have $\Delta\beta_T = \frac{(\rho_2-\rho_1)\lambda\rho_1}{1-\gamma\lambda} \geq 0$ and $\Delta Var(y_T) = \rho_2^2 - \rho_1^2 + \frac{2\gamma(\rho_2-\rho_1)\lambda\rho_1}{1-\gamma\lambda} \geq 0$. It can further be shown that $\beta_T^2 - \beta_{T-1}^2 = (\frac{(\rho_2-\rho_1)\lambda\rho_1}{1-\gamma\lambda})^2 + 2(\frac{(\rho_2-\rho_1)\lambda\rho_1}{1-\gamma\lambda})\beta_{T-1}$. Note that $\Delta\beta_T = 0$ but $\Delta Var(y_T) > 0$ if either $\rho_1 = 0$ or $\lambda = 0$. More generally, condition (A.23) holds if either λ or ρ_1 are small.[39] See also Figure 2, which contains several examples in which $\Delta\beta_T > 0$ and $\Delta r_T < 0$ after an increase in skill returns in a multi-skill model.

Case 5. For an increase in heritability, $\lambda_2 > \lambda_1$, $sign(\Delta r_T) \neq sign(\Delta\beta_T)$ is possible if γ is very large and ρ is sufficiently small. From Table A.1 we have $\Delta\beta_T = \frac{(\lambda_2-\lambda_1)\rho^2}{1-\gamma\lambda_1} > 0$ and $\Delta Var(y_T) = \rho^2(\lambda_2^2 - \lambda_1^2) + 2\gamma(\lambda_2 - \lambda_1)\rho^2/(1 - \gamma\lambda_1) > 0$ (assuming $\rho > 0$). For instance, the condition holds when $\gamma = 0.9$, $\rho = 0.2$,

[39]We can use numerical analysis to explore the regularity with which the condition holds. Specifically, we simulate a joint uniform distribution of $(\gamma, \lambda, \rho_1, \rho_2)$ with each parameter being strictly between zero and one. We exclude combinations that imply that β_{T-1} or β_T is larger than one and those which require $\sigma^2 < 0$ for $Var(y_{T-1}) = 1$. Focusing on cases with $\rho_2 > \rho_1$, the condition $sign(\Delta r_T) \neq sign(\Delta\beta_T)$ holds in 68 percent of cases across the parameter space. The condition is more likely to hold when λ or ρ_1 are low.

$\lambda_1 = 0.1$, $\lambda_2 = 0.5$. Similar to the previous case, we can use numerical analysis to explore the regularity with which the condition holds for different parameter values.[40]

□

A.4.5 Proposition 5 and Cohort Dynamics

Proposition 5 states that structural changes can have a sudden impact on mobility in the first affected generation, but will have only a gradual effect on mobility trends over cohorts in subsequent generations. To show this, consider the following notation to distinguish cohorts and generations. Let $C(t)$ denote the cohort into which generation t of a family is born. Let $A(t-1, C(t))$ denote the age at birth of the corresponding parent. Member $t-j$ of a family is then born in cohort

$$C(t-j) = C(t) - A(t-1, C(t)) - ... - A(t-j, C(t-j+1)). \tag{A.25}$$

Denote realizations of these random variables by lower case letters and abstract from life-cycle dynamics, such that a parameter with subscript $c(t)$ represents the average economic environment over the lifecycle of cohort c.[41] The (scalar) equivalent to equation (7) is then

$$\beta_{c(t)} = \frac{Cov\left(y_{c(t)}, y_{C(t-1)}\right)}{Var\left(y_{C(t-1)}\right)} = \gamma_{c(t)} + \rho_{c(t)}\lambda_{c(t)}\frac{Cov\left(e_{C(t-1)}, y_{C(t-1)}\right)}{Var\left(y_{C(t-1)}\right)}, \tag{A.26}$$

The IGE for a given cohort depends on cohort-specific transmission mechanisms and the variance and covariance of income and endowments in the parent generation. Sudden changes in the economic environment as represented by the parameters γ_c, ρ_c and λ_c therefore have a sudden impact on the IGE in the first generation (while changes in market luck σ_c^2 or endowment luck Φ_c^2 shift the IGE only from the second generation onwards, as illustrated in Figure 1d). However, the variance and covariance of income and endowments in the parent generation may vary with the timing of fertility among parents, because different parental cohorts may have been subject to different policies and institutions. For example, using the

[40] Simulating the analog parameter distribution and focusing on cases with $\lambda_2 > \lambda_1$, the condition for $sign(\Delta r_T) \neq sign(\Delta\beta_T)$ holds in about 6 percent of cases across this parameter space. The condition is more likely to hold when ρ is low and when γ is high. In fact, the condition can only hold when ρ is about 0.4 or lower and γ is at least 0.75 (suggesting implausibly high values of β). The interpretation is that if the IGE is determined almost entirely by the direct effect of parental income, an increase in the heritability of skills can satisfy condition (A.23) since the relative effect on the variance of income dominates the relative effect on the (squared) covariance between parental and child income.

[41] For a consideration of lifecycle dynamics see for example Conlisk (1969) or Heckman and Mosso (2014).

law of iterated expectations we have

$$Cov\left(e_{C(t-1)}, y_{C(t-1)}\right) = \sum_{a(t-1)} f_{c(t)}\big(a(t-1)\big) Cov\big(e_{c(t)-a(t-1)}, y_{c(t)-a(t-1)}\big), \qquad \text{(A.27)}$$

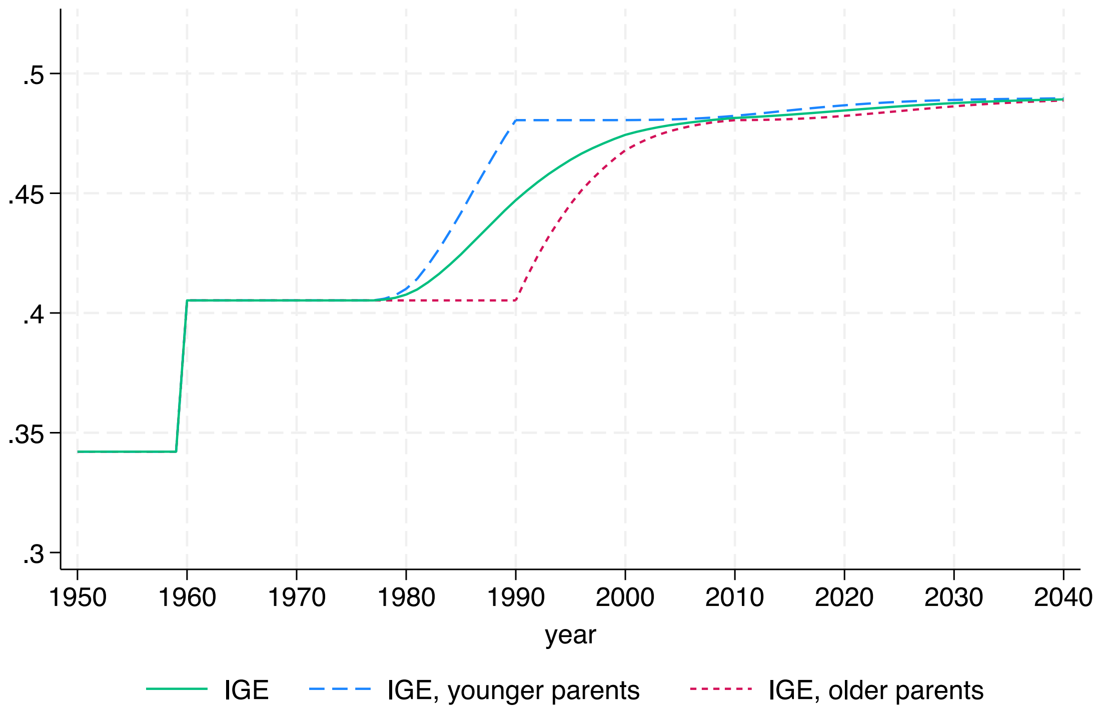
where $f_{c(t)}$ is the probability mass function for parental age at birth of cohort c_t (and where we abstracted from mean changes across cohorts, which would enter as an additional term). The IGE depends on current mechanisms and a *weighted average* of the cross-covariances of income and endowments in previous cohorts, where the weights are given by the cohort-specific distribution of timing of fertility (i.e. parental age at birth). Each covariance can be iterated backwards to show that it is a differently weighted function of past parameter values, and thus past structural changes, implying that past structural changes have only a gradual impact on current mobility trends.

Figure A.1 provides an illustration of these arguments. Specifically, we plot the transition path of the IGE in response to an increase in the returns to skills ($\rho \uparrow$). For comparability we pick the same parameter values as in Section 3, such that Figure A.1 plots the "cohort-level" counterpart to the "generation-level" transition path shown in Figure 1b. We assume that the structural change (the increase in ρ) occurs suddenly in 1960; many types of structural changes would result in more gradual parameter changes, but we focus on a sharp change here as it results in a starker contrast between the initial shift of the IGE and its subsequent transition path. Moreover, we assume that the age-of-parent at birth follows the observed age distribution of fathers of children born in Sweden in 1960, with a median age of 31.

The simulation illustrates that even though the first-generation shift is sudden, the second-generation shift in the IGE is spread out over many cohorts, as there is a substantial variation in the age-at-birth among parents. To better illustrate this pattern, the figure also contains the implied IGE for children with younger parents (up to age 30 at the birth of their child) and older parents (aged 31 and above). For example, among children born in 1990, the younger parents were already exposed to higher skill returns, so the second-generation shift in the IGE is already visible; in contrast, the older parents of children in that same birth cohort were still subject to low skill returns, so the second-generation shift in the IGE has not occurred yet within this group. While highly stylized, the example illustrates how IGE trends on the cohort level "smear" across generations, and the influence of economic environments in different calendar years, due to staggered life cycles.

Equation (A.27) and Figure A.1 also point to a simple diagnostic test that practitioners can implement to distinguish whether an observed shift in an intergenerational coefficient of interest is due to a structural change in the current generation, or a structural change that occurred in past generations. Rather than estimating a single pooled regression, researchers should study how the IGE varies with the age of parents at birth. Figure A.2 illustrates this argument in our Swedish application, reporting the coefficient estimates from a regression

Figure A.1: Increasing Returns to Skills (Cohort Dynamics)

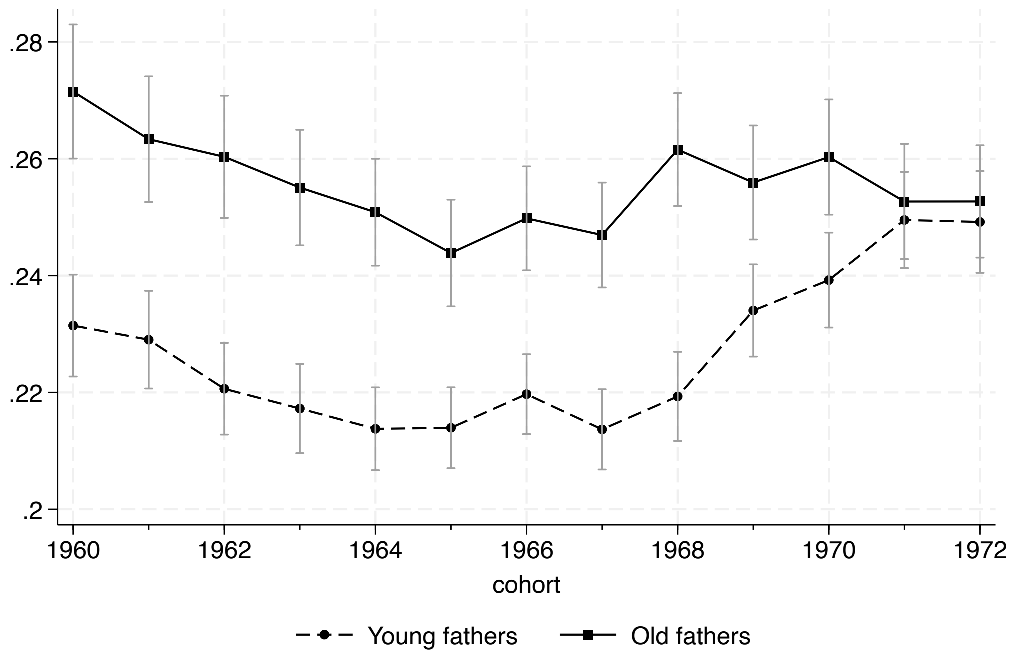


Note: Simulated transition path of the intergenerational elasticity (IGE) across birth cohorts, assuming that for children born in 1960 the returns to skills increase from $\rho_1 = 0.2$ to $\rho_2 = 0.5$ (assuming $\gamma = 0.3$ and $\lambda = 0.8$). We assume that the age-of-parent at birth follows the observed age distribution of fathers of children born in Sweden in 1960 (median age: 31). In addition, we also plot the IGE for the subset of children from parents below or above median parental age at birth, respectively.

of years of schooling of children born in the indicated birth cohort on years of schooling of their fathers, separately for "young" and "old" fathers. We find that the increase in the intergenerational coefficient among cohorts born in the late 1960s and early 1970s happens only among "young" parents, some of whom were themselves exposed to the school reform that affected birth cohorts born from the late 1940s onwards.

Of course, this simple diagnostic test is not foul-proof, as the selection into fertility might change over time in ways that differentially affect the intergenerational coefficient for "young" or "old" parents. If researchers study a particular structural change, a more targeted approach would exploit direct information on whether a given parent was or was not exposed to that change (as we do in Section 5.2). And in settings with limited sample size and less precise intergenerational estimates, as in our US application based on the PSID, it can still be fruitful to study whether the covariance between different parental characteristics changes over birth cohorts (as in Table 1).

Figure A.2: The Intergenerational Education Coefficient in Sweden, by Parental Age



Note: Each dot represents the coefficient from a regression of years of schooling of the offspring in the respective birth cohort on years of schooling of their fathers, separately for fathers aged 28 or below at the birth of their child (young fathers, dashed line) or above (old fathers, solid line). Grey bars: 95% confidence intervals.

A.5 Sibling Correlations: A Simple Illustration

We here illustrate how the sibling correlation, denoted r^S, shifts along its transition path in response to a change in skill returns ρ, as in Case 1 (Section 3) and Case 4 (Section 4). Our objective is to further illustrate that the dynamics in different measures of the importance of family background can differ. Consider first the simplified version of our baseline model, with a single endowment e_t and scalar versions of equations (5) and (6), such that

$$y_{tij} = \gamma_t y_{t-1j} + \rho_t e_{tij} + u_{tij} \tag{A.28}$$

$$e_{tij} = \lambda_t e_{t-1j} + \Phi_t v_{tij}. \tag{A.29}$$

where j denotes the parent and i the sibling. The luck term u_{tij} may consist of two parts: a shock that is common to siblings and one that is not, i.e. $u_{tij} = c_{tj} + \tilde{u}_{tij}$. While not affecting our main qualitative findings, this decomposition illustrates that the sibling correlation may capture a broader concept of the role of family background than standard intergenerational measures. Most commonly, the sibling correlation is based on the variance decomposition $Var(y_{tij}) = Var(a_{tj}) + Var(b_{tij})$, where $Var(a_{tj})$ denotes the variance between families whereas $Var(b_{tij})$ denotes the variance across individuals within families. Specifically, in our model (suppressing the i and j subscripts)

$$r_t^S = \frac{Var(a_t)}{Var(a_t) + Var(b_t)} = \frac{Var(\gamma_t y_{t-1} + \rho_t \lambda_t e_{t-1} + c_t)}{Var(\gamma_t y_{t-1} + \rho_t \lambda_t e_{t-1} + c_t + \rho_t \Phi_t v_t + \tilde{u}_t)}, \tag{A.30}$$

Focus first on the compensated response, such that $Var(y_t)$ is held constant by a compensating shift in the variance of $\tilde{u}_t$. A change in returns from $\rho_{t<T} = \rho_1$ to $\rho_{t \geq T} = \rho_2$ shifts the sibling correlation in the first affected generation according to

$$\Delta r_T^S = r_T^S - r_{T-1}^S = (\rho_2^2 - \rho_1^2)\lambda^2 + \frac{2\gamma\lambda\rho_1(\rho_2 - \rho_1)}{1 - \gamma\lambda}, \tag{A.31}$$

and in the second generation according to

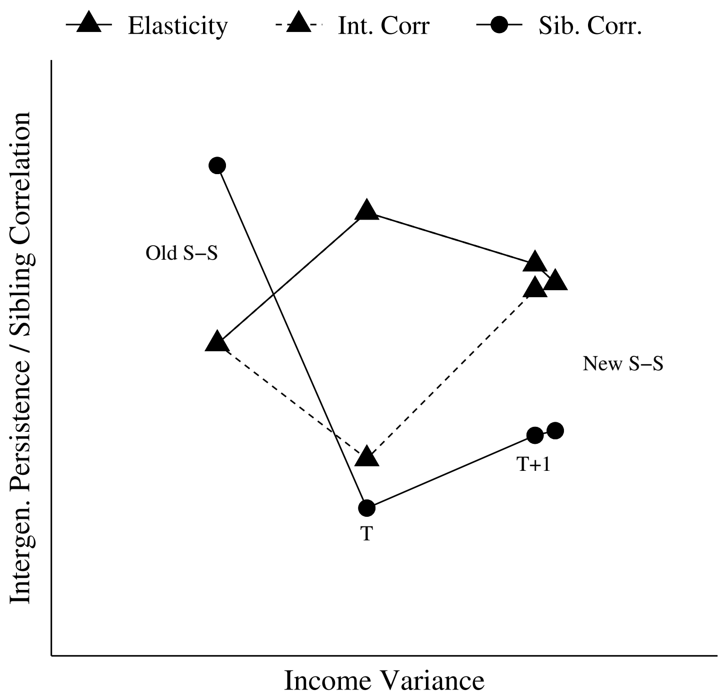
$$\Delta r_{T+1}^S = 2\gamma\rho_2\lambda\Delta Cov(e_T, y_T) = 2\gamma\lambda\rho_2(\rho_2 - \rho_1), \tag{A.32}$$

In a simple meritocratic economy ($\gamma = 0$), the sibling correlation only shifts in the first affected generation and is unaffected thereafter. In contrast, equation (12) in Section 3 shows that the IGE does continue to shift in the second generation, and that this shift can very well be larger than its corresponding first-generation shift. More generally, the second-generation shift of the sibling correlation depends on the causal effect of parental income γ, which is likely to be small, while the second-generation shift of the IGE does not.[42] The sibling correlation responds therefore more immediately (supporting Proposition 4).

We further consider the dynamics of the sibling correlation in the model with multiple skills. Figure A.3, based on a similar parametrization to case (a) in Figure 2, illustrates that the IGE and the sibling correlation tend to exhibit quite different transition paths also in this more general model. As for the intergenerational measures, their transition paths can be non-monotonic and span over multiple generations. However, the initial shift in the sibling correlation is comparatively large, while the subsequent shifts (driven by dynamics in the cross-sectional variance of income and its causal impact γ) are much smaller in size. In the specific parameterization illustrated in Figure 2, the sibling correlation declines while intergenerational persistence increases in steady state. The intuition is that the steady-state responses depend on the heritability of the endowment for which returns increase relative to other components of family background; the sibling correlation captures a broader concept of family background (i.e. the term c_t in eq. (A.30)) and may thus decrease if the endowment's heritability is not similarly high.

[42] The empirical literature rarely finds *bivarate correlations* between log incomes of parents and children higher than 0.5 and the causal effect of parent on child income is believed to be much smaller.

Figure A.3: An Increase in the Returns to a Single Skill



Note: Transitional dynamics of the intergenerational elasticity, intergenerational correlation, sibling correlation and the average of the variance of income in the parent and child generation. Parameters are $\gamma = 0.2$, $\lambda_k = 0.3$, $\lambda_l = 0.8$, $Var(u_t) = 0.5$, $Var(c_t) = 0.4$, and $\rho_l = 0.4$. In generation T, the returns to skill k increase from $\rho_{k,1} = 0.3$ to $\rho_{k,2} = 0.8$.

A.6 Choice of Parameter Values

Our main findings do not rely on specific parameter choices, but our numerical examples will benefit from parametrizations that are consistent with the empirical literature. One difficulty is that some variables in our model represent broad concepts (e.g., human capital $\boldsymbol{h}_t$ may include any productive characteristic of an individual), which are only imperfectly captured by data. In addition, the parameters of the model reflect total effects from those variables. While estimates of (intergenerational) correlations and other moments are widely reported, there exists less knowledge about the relative importance of the various underlying causal mechanisms. Although only indicative, we can at least choose parameter values that are consistent with the available evidence.

Lefgren et al. (2012) examine the relative importance of different mechanisms in a transmission framework that is similar to ours. Using imperfect instruments that are differentially correlated with parental human capital and income they estimate that in Sweden the effect from parental income (captured by the parameter γ) explains about a third of the intergenerational elasticity, while parental human capital explains the remaining two thirds. In our model we further distinguish between a direct and indirect (through human capital accumulation) effect from parental income, as captured by the parameters γ_y and γ_h, but the total effect is sufficient for the parameterization of our examples.

The literature provides more guidance on the transmission of physical traits such as height or cognitive and non-cognitive abilities, for which we use the term *endowments*. Common to these are that genetic inheritance is expected to play a relatively important role. From the classic work of Galton to more recent studies the evidence implies intergenerational correlations in the order of magnitude of about 0.3-0.4 when considering one and much higher correlations when considering both parents.[43] Those estimates may reflect to various degrees not only genetic inheritance but also correlated environmental factors; we capture both in the *heritability* parameter λ (estimates of genetic transmission are then a lower bound), for which values in the range 0.5-0.8 seem reasonable. Note that we use the term "heritability" in a broad sense, while the term refers only to genetic inheritance in the biological literature.

Finally, a reasonable lower-bound estimate of the *returns* ρ to endowments and *human capital* can be approximated by evidence on the explanatory power of earnings equations. Studies that observe richer sets of covariates, including measures of cognitive and non-cognitive ability, typically yield estimates of R^2 in the neighborhood of 0.40.[44] On the one hand, such estimates are likely to underestimate the explanatory power of (broadly defined) human capital as of imperfect measurement and omitted variables. On the other hand, we want to only capture returns to the component of human capital that is not due to parental income and investment; we capture the latter channel instead in the parameter γ_h (and its contribution to offspring income in γ). In any case, values of ρ in the range of 0.6-0.8 should be at least roughly consistent with the empirical evidence.[45]

These parameter ranges are consistent with recent estimates of the intergenerational income elasticity β in the US, which are typically in the range of 0.45-0.55 (see Black and Devereux, 2011). Given reliable elasticity estimates we can also cross-validate and potentially narrow down the implied range for the structural parameters of the model. We write each parameter as a function of the others in steady state,

$$\beta = \gamma + \frac{\rho^2\lambda}{1-\gamma\lambda} \qquad \gamma = \frac{\beta\lambda + 1 \pm \sqrt{\beta^2\lambda^2 - 2\beta\lambda + 4\lambda^2\rho^2 + 1}}{2\lambda}$$
$$\rho = \sqrt{\frac{(\beta-\gamma)(1-\gamma\lambda)}{\lambda}} \qquad \lambda = \frac{\beta-\gamma}{\beta\gamma + \rho^2 - \gamma^2}, \tag{A.33}$$

and plug in the discussed values on the right-hand sides to impute parameter ranges that

[43] For estimates of correlations in measures of cognitive ability, see Bowles and Gintis (2002) and the studies they cite; for measures of both cognitive ability and non-cognitive ability, see Grönqvist et al. (2017).

[44] See for example Lindqvist and Vestman (2011) for Sweden. Fixed-effects models yield higher estimates, although some of the difference may be capturing persistent luck rather than unobserved characteristics.

[45] In the initial steady state we standardize $Var(y) = Var(e) = 1$, such that $R^2 = 0.4$ translates into $\rho \approx 0.63$.

are consistent with our reading of the empirical literature. Specifically we rule out too high values of λ and ρ as they cause γ to approach zero, to arrive at

$$0.45 \leq \beta \leq 0.55, \quad 0.15 \leq \gamma \leq 0.25, \quad 0.60 \leq \rho \leq 0.70, \quad 0.50 \leq \lambda \leq 0.65.$$

These implied ranges should not be taken literally, but are sufficient to provide a reasonable illustration of the potential quantitative implications of our findings.

A.7 Correlated endowments

We revisit Case 3 under the assumption that $\mathbf{\Lambda}_t$ is not diagonal, such that elements of the endowment vector $\boldsymbol{e}_t$ are potentially correlated. Suppose that at generation T the returns to human capital change from $\boldsymbol{\rho}_1$ to $\boldsymbol{\rho}_2$ but that the steady-state variance of income remains unchanged.

By substituting equation (5) for y_{T-1} and income in previous generations we can express the pre-shock elasticity as

$$\beta_{T-1} = Cov(y_{T-1}, y_{T-2}) = \gamma + \boldsymbol{\rho}_1' Cov(\boldsymbol{e}_{T-1}, y_{T-2}) = \gamma + \boldsymbol{\rho}_1' \mathbf{\Gamma} \boldsymbol{\rho}_1 \tag{A.34}$$

where

$$\mathbf{\Gamma} = \sum_{l=1}^{\infty} \gamma^{l-1} Cov(\boldsymbol{e}_{T-1}, \boldsymbol{e}_{T-1-l}) \tag{A.35}$$

is the cross-covariance between the endowment vectors of offspring and parents (if $\gamma = 0$), or a weighted average of the endowment vectors of parents and earlier ancestors ($0 < \gamma < 1$). These cross-covariances measure to what degree each offspring endowment is correlated with the *same* endowment in previous generations (the diagonal elements) and each of the *other* $K - 1$ endowments (the off-diagonal elements). Note that $\mathbf{\Gamma}$ does not depend on t if these cross-covariances are in steady state.

We can similarly derive the elasticity in the first affected generation and in the new steady state as

$$\beta_T = \gamma + \boldsymbol{\rho}_2' \mathbf{\Gamma} \boldsymbol{\rho}_1 \tag{A.36}$$

$$\beta_\infty = \gamma + \boldsymbol{\rho}_2' \mathbf{\Gamma} \boldsymbol{\rho}_2. \tag{A.37}$$

The conditions under which a change in skill prices leads to a non-monotonic response in mobility can be easily summarized if the cross-covariances $Cov(\boldsymbol{e}_{T-1}, \boldsymbol{e}_{T-j})\ \forall j > 1$ are symmetric. Symmetry requires the correlation between offspring endowment k and parent endowment l to be as strong as the correlation between offspring endowment l and parent endowment k, $\forall k, l$. We can then note that

$$\begin{aligned} 2\beta_T &= 2\left(\gamma + \boldsymbol{\rho}_2'\boldsymbol{\Gamma}\boldsymbol{\rho}_1\right) \\ &= \gamma + \boldsymbol{\rho}_1'\boldsymbol{\Gamma}\boldsymbol{\rho}_1 + (\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')\boldsymbol{\Gamma}\boldsymbol{\rho}_1 + \gamma + \boldsymbol{\rho}_2'\boldsymbol{\Gamma}\boldsymbol{\rho}_2 + \boldsymbol{\rho}_2'\boldsymbol{\Gamma}(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2) \\ &= \beta_{T-1} + \beta_\infty + (\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')\boldsymbol{\Gamma}\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2'\boldsymbol{\Gamma}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1) \\ &= \beta_{T-1} + \beta_\infty - (\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')'\boldsymbol{\Gamma}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1), \end{aligned} \tag{A.38}$$

where we expanded and subtracted $\boldsymbol{\rho}_1'$ and $\boldsymbol{\rho}_2$, substituted equations (A.34) and (A.37), and finally took the transpose and used the symmetry of $\boldsymbol{\Gamma}$ to collect all remaining terms in a quadratic form.

Let S denote the subset of prices that do not change in generation T, and denote by $\boldsymbol{\Gamma}_S$ and $\boldsymbol{\Lambda}_S$ the minors of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ that are formed by deleting each row and column that correspond to an element in S. The quadratic form $(\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')'\boldsymbol{\Gamma}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1)$ is greater than zero for $\boldsymbol{\rho}_2 \neq \boldsymbol{\rho}_1$ if $\boldsymbol{\Gamma}_S$ is positive definite. A sufficient condition for $\boldsymbol{\Gamma}_S$ to be positive definite is diagonality of the heritability matrix $\boldsymbol{\Lambda}_S$, with positive diagonal elements. More generally, the matrix $\boldsymbol{\Gamma}_S$ is positive definite if the respective minors of the cross-covariances $Cov(\boldsymbol{e}_{T-1}, \boldsymbol{e}_{T-j})\ \forall j > 1$ are strictly diagonally dominant. Strict diagonal dominance requires that the correlation between offspring endowment k and parent endowment k is stronger than the sum of its correlation to all other relevant parent endowments $l \neq k, l \in S$ (i.e., offspring are similar instead of dissimilar to their parents).

Price changes then increase intergenerational mobility temporarily (β_T is below *both* the previous steady state β_{T-1} and the new steady state β_∞) as long as the steady-state elasticity shifts not too strongly, specifically iff

$$|\beta_\infty - \beta_{T-1}| < (\boldsymbol{\rho}_2' - \boldsymbol{\rho}_1')\boldsymbol{\Lambda}\left(\boldsymbol{I} - \gamma\boldsymbol{\Lambda}\right)^{-1}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1). \tag{A.39}$$

A.8 Data: The Panel Study of Income Dynamics

In Section 4 we estimate mobility trends and the components of the IGE for the US using data from the Panel Study of Income Dynamics (PSID). We use data from all PSID waves released between 1968 and 2017. The survey was annual up until 1997 and has been biannual thereafter. The PSID is a highly useful resource for intergenerational mobility research because it follows children from the original sample as they grow older and form their own households.

Apart from a few exceptions, we follow the sampling procedure and variable definitions used by Lee and Solon (2009). As such, we use only the core sample, also known as the Survey Research Center component of the PSID. We focus on sons born starting from 1952. To measure parental income, we average log annual family income in the childhood home over the three years when the child was 15-17 years old, which is also similar to the measure in Chetty et al. (2014b). We measure the children's adult income by the (log) annual family income in the household in which they were the household head or head's spouse. While Lee and Solon use all annual child income observations from age 25 and later, we focus only on incomes between age 30 and 35, for two reasons. First, we keep the child age (of income observation) composition constant over time, which is important due to concerns about life-cycle bias and the fact that we use almost 20 more years of income data. Second, focusing on the 30-35 age range makes the estimates more comparable to Chetty et al. (2014b), who measure child income in the early 30s. Similar to Lee and Solon we exclude outlier observations (using the same income thresholds) and treat each child-year combination as a separate observation (within the 30-35 age range).

We estimate by OLS the regression model

$$INC_{itc} = \eta_{tc} + \eta_{ac} + \beta_c PINC_{ic} + \gamma_c PINC_{ic} AGE_{itc} + \lambda_c PAGE_{ic} + \mu_c PAGE_{ic}^2 + \varepsilon_{itc}, \tag{A.40}$$

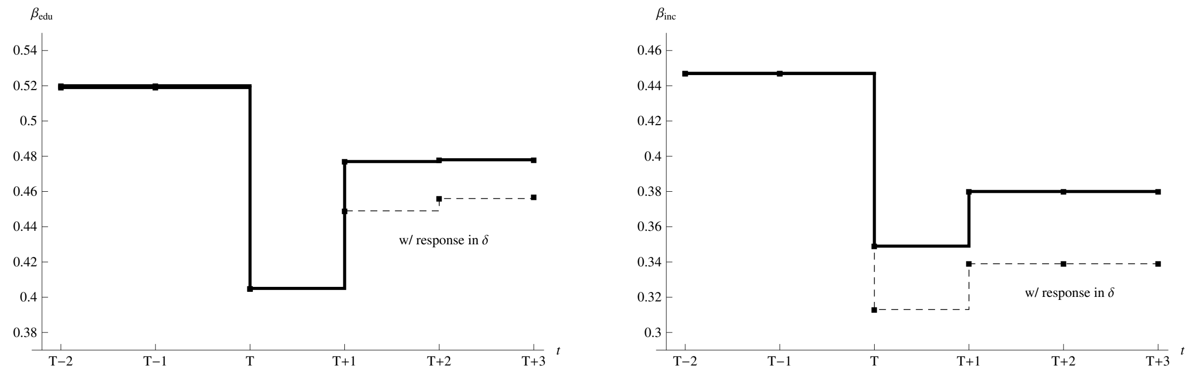
where INC_{itc} is log family income in year t for child i in decade of birth c = *1950, 1960, 1970, 1980* and $PINC_{ic}$ is log parental income. We condition on fixed effects for calendar year η_{tc} and the age η_{ac} at which child income is observed, a linear interaction between child age AGE_{itc} and log parental income, and a quadratic function in parental age at the time their income is observed ($PAGE_{ic}$). We estimate the regression separately by decade of birth, thus the c-subscripts on the coefficients. We normalize AGE_{itc} such that it equals 0 at age 33, which implies that we can interpret β_c as the intergenerational elasticity at child's age 33 for the different birth cohorts. The normalization has no effect on the variation in β_c over cohorts, but due to life-cycle effects leads to slightly lower estimates as compared to the normalization to age 40 in Lee and Solon (2009). We cluster standard errors on the individual child level. Given available survey years and our income measures, we are able to estimate this model for cohorts born in the 1950s, 1960s, 1970s, and 1980s (specifically for birth years 1952-1986).

A.9 Compulsory Schooling in the Intergenerational Model

To predict the impact of a compulsory schooling policy on educational and income mobility, first include constants α_y and α_h into the scalar variants of our baseline equations (2)-(3), thus allowing for mean changes in income and education. The school reform raises schooling of individuals with particularly low educational attainment. This "mechanical" shift may in turn affect the attainment of others via potential general equilibrium responses. Compositional changes may generate peer effects, and changes in supply may alter the returns to

Figure A.4: Raising the compulsory schooling level

(a) Intergenerational educational coefficient   (b) Intergenerational income elasticity



Note: Income and educational mobility trends in numerical example, with $x = 9$, $\alpha_y = 9$, $\gamma_y = 0$, $\delta = 0.2$ (dashed line: $\delta = 0.18$), $\alpha_h = 0$, $\gamma_h = 1$, $\theta = 2$, $\lambda = 0.6$, and (u_y, u_h, v) normally distributed with variances $(0.1, 3, 0.64)$.

schooling and thus schooling decisions.[46] However, a theoretical discussion of the numerous responses that may occur over such long time intervals can be only incomplete and speculative. We instead focus on the main "mechanical" effect of the school reform, which explains the observed empirical pattern well. To capture it assume that eq. (3) determines *intended* schooling h^*, while from generation T onwards *actual* schooling h_t is compulsory until x years, such that

$$h_t = \begin{cases} h_t^* & \text{if } t < T \\ max(h_t^*, x) & \text{if } t \geq T \end{cases} . \tag{A.41}$$

Consider the dynamic response in the most popular measure of income and educational mobility, the intergenerational elasticity of income β_{inc} and educational coefficient β_{edu},

$$\beta_{inc,t} = \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} \quad \text{and} \quad \beta_{edu,t} = \frac{Cov(h_t, h_{t-1})}{Var(h_{t-1})}. \tag{A.42}$$

In our main model, we derived the transitional dynamics by repeated insertion of the structural equations of our model, using linearity of the expectation operator to solve for the required moments. But the compulsory schooling requirement generates non-linear relationships that depend also on the distributions of the errors in these equations.

Figure A.4 provides a simulated numerical example based on simple parametric assumptions (e.g., normally distributed errors). From generation T schooling becomes compulsory until $x = 9$ years. We assume that parental schooling has only modest indirect intergenerational spillover effects ($\gamma_h = 1$) and choose other parameters such to generate pre-reform first and second moments for schooling h_t that are similar to the observed moments in the

[46]Spillover effects on educational attainment of individuals not directly affected by the reform were found to be small in Holmlund (2007).

Swedish data.

Panel A plots the response of the intergenerational educational coefficient β_{edu}. In offspring generation T the reform compresses the variance of schooling strongly, which decreases the numerator of β_{edu} – differences in schooling between parents result into smaller differences among their offspring. However, from generation $T + 1$ the variance of schooling is also compressed among *parents*, who were already subject to the school reform in the previous generation. The coefficient β_{edu} is inversely scaled by this variance, and thus tends to rise. The non-monotonic response is thus mainly a consequence of strong changes in the variance of the marginal distributions (a direct and mechanical effect of the reform).

The reform could lead to further substantial compressions of educational attainment in subsequent generations if schooling has very strong causal effects on offspring outcomes ($\gamma_h \gg 1$). However, the existing empirical literature points to modest intergenerational "multiplier" effects of education (Holmlund et al., 2011). The dashed line illustrates one potentially important general equilibrium response: increased supply of formal schooling may decrease its returns on the labor market (a decrease in δ), decreasing inequality in income and thus (if human capital accumulation is subject to parental investments) educational inequality and intergenerational persistence. Assuming that formal schooling improves an individual's earnings potential, the pattern in the income elasticity β_{inc} tends to be similar but differs in some aspects (Panel B in Figure A.4). For example, the potential general equilibrium response the returns to formal schooling might affect income and thus the income elasticity already in generation T (dashed line).

A.10 The Swedish Compulsory School Reform

In the aftermath of World War II, many European countries implemented large-scale educational reforms with the main purpose of extending the level of schooling. The rapid post-war economic development in Europe increased the demand for better educated and more skilled workers, and especially in the Scandinavian countries there was also a strong desire to reform the education system as a means to increase equality of opportunity.

The Swedish compulsory school reform is comprehensively discussed in Holmlund (2007), so we describe here only its most important elements. Gradually implemented across municipalities from the late 1940s, the reform's two main components were to *raise compulsory schooling* from seven or eight to nine years, and to *postpone tracking* decisions. The reform also prescribed a unified national curriculum and municipalities received additional state funding to cover costs from its implementation. One of the Swedish reform's main objectives was to increase educational attainment among students from less advantaged backgrounds (Erikson and Jonsson, 1996).

In the pre-reform school system, students typically went through grades one to six in a basic compulsory school (*folkskolan*). In the final year, more able students were selected for an academically oriented junior secondary school (*realskolan*), while remaining students stayed in the non-academic basic school until completion of compulsory education. Compulsory education was typically seven years long, although in some municipalities (mainly the bigger cities) the minimum was eight years. Upon completion of basic compulsory school, students went on to either full-time vocational education or to work. Students completing junior secondary school instead typically went on to higher education.

In 1948, a proposal was made to replace the old system with a nine-year compulsory comprehensive school. Students were allowed to choose between three different programs with varying academic content after sixth grade but there would be no selection based on grades and all pupils would attend the same schools. All schools would also share a unified national curriculum. This new system was implemented as a nation-wide experiment between 1949 and 1962, in which the proposed schools were introduced municipality by municipality, rather than by separate schools or classes. New municipalities were successively added each year. The new school system was finally nationally implemented in 1962, following parliamentary approval.

The reform does not constitute a fully randomized experiment; while their representativeness was a criterion for eligibility, the timing of the reform was not independent of municipality characteristics. But the *gradual* implementation of the reform provides a source of variation that enables researchers to control for both regional and cohort-specific effects. Similar expansion schemes were also adopted in Norway and Finland, and the design of these reforms has inspired several important studies that exploit the gradual implementation as a source of exogenous variation in education (Meghir and Palme, 2005; Black et al., 2005; Pekkarinen et al., 2009).

A.11 Data, Descriptives, and Basic Evidence

Our source data set is based on a 35 percent random sample of the Swedish population born between 1932 and 1967. Using information based on population registers, we construct an *intergenerational sample* by linking these sampled individuals to their biological parents and children. We then individually match data on personal characteristics and place of residence based on bi-decennial censuses starting from 1960, as well as education and income data stemming from official registers. For our main analysis of the effect of the school reform on mobility, however, we restrict this sample further. For our first-generation analysis, we focus on those born 1943-1955 (the cohorts that were directly affected by the reform introduction) and their parents. Each observation thus consists of the schooling, income and other relevant characteristics of a child in the directly affected generation (born 1943-1995) and of that child's father. For our second-generation analysis, an observation is based on the same variables for children born 1966-1972 and their fathers, some of which belong to the directly affected generation.

Educational registers were compiled in 1970, 1990 and about every third year thereafter, containing detailed information on each individual's educational attainment. We consider for each individual the highest attainment recorded across these years. The information on schooling levels is translated into years of education with 7 years for the old compulsory school being the minimum, and 20 years for a doctoral degree the maximum. Education data in 1970 is available only for those born 1911 and later. We can therefore not observe schooling for parents who were 33 years or older at their child's birth in 1943 (at the onset of the reform implementation). This age limit increases by a year for each subsequent offspring cohort, potentially creating a confounding trend in mobility measures over cohorts due to non-random sample selection. For comparability we thus restrict our *intergenerational sample* to parent-child pairs in which parents were no older than 32 years when their child was born. Educational data may also be missing for other reasons, in particular if parents had died or emigrated before 1970. The probability of such occurrences is potentially related to individual characteristics, but the share of affected observations is small.[47] As the data are collected from official registers there are no standard non-response problems.

The most recent educational register was compiled in 2007, which allows us to consider mobility trends in terms of years of education for cohorts born from the early 1940s up until 1972. Attainment of individuals at the top of the educational distribution is not reliably covered for more recent cohorts; only a small population share is affected, but measurement error in the tails of the distribution could have a disproportionately large effect on intergenerational mobility measures.

For studying mobility in terms of income, we construct a measure of long-run income status based on age-specific averages of annual incomes, which are observed for the years 1968-2007. We use total (pre-tax) income, which is the sum of an individual's labor (and labor-related) earnings, early-age pensions, and net income from business and capital realizations. We express all incomes in 2005 prices and exclude observations with average incomes below 10000 SEK (equivalent to about 1300 US dollars). Incomes for parents are necessarily measured at a later age than incomes for their offspring, which may bias estimates of the intergenerational elasticity of lifetime income. Such bias is less problematic for our purposes as we are interested in mobility differences between groups instead of the overall level of income mobility in the population. For estimation of the intergenerational elasticity, we use the log of the age-specific averages of annual incomes. For estimating intergenerational correlations we standardize our log income measures by birth year, while for rank correlations

[47] Educational information are less often missing among offspring, due to their younger age and the more frequent measurement of education after 1990. The share of missing observations does not vary with reform status (conditional on municipalities and offspring cohorts), and has thus little effect on our causal analysis.

Table A.2: Sample Statistics by Birth Cohort

	Source data			Intergenerational samples				
	# obs.	reform shares		# obs.	with non-missing		reform shares	
		(offspring)	(fathers)		(educ.)	(inc.)	(offspring)	(fathers)
1943	42,138	0.04	0.00	17,211	15,008	7,912	0.05	0.00
1944	44,715	0.06	0.00	18,425	16,179	10,607	0.07	0.00
1945	44,682	0.06	0.00	18,604	16,441	12,702	0.07	0.00
1946	44,299	0.11	0.00	19,124	17,101	15,113	0.11	0.00
1947	43,288	0.18	0.00	19,078	17,103	16,873	0.18	0.00
1948	42,527	0.31	0.00	19,063	17,192	16,968	0.31	0.00
1949	40,628	0.39	0.00	18,449	16,768	16,489	0.40	0.00
1950	38,854	0.53	0.00	19,421	17,657	17,348	0.54	0.00
1951	36,951	0.56	0.00	18,644	17,016	16,750	0.57	0.00
1952	37,031	0.69	0.00	19,102	17,442	17,135	0.70	0.00
1953	37,537	0.79	0.00	19,452	17,904	17,614	0.80	0.00
1954	35,668	0.86	0.00	18,453	16,955	16,625	0.86	0.00
1955	36,440	0.95	0.00	19,122	17,569	17,204	0.96	0.00
1956	36,666	1.00	0.00	20,942	19,217	18,763	1.00	0.00
...	...	...	...	...	...	...	...	...
1965	42,909	1.00	0.01	28,447	26,762	26,203	1.00	0.01
1966	43,050	1.00	0.01	29,043	27,415	26,710	1.00	0.02
1967	42,686	1.00	0.02	28,897	27,366	26,643	1.00	0.03
1968	54,105	1.00	0.04	33,526	32,524	31,686	1.00	0.05
1969	52,317	1.00	0.05	32,157	31,315	30,460	1.00	0.06
1970	53,908	1.00	0.07	32,508	31,788	30,670	1.00	0.08
1971	56,493	1.00	0.09	33,251	32,539	31,343	1.00	0.12
1972	57,035	1.00	0.12	33,081	32,409	31,123	1.00	0.16

Note: Father-child pairs are included in the intergenerational sample if father's age at birth of the child is below 33.

we use income ranks by birth year.

While we present evidence on mobility in father-child pairs, the consideration of maximum parental education and income yields similar results. We test the robustness of our results using other samples with no or different restrictions on parental age, or alternative measures of parental education and income, some of which we will also report below.

To construct the reform dummy, which indicates whether an individual was subject to the new system of comprehensive schooling, we follow the procedure first used by Holmlund (2008). Reform status can be approximated using information on an individual's birth year (from the administrative register) and place of residence during school age (from the censuses).[48] The gradual implementation of the reform affected cohorts born between 1938 and

[48] Reform status across cohort-municipality cells can be inferred by tracing in which cohort, for each municipality, the share graduating from the old school system discontinuously drops to zero (or close to zero). Helena Holmlund has kindly provided us with her coding, and we refer to Holmlund (2007) for further details on the coding procedure and potential measurement issues.

1955, but the school municipality cannot be reliably determined for individuals born before 1943. As the share of individuals affected by the reform was very small we set the reform dummy to zero for all cohorts before 1943 (and one for all cohorts after 1955).[49]

Table A.2 describes, by birth cohort, both the source data and the intergenerational sample, which was drawn according to the conditions described above. The number of observations for each cohort are listed in columns 2 and 5. Columns 6 and 7 describe the number of observations with non-missing education or income information. Columns 3-4 and 8-9 describe how the share of offspring and fathers attending reformed schools increases over cohorts. It increases faster among fathers in the intergenerational sample than in the source data, due to oversampling of younger parents in the former.[50]

The reform had a direct impact on educational attainment, which can be also measured with high precision over long time intervals.[51] Figure A.5 plots the mean and variance of years of schooling of offspring cohorts (1933-1972) and their fathers (1911-1935) in our intergenerational sample. Vertical bars at the 1943 and 1955 cohorts indicate the start and end point of the reform's implementation. A reform effect on *average* years of schooling is not easily discernible from panel (A). Indeed, Holmlund (2007) finds the reform effect on mean schooling to be small (lower bound estimate of 0.19 years), as only a share of children are affected by the compulsory requirement. In contrast, the shift in the *variance* of schooling is more striking: the reform period coincides with a sudden and strong compression of the distribution of schooling. Comparison with earlier trends in the first half of the 20th century illustrates the exceptional magnitude of those changes.

Figure A.6 provides more direct evidence on the reform impact. Recentering the data within each municipality, we compare educational attainment and the intergenerational educational coefficient before and after a cohort was first subject to the new school type. The share of individuals with less than 9 years, the variance of schooling and the intergenerational schooling coefficient all drop strongly with local reform implementation.

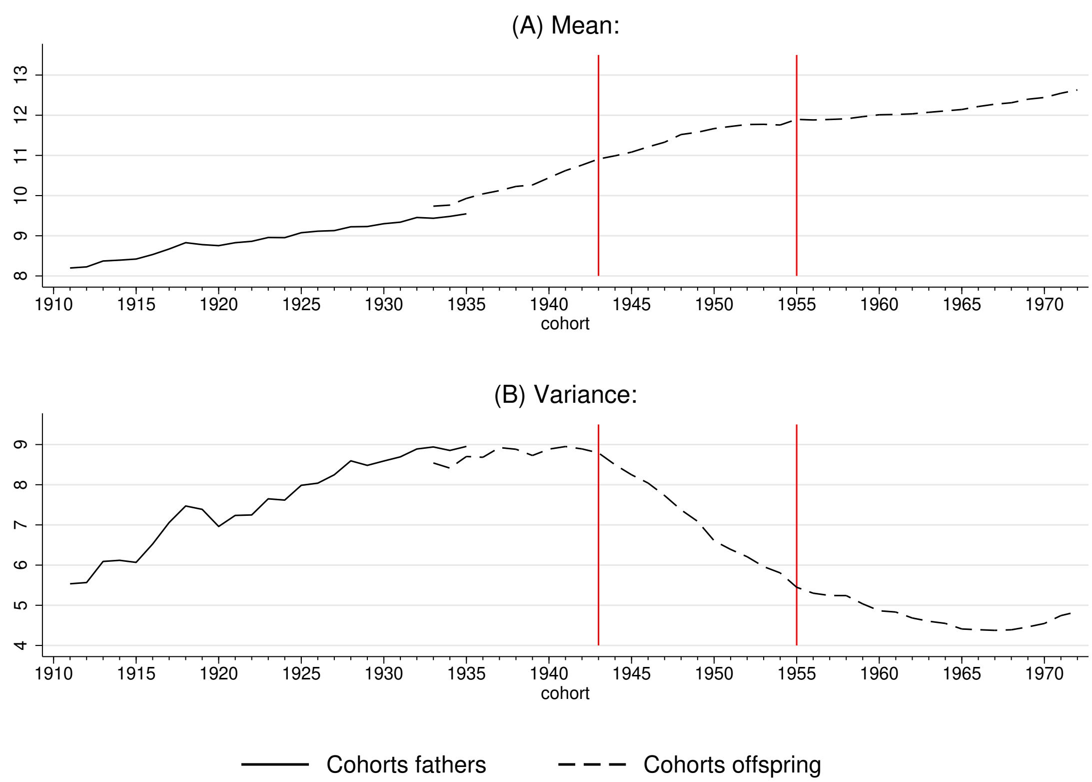
In Nybom and Stuhler (2014) we show that the reform effect was strongly heterogeneous across cohorts. The reform reduced the intergenerational schooling coefficient by almost 25 percent in those municipalities that were subject to the reform already in the early 1940s.

[49]Cohorts born before 1943 were subject to the new school system in 33 out of a total of 1034 municipalities. With the exception of less than a handful mid-sized urban municipalities, all of these were small, rural municipalities. We further drop a small number of municipalities for which the implementation date is unclear.

[50]A smaller share of individuals from the raw data are sampled among earlier cohorts, as their fathers are less likely to be identified in the source data. Identification of the reform effect requires that the probabilities that fathers, education and income are observed do not change systematically with introduction of the reform. While sampling probabilities differ across birth cohorts and municipalities, the correlation with reform status is negligible.

[51]A measure of education in later life is likely to capture an individual's entire educational attainment, as most people complete schooling in early life. In contrast, differences in current incomes are poor proxies of differences in lifetime income, such that measures of income mobility (in particular of mobility trends) are sensitive even to small changes in the age at which incomes are observed (the *life-cycle bias* problem, see Jenkins, 1987, Haider and Solon, 2006, and Nybom and Stuhler, 2016).

Figure A.5: Mean and Variance of Years of Schooling over Cohorts



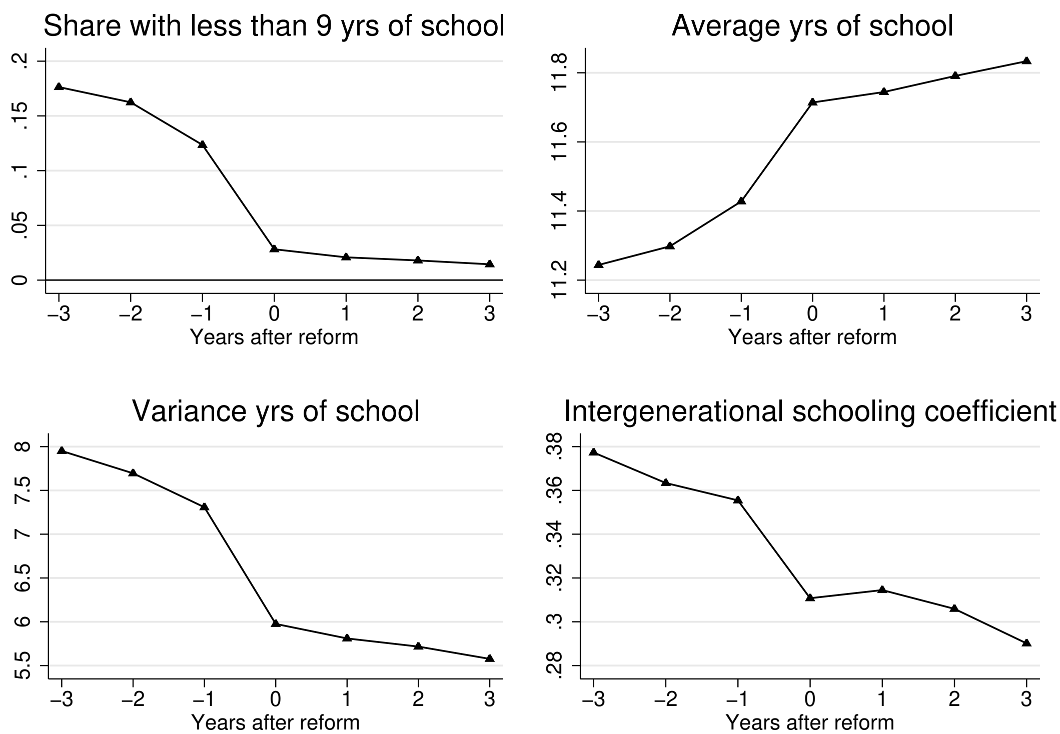
Note: Moments of years of schooling over cohorts of offspring (dashed line) and their fathers (solid line) in intergenerational sample.

But its impact shrinks over cohorts, and becomes indistinguishable from zero after the 1951 cohort. The reason becomes clear from Figure A.5. The secular rise of educational attainment made the reform's compulsory schooling requirement less consequential, and by the early 1950s most pupils were attending school for at least nine years anyways. Our estimates can thus be interpreted as representing an intention-to-treat effect, with the share of compliers diminishing over cohorts.

A.12 Robustness of Empirical Results

We perform a number of tests to probe the robustness of our results. Table A.3 compares our baseline estimates of the reform effect on the intergenerational educational coefficient and income elasticity with estimates from six alternative specifications. First, we include matched siblings in our sample, which increases its size but also diminishes representativeness for some cohorts (see data subsection). Second, we restrict the sample to younger fathers with age at birth below 30, to probe the sensitivity of our results to such age restrictions. Our third robustness tests address measurement error in the reform indicator. Individuals who have been in a lower than expected grade from delayed school entry or grade repetition may have

Figure A.6: Educational Attainment and Intergenerational Mobility, Pre- vs. Post-Reform



Note: We recenter the data such that the reform occurs at time zero for each municipality. Panels (a)-(c) summarize the distribution of offspring educational attainment. Each dot in panel (d) represents the coefficient from a regression of years of schooling of offspring on years of schooling of their fathers. Based on intergenerational sample (fathers aged below 33). Grey bars: 95% confidence intervals.

been subject to the reform before others from the same birth cohort (see Holmlund, 2007). The resulting attenuation bias can be reduced by dropping all individuals born in the cohort just preceding local implementation of the reform. Fourth, we use the maximum of both parents' (instead of the father's) educational attainment or income. Fifth, we include additional controls for the birth cohort of fathers (first generation) or offspring (second generation estimates). Finally, we include municipality-specific linear time trends to support the common trends assumption that is underlying our difference-in-differences analysis.

Our estimates of the reform effect on the intergenerational educational coefficient remain statistically significant on the $p < 0.001$ level across all specifications. Their sizes vary either very little or as expected. In particular, they increase in absolute size when measurement error in the reform indicator is being addressed (column 4). Estimates differ slightly also when we estimate a parent-offspring (instead of father-offspring) measure of persistence, using maximum education among both mothers and fathers as independent variable (column 5). Estimates of the reform effect on the intergenerational income elasticity have always the same sign, but vary more strongly and are not always statistically significant on the $p < 0.05$ or even $p < 0.1$ level. Two factors reduce precision. First, long-run income is measured

Table A.3: Robustness Tests

	baseline	with siblings	fathers below 30	pre-reform dropped	parental max.	cohort controls	municip. time trends
Education:							
1st gen.	-0.0371***	-0.0393***	-0.0408***	-0.0434***	-0.0357***	-0.0387***	-0.0364***
	(0.0072)	(0.0054)	(0.0089)	(0.0083)	(0.0064)	(0.0073)	(0.0074)
2nd gen.	0.0655***	0.0651***	0.0655***	0.0710***	0.0307***	0.0655***	0.0622***
	(0.0128)	(0.0122)	(0.0128)	(0.0139)	(0.0093)	(0.0126)	(0.0131)
Income							
1st gen.	-0.0196*	-0.0078	-0.0181	-0.0195*	-0.0210**	-0.0233**	-0.0239**
	(0.0100)	(0.0068)	(0.0115)	(0.0118)	(0.0088)	(0.0095)	(0.0097)
2nd gen.	0.0410*	0.0148	0.0410*	0.0492*	0.0344**	0.0418**	0.0363*
	(0.0216)	(0.0165)	(0.0216)	(0.0238)	(0.0155)	(0.0212)	(0.0219)

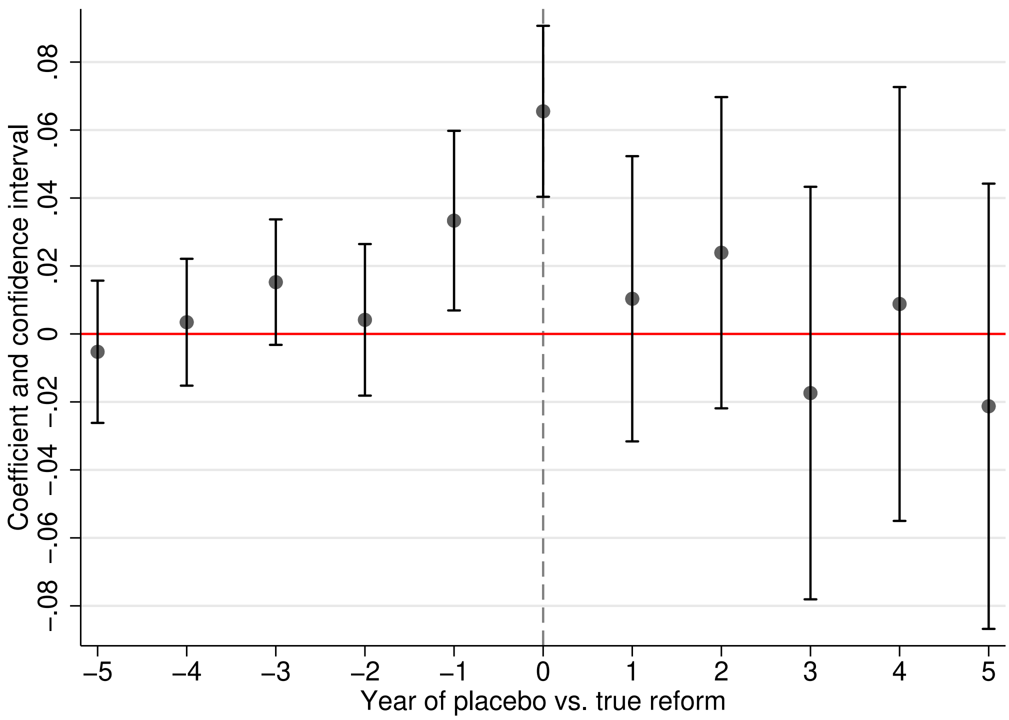
Note: Sensitivity analyses reporting the coefficient on the interaction between reform dummy and parental education and income and clustered standard errors (in parentheses), * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1 contains the baseline specification. For the next columns we include the sibling subsample, restrict the sample to fathers with age at birth below 30, drop those born in the cohort preceeding the reform implementation, use the maximum of mother's and father's education or income, include father (rows 1 and 3) or offspring cohort dummies (rows 5 and 7), or include municipality-specific linear trends.

with much larger error than educational attainment. Second, the reform had a mechanic and strong effect on the distribution of educational attainment, while incomes were only indirectly affected.

Overall the tests corroborate the existence and the direction of reform effects on the intergenerational persistence in both education and income, but underscore that the former is more precisely estimated. We provide further evidence on the suitability of our identification strategy and the common trends assumption by performing a number of placebo tests. Following Meghir et al. (2011) we falsely assume that the reform took place before or after the actual implementation date. We first sample only those offspring born in 1966 to 1972 whose fathers were subject to the reform and generate a placebo "non-treated" group by pretending that the school reform was implemented one year later, two years, three years, and so on. Similarly, we sample only those fathers who were not treated and pretend that the reform was implemented earlier, thus generating a placebo "treated" group.

Each dot represents the estimate of the reform effect on the intergenerational educational coefficient assuming the reform took place at the specified period before or after the actual implementation date. The largest estimate is obtained when we use the correct timing for the reform assignment (at zero). We find small and insignificant estimates in all other cases, except when we assume that the reform was implemented one year before the actual date. Measurement error in reform status is a potential explanation for this observation, as discussed above and also visible from Figure A.6 – those in a lower than expected grade may have been subject to the reform even though not captured by our reform indicator (see Holmlund, 2007).

Figure A.7: Placebo Test: Second Generation



Note: Each dot represents an estimate of the reform effect on the intergenerational educational coefficient in cohorts 1966-72 under the assumption that the reform took place at the specified period before or after the actual implementation date. Based on intergenerational sample (fathers aged below 33). Grey bars: 95% confidence intervals.

The resulting estimates are plotted in Figure A.7. Each dot represents the estimate of the reform effect on the intergenerational educational coefficient assuming the reform took place at the specified period before or after the actual implementation date. The largest estimate is obtained when we use the correct timing for the reform assignment (at zero). We find small and insignificant estimates in all other cases, except when we assume that the reform was implemented one year before the actual date. Measurement error in reform status is a potential explanation for this observation, as discussed above and also visible from Figure A.6 – those in a lower than expected grade may have been subject to the reform even though not captured by our reform indicator (see Holmlund, 2007).