Interpreting Trends in Intergenerational Mobility

Martin Nybom*, Jan Stuhler†‡

November 25, 2016

Abstract

We argue that events in previous generations are important determinants of contemporaneous shifts in intergenerational mobility. To show this, we first study the dynamic response of income mobility to structural changes in a model of intergenerational transmission. Mobility depends on past policies and institutions, such that reforms may generate long-lasting mobility trends over multiple generations. These trends are often non-monotonic. As mobility tends to be highest when a structural change occurs, declining mobility may reflect past gains rather than a recent deterioration of “equality of opportunity”. Exploiting administrative data over three generations, we study a compulsory school reform in Sweden to illustrate some of the implications of our model. The reform reduced the transmission of income and educational inequalities from parents to their offspring in the directly affected generation, but increased their persistence in the next.

*Stockholm University, SOFI and IFAU (martin.nybom@sofi.su.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 Fernandez, Raffaella Giacomini, Stephen Jenkins, Mikael Lindahl, Matthew Lindquist, Steve Machin, Magne Mogstad, Lars Nesheim, Marieke Schnabel, Gary Solon, and seminar participants at University College London, the 2013 RES Conference at Royal Holloway, the Workshop on Intergenerational Mobility at the University of Copenhagen in June 2013, the Swedish Institute for Social Research, the 2013 EALE Conference at the University of Torino, the Universita Cattolica in Milan, 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.
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 most 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, and Black and Devereux, 2011 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 necessitates a dynamic perspective, but existing theoretical work focuses instead on the relation between transmission mechanisms and the long-run or 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 human capital through 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

\(^1\)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 “Ever Higher Society, Ever Harder to Ascend” in The Economist (Dec. 2004), “Moving Up: Challenges to the American Dream” in the Wall Street Journal (May 2005), “The Mobility Myth” in The New Republic (Feb. 2012), 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).
trends. Conversely, 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 fairly broad class of structural changes cause non-monotonic transitions between steady states. The response in mobility switches at some point its sign, and mobility in the first affected generation and in steady state may shift in different directions. Such non-monotonic transitions can also occur in conjunction with changes in cross-sectional inequality, which interact with the transitional dynamics in mobility. Only accounting for the initial response of mobility may thus lead to misleading interpretations about the long-run consequences of structural change.

Third, a change in the strength of one transmission mechanism relative to another can generate transitional mobility, as some families gain while others lose. 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 offspring from poor families. But while mobility increases in the first affected generation, it is bound to decline again in subsequent generations, as 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. As we show, this insight applies more generally in a model with multiple skills.

Finally, we examine a Swedish compulsory school reform to illustrate the first two implications empirically. Exploiting its gradual implementation across areas, and the availability of registry 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. The timing of fertility proves central to understand this pattern. Because this timing varies so strongly, the second-generation effect is likely to persist up to recent cohorts.

Our work relates to both the theoretical and the growing empirical literature on trends in intergenerational mobility. However, most theoretical studies examine only the relationship between causal transmission mechanisms and the implied steady-state level of mobility. An early exception is Atkinson and Jenkins (1984). While they show that failure of the steady-state assumption impedes identification of structural parameters, we instead consider the dynamic effects of changes in such parameters on mobility. Solon (2004) notes that the interpretation of mobility trends would benefit from a theoretical perspective, and 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 alterna-

\[^2\]Moreover, Jenkins (1982) discusses stability conditions for systems of stochastic linear difference equations with constant coefficients, Conlisk (1974b) derives stability conditions for systems with random coefficients.
tive causes of rising cross-sectional inequality. Our paper also builds on Becker and Tomes (1979) and the large literature on individual income processes (e.g. Meghir and Pistaferri, 2011). While Becker and Tomes analyze the dynamics of individual outcomes within families, our objective is instead to study how such processes relate to the dynamics of aggregate moments that capture intergenerational mobility.\(^3\)

The empirical literature is broad. A large and long-standing, originally sociological, literature documents occupational and class mobility over time (see Breen, 2004, Hauser, 2010, Long and Ferrie, 2013, and Modalsli, 2015). 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 (see Black and Devereux, 2011). Such studies face substantial data requirements, and some of the emerging evidence appears conflicting.\(^4\) 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, or Breen and Jonsson, 2007). Common to all 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 structural parameters and summarize our main propositions on the dynamic content of the model in Section 2. We study a set of theoretical cases to illustrate our main results in Section 3 and, with a focus on the dynamics in cross-sectional inequality, in Section 4. Section 5 presents our empirical application, and Section 6 concludes.

### 1 A Model of Intergenerational Transmission

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

\(^3\)The dynamics in individual processes can have interesting implications. For example, Solon (2014) notes that the transmission dynamics within families can be more complex than a Markov representation suggests if grandparents have an independent causal effect on their grandchildren. While he considers the consequences for the steady-state level of mobility, such higher-order effects would also lead to prolonged transitional dynamics.

\(^4\)Hertz (2007), Lee and Solon (2009) and Chetty et al. (2014a) find no evidence of a major trend in the second half of the 20th century in the US, but cannot reject more gradual changes over time. In contrast, Levine and Mazumder (2007) as well as Aaronson and Mazumder (2008) argue that mobility has fallen in recent decades. A decline has also been found for the UK (Blanden et al., 2004; Nicoletti and Ermisch, 2007), while mobility increased in the Nordic countries (Bratberg et al., 2007; Pekkala and Lucas, 2007; Björklund et al., 2009; Pekkarinen et al. forthcoming).
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. The IGE is given by the slope coefficient in the linear regression

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

The IGE \( \beta_t \) captures a statistical relationship so the error \( \epsilon_{i,t} \) is uncorrelated with the regressor by construction. Under stationarity of the variance of \( y_{i,t} \) 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 or sibling correlation.

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). We show in Appendix A.1 that the pathways represented by these equations can be derived from the optimizing behavior of parents in an underlying utility-maximization framework (Becker and Tomes, 1979; Goldberger, 1989; and Solon, 2004).\(^5\)

The structural equations of our model are

\[
y_{it} = \gamma_{y,t} y_{it-1} + \delta t h_{it} + u_{y,it} \tag{2}
\]

\[
h_{it} = \gamma_{h,t} y_{it-1} + \Theta_t e_{it} + u_{h,it} \tag{3}
\]

\[
e_{it} = \Lambda_t e_{it-1} + \Phi t w_{it}. \tag{4}
\]

From equation (2), income \( y_{it} \) in generation \( t \) of family \( i \) is determined by parental income \( y_{it-1} \), own human capital \( h_{it} \), 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 \( h_{it} \), 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 \( \delta t \). The random shock term \( u_{y,it} \) captures factors that do not relate to

\(^5\)This and all subsequently referred to appendix sections are located online at [insert url].
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 $e_{it}$, and chance $u_{h,it}$. A role for parental income through the $J \times 1$ vector $\gamma_{h,t}$ may stem from parental investment into offspring human capital, and the elements of $\gamma_{h,t}$ may differ if parental investments are more targeted or more effective on some types of human capital than others. Parental income may thus affect offspring income directly (through $\gamma_{y,t}$) or indirectly (through $\gamma_{h,t}$). The $J \times K$ matrix $\Theta_t$ governs how endowments such as abilities or preferences, represented by the $K \times 1$ vector $e_{it}$, affect the accumulation of human capital. Endowments are partly inherited from parental endowments $e_{it-1}$ through the $K \times K$ heritability matrix $\Lambda_t$, and partly due to chance $v_{it}$. We consider heritability in a broad sense, potentially reflecting both genetic inheritance and family environment. Market luck $u_{y,it}$ and the elements of $u_{h,it}$ and $v_{it}$ 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 $\gamma_{h,t}$ and $\delta'_t \Theta_t$ are non-negative) and that parent and offspring endowments are not negatively correlated (elements of $\Lambda_t$ are non-negative), for all $t$.

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

$$y_t = \gamma_t y_{t-1} + \rho_t e_t + \sigma_t u_t$$  \hspace{0.5cm} (5)

$$e_t = \Lambda_t e_{t-1} + \Phi_t v_t,$$  \hspace{0.5cm} (6)

where the parameter $\gamma_t = \gamma_{y,t} + \delta'_t \gamma_{h,t}$ aggregates the direct and indirect effects of parental income, the $1 \times K$ vector $\rho_t = \delta'_t \Theta_t$ 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 $\sigma_t u_t = u_{y,t} + \delta'_t u_{h,t}$ aggregates the luck terms related to income and human capital. As $\rho_t$ captures returns to both endowments and skills we use these terms interchangeably below. We normalize the variance of $u_t$ to one in all periods, such that changes in the importance of market luck are captured by $\sigma_t$.

Our model has a similar structure as the model in Conlisk (1974a), but in contrast to the previous literature we assume that income depends on human capital through a vector of

---

6The 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.
distinct skill dimensions. This generalization is central for some of our findings. 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 $\xi_t$ consists of the set of all parameter values that generation $t$ is subject to, i.e. $\xi_t = \{\gamma_t, \rho_t, \sigma_t, \Lambda_t, \Phi_t\}$. A structural change is a permanent change in any of the features of the environment in generation $t = T$, such that $\xi_{t<T} = \xi_1 \neq \xi_{T+s} = \xi_2 \forall s \geq 0$.

For simplicity, we assume that the moments of all variables were in steady-state equilibrium before the structural change occurs, and that the process is stable (implicitly restricting the parameter space, see Appendix A.2). These assumptions simplify the discussion and facilitate comparisons to the existing literature. We also normalize the variances of $y_t$ and elements of $e_t$ 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

$$
\beta_t = \frac{Cov(y_t, y_{t-1})}{Var(y_{t-1})} = \gamma_t + \frac{\rho_t' \Lambda_t Cov(e_{t-1}, y_{t-1})}{Var(y_{t-1})}.
$$

Thus, $\beta_t$ depends on current transmission mechanisms (parameters $\gamma_t$, $\rho_t$ and $\Lambda_t$), but also 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). Expression (7) also illustrates that two populations currently subject to the same transmission mechanisms can 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 \( \{ t, \rho_t, \Lambda_t \} \). We can
iterate equation (7) backwards to express \( \beta_t \) in terms of parameter values,

\[
\beta_t = \gamma + \rho' \Lambda \rho_t - 1 + \rho' \Lambda_t \left( \sum_{r=1}^{\infty} \left( \prod_{s=1}^{r-1} \gamma_t \Lambda_{t-s} \right) \rho_{t-r} - 1 \right), (8)
\]

where for simplicity we assumed that \( \Lambda_t \) is diagonal, 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.\(^7\) The
current level of intergenerational mobility thus depends on current and past transmission
mechanisms. If no structural changes occur, \( \xi_s = \xi \forall s \leq t \), then equation (8) simplifies to
the steady-state IGE

\[
\beta = \gamma + \rho' \Lambda \sum_{s=0}^{\infty} (\gamma \Lambda)^s \rho = \gamma + \rho' \Lambda (I_{KK} - \gamma \Lambda)^{-1} \rho, (9)
\]

where the second step follows since the geometric series \( \sum_{s=0}^{\infty} (\gamma \Lambda)^s \) converges (the absolute
value of each eigenvalue of \( \gamma \Lambda \) is below one). The literature has almost exclusively focused
on how changes in structural parameters affect the IGE in steady state, as given by (9). We
will instead analyze its transition path as determined by equations (7) and (8).

2.1 Dynamic Properties of the System

Following a structural change of the economic environment \( \xi_t \), what can we say about the
transition path of the IGE towards the new steady state? We list our main propositions here.
To provide a concise summary, we relegate most of the analytical derivations to Section 3,
before illustrating them further in Sections 4 and 5. The first establishes the existence of
prolonged mobility trends:

**Proposition 1. TRANSITIONAL DYNAMICS.** Following a structural change in the economic
environment \( \xi_t \), the intergenerational elasticity \( \beta_t \) may shift repeatedly over multiple subse-
quent generations while converging to its new steady state.

**Proof.** From equation (8), a variance-preserving change from \( \xi_{t<T} = \xi_1 \) to \( \xi_{t>T} = \xi_2 \) shifts
the IGE in generation \( T \) by \( \Delta \beta_T = \gamma_2 - \gamma_1 + (\rho'_2 \Lambda_2 - \rho'_1 \Lambda_1) (I - \gamma_1 \Lambda_1)^{-1} \rho_1 \), in generation

\(^7\)For a finite process, \( \beta_t \) will also depend on the initial condition \( Cov(e_0, y_0) \). If cross-sectional inequality
varies over generations, or if \( \Lambda_t \) is not diagonal, the derivation of equation (8) requires backward iteration of
\( Var(y_t) \) and the variance-covariance matrix of \( e_t \).
For example, a policy change may raise mobility in steady state, but it may take several generations until the new steady state is reached. Mobility shifts repeatedly, even if no other changes in the economic environment occur. The longevity of this convergence process depends on the parameters of the model. We show that certain structural changes cause longer transitions than others, and illustrate that the IGE tends to shift more strongly in the first two generations. Moreover, even adjustments that fully materialize within two generations can generate long-lasting transitional dynamics over cohorts, whose shape can be characterized further (see Proposition 5). Proposition 1 thus has important implications for the interpretation of mobility trends over time:

**Corollary to Proposition 1.** An observed shift in the IGE in generation \( t \) may be due to a structural change that occurred in a previous generation \( s < t \). Conversely, IGE trends may fail to reflect the impact of a contemporaneous structural change if they are dominated by the response to a structural change that occurred in a previous generation. Finally, this contemporaneous impact may deviate from the long-term impact on mobility.

The literature often relates observed shifts in mobility measures to recent institutional reforms or policies. However, mobility may fail to respond to an apparent change in the economic environment, or shift in response to previous structural changes, partly because past conditions can affect several consecutive generations of a given family (as illustrated in Section IV of Becker and Tomes, 1979). As such, the first-generation impact of a structural change can give a false impression of its long-term impact on mobility – in particular because the transition path can be non-monotonic:

**Proposition 2. Non-monotonicity.** The dynamic response of the IGE to a change in the economic environment can be non-monotonic. The slope of its transition path over generations may at some point change sign (“weak non-monotonicity”). In addition, its direction in the first affected generation may differ from the direction of its steady-state shift (“strong non-monotonicity”).

*Proof. Case 2 in Section 3 demonstrates weak non-monotonicity. Case 4 in Section 4 shows strong non-monotonicity and illustrates that the proposition extends to other mobility measures, such as the intergenerational correlation.*

The response of the IGE to a structural change can be v-shaped (or inversely v-shaped), such that its initial response overstates its long-term response. Moreover, under strong non-monotonicity, the initial shift in the IGE gives a qualitatively incorrect picture of the effect on long-run mobility. Non-monotonic transitions can be caused by variance changes across
generations, which directly scale the IGE (see eq. 7). However, they also arise because changes in the economic environment tend to raise mobility temporarily:

**Proposition 3. 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 response in mobility is non-monotonic if the difference between mobility in the old and new steady state is not too large.

*Proof.* See Case 3 in Section 3.

Intuitively, changes in the economic environment alter the long-run prospects of some families relative to others, such that mobility is particularly high in the generation in which this reshuffling of prospects takes place. When incomes depend on the transmission of multiple characteristics, and these characteristics are differently distributed across families, then a change in the relative importance of one characteristic has a stronger effect on some families than others. For example, if the returns to a particular skill rise then the income prospects of families in which this skill is comparatively abundant will rise. Hence, 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. Propositions 2 and 3 thus have important implications for the interpretation of mobility trends:

**Corollary to Propositions 2 and 3.** 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. In particular, a decline in mobility may reflect past gains rather than a recent deterioration of “equality of opportunity”.

### 2.2 Other Measures of Intergenerational Persistence

Apart from the intergenerational elasticity, our qualitative arguments also apply to other measures of the importance of family background, such as the intergenerational or Pearson correlation (e.g., Hertz et al., 2008), the Spearman rank correlation (Chetty et al., 2014b), or sibling correlations (Solon et al., 1991; Björklund et al., 2009). Their transitional dynamics can however differ substantially. This is particularly easy to show for the intergenerational correlation, which responds more immediately because it weights the transition path of the variance of income differently. Sibling correlations depend less directly on conditions in the parent generation and thus respond even more immediately to changes in the economic environment.
Proposition 4. TRANSITIONAL DYNAMICS IN ALTERNATIVE MOBILITY MEASURES.
Different mobility measures can exhibit different non-steady state dynamics, even when their steady-state response is similar.

Proof. For 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 $Var(y_t) \neq Var(y_{t-1})$. We illustrate this result in Section 4 and Appendix A.3. For an illustration of the sibling correlation see Appendix A.4. □

2.3 Non-steady State Dynamics over Cohorts

While the theoretical literature models transmission between generations, empirical studies estimate mobility trends over cohorts. This distinction is less relevant for steady state analysis, and has thus been ignored in the theoretical literature. But it does matter for the transitional dynamics and thus the interpretation of mobility trends. Most importantly, the effect of structural changes on mobility trends will be smoothed out by variation in the timing of fertility among parents. Mobility may shift steadily over multiple decades even when the system returns to steady state within the next generation. In contrast, sudden shifts in mobility must be due to contemporaneous events. As a consequence, the timing of fertility is key for the analysis of mobility trends, and may help to identify the contribution of past structural changes to current trends. We summarize and derive these arguments in the following proposition, and illustrate them 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. 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

$$C(t-j) = C(t) - A(t-1, C(t)) - ... - A(t-j, C(t-j+1)).$$ (10)

Denote realizations of these random variables by lower case letters and abstract from lifecycle dynamics, such that a parameter with subscript $c(t)$ represents the average economic environment over the lifecycle of cohort $c$.\textsuperscript{8} The (scalar) equivalent to equation (7) is then

$$\beta_{c(t)} = \frac{Cov(y_{c(t)}, y_{C(t-1)})}{Var(y_{C(t-1)})} = \gamma_{c(t)} + \rho_{c(t)} \lambda_{c(t)} \frac{Cov(e_{C(t-1)}, y_{C(t-1)})}{Var(y_{C(t-1)})},$$ (11)

\textsuperscript{8}For a consideration of lifecycle dynamics see for example Conlisk (1969) or Heckman and Mosso (2014).
The IGE for a given cohort depends on cohort-specific transmission mechanisms (and may thus change suddenly) and the variance and covariance of income and endowments in the parent generation. However, these moments 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 law of iterated expectations we have

\[ Cov(e_c(t-1), y_c(t-1)) = \sum_{a(t-1)} f_{c(t)}(a(t-1)) Cov(e_c(t-a(t-1)), y_c(t-a(t-1))) \],

(12)

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.\(^9\)

3 Transmission Dynamics

In this section, we derive how the IGE shifts in response to various structural changes in the economic environment \( \xi \). Our objective is twofold. First, the cases prove and illustrate our main propositions. Second, we provide economic intuition for our analytical results and relate them to the empirical literature. We present here three cases that support our first three propositions. We cover a fourth case in more detail in Section 4 and an empirical application in Section 5, which will illustrate our remaining two propositions. While our choice of cases here is necessarily selective, we provide an overview of other cases in Appendix A.3.

Assume that a structural change occurs in generation \( t = T \), such that \( \xi_{t<T} \neq \xi_{t\geq T} \), and that the moments of all variables were in steady-state equilibrium in \( t = T-1 \). It is useful to distinguish two classes of structural changes. First, an uncompensated (or absolute) change in the importance of a transmission channel, reflected by a single parameter shift. Second, a compensated (or relative) change, in which its importance relative to factors that are not transmitted within families changes such that all variances remains constant. While the latter requires shifts in at least two parameters, it leads to a more intuitive interpretation of mobility trends in later generations (the uncompensated and compensated responses are identical in

\(^9\)The covariance will vary by parental age also because of the selective nature of fertility. As we show in Section 5, such selectivity affects the transitional dynamics in response to structural changes further. A number of other interesting but less central 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 mechanisms.
generation $T$). We focus on the compensated response here and highlight how the non-steady state dynamics of inequality and mobility are intertwined in the next section.

We start with simplified versions of our baseline model, considering a single endowment $e_t$ and scalar versions of equations (5) and (6), such that

$$y_t = y_{t-1} + \rho_1 e_t + \sigma_t u_t$$  \hspace{1cm} (13)

$$e_t = \lambda_1 e_{t-1} + \phi_t v_t.$$  \hspace{1cm} (14)

Our main qualitative findings do not rely on specific parametrizations. However, we also illustrate the transitional dynamics numerically, for which we choose parameter values in line with the empirical literature and cross-validation within our model (see Appendix A.5).

**Case 1. The heritability of and returns to endowments. Assume that the heritability of endowments ($\lambda_t$) or the returns to endowments ($\rho_t$) change.**

Consider the compensated response in a simple meritocratic economy ($\gamma = 0$). Derivations for more general cases are described in Appendix A.3. A change in returns from $\rho_{t<T} = \rho_1$ to $\rho_{t\geq T} = \rho_2$ shifts the IGE in the first affected generation according to

$$\Delta \beta_T = \beta_T - \beta_{T-1} = (\rho_2 - \rho_1) \lambda \rho_1,$$  \hspace{1cm} (15)

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

$$\Delta \beta_{T+1} = \rho_2 \lambda \text{Cov}(e_T, y_T) = \rho_2 \lambda (\rho_2 - \rho_1),$$  \hspace{1cm} (16)

is induced by the change in the correlation between income and endowments among the parents of offspring generation $T+1$, in turn caused by changing returns to those endowments in generation $T$. A change in the heritability from $\lambda_{t<T} = \lambda_1$ to $\lambda_{t\geq T} = \lambda_2$ instead shifts the IGE only in the first generation,

$$\Delta \beta_T = (\lambda_2 - \lambda_1) \rho^2,$$  \hspace{1cm} (17)

and mobility remains constant afterwards. Figure 1a gives a numerical example.

A change in the heritability of endowments $\lambda$ has therefore a more immediate effect on the IGE than a change in their returns $\rho$. This observation extends to the more general case in which parental income has a causal effect on child income ($\gamma \neq 0$). While a change in $\lambda$ then affects the covariance between income and endowments and thus the IGE across multiple generations, the magnitude of these shifts declines rapidly (by the fraction $\gamma \lambda_2$, see Table A.1). In contrast, the IGE may shift more strongly in the second than in the first generation after an increase in $\rho$. To illustrate why such results matter, we relate them in Section 4 to the
evidence on growing skill premia from around 1980 in the US and other OECD countries. In that section, we also illustrate how mobility dynamics interact with dynamics in cross-sectional inequality.

We move next to the more general model that allows for parental income to have causal effects ($\gamma \neq 0$). Consider first an example of “equalizing opportunities”:\(^{10}\)

**Case 2.** **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 jobs and occupations. Mobility then shifts in the first affected generation,

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

(18)
due to both the declining importance of parental income, $\Delta \gamma = \gamma_2 - \gamma_1 < 0$, and the increasing returns to endowments, $\Delta \rho = \rho_2 - \rho_1 > 0$. 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 $\text{Cov}(e_{T-1}, y_{T-1}) < 1$. Mobility thus tends to initially increase. Mobility also shifts in the second generation,

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

(19)
due to changes in covariance between parental income and endowments, and possibly also changes in the variance of income (see Appendix A.3). The relative impact of parameter changes is now reversed, with the decline in $\gamma$ being attenuated by $\lambda \text{Cov}(e_{T-1}, y_{T-1})$. 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. Figure 1b plots a numerical example, illustrating that this transition can be long-lasting; mobility shifts become insignificant only in the third generation, or more than half a century after the structural change.

The dynamic response of the IGE thus tends to be *non-monotonic*, with an initial decline and a subsequent increase. Specifically, if the variance of income remains constant, it will be non-monotonic if $-\Delta \gamma / \Delta \rho > \lambda \text{Cov}(e_{T-1}, y_{T-1}) < -\Delta \rho / \Delta \gamma$, which holds if $\Delta \gamma$ and $\Delta \rho$ are sufficiently similar in absolute size. Whether the response is “weakly” or “strongly” non-monotonic (see Proposition 2) depends on parameter values; strong non-monotonicity is more likely when $\lambda$ is high. The insight extends to the uncompensated case (Appendix A.3).

---

\(^{10}\)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 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 offspring with high-income, low-skill parents. In contrast, it benefits talented offspring 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 offspring 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. From equation (19), mobility declines more strongly if endowments are more strongly inherited.\(^{11}\)

The example 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.

We argue that such patterns are plausible, not that they are particularly likely: transitional dynamics over such long time periods would also be subject to various behavioral responses that we do not model here (see for example Mora and Watts, 2015).

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 3. Changing returns to skills.** Assume that the returns to different types of skills or endowments change on the labor market \(\rho_1 \neq \rho_2\).

Changes in the returns to different skills could stem from changes in relative supplies (e.g., as of immigration or educational reforms) 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).

We start with 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 \(\Lambda\) is given in Appendix A.6. The steady-state IGE before the structural change is then equal to

\[
\beta_{t-1} = \gamma + \rho_1 \Lambda (I - \gamma \Lambda)^{-1} \rho_1, \tag{20}
\]

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

\[
\beta_{t\to \infty} = \gamma + \rho_2 \Lambda (I - \gamma \Lambda)^{-1} \rho_2. \tag{21}
\]

\(^{11}\)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.
The IGE in generation $T$ can then be expressed as

$$
\beta_T = \frac{1}{2} (\beta_{T-1} + \beta_{t\to\infty}) - \frac{1}{2} (\rho'_2 - \rho'_1) \Lambda (I - \gamma \Lambda)^{-1} (\rho_2 - \rho_1),
$$

(22)

where the quadratic form in the last term is greater than zero for $\rho_2 \neq \rho_1$ since $\Lambda (I - \gamma \Lambda)^{-1}$ is positive definite. The IGE in the first affected generation can therefore be decomposed into two parts, the average of the old and the new steady-state IGE (first term) plus a purely transitional gain (second term). This result supports the first part of Proposition 3. For its second part, note that changes in returns thus cause a temporary spike in mobility ($\beta_T$ is below $\beta_{T-1}$ and $\beta_{t\to\infty}$) as long as the steady-state IGE does not shift too strongly, specifically if

$$
|\beta_{t\to\infty} - \beta_{T-1}| < (\rho'_2 - \rho'_1) \Lambda (I - \gamma \Lambda)^{-1} (\rho_2 - \rho_1).
$$

(23)

This argument also holds if cross-sectional inequality is lower in the new than in the old steady state. Eq. (22) then includes the additional term $\rho'_2 \Lambda (I - \gamma \Lambda)^{-1} \rho_2 (1 - \frac{1}{\text{Var}(y_{T\to\infty})})$, which is negative if $\text{Var}(y_{T\to\infty}) < \text{Var}(y_{T-1}) = 1$.

Figure 1c illustrates 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.\footnote{We have $\Delta \beta_T = - (\rho_{k,2} - \rho_{k,1})^2 \lambda / (1 - \gamma \lambda)$, which is negative, and $\Delta \beta_{T+1} = \lambda (\rho_{k,2} - \rho_{k,1})^2 + \lambda (\rho_{k,2}^2 + \rho_{k,1}^2 + (2 \rho_{k,1} \rho_{k,2} \lambda \gamma) / (1 - \gamma \lambda) ) / (1 / \text{Var}(y_T) - 1)$, which is positive since $\text{Var}(y_T) = 1 - 2 \gamma \lambda (\rho_{k,2} - \rho_{k,1})^2 / (1 - \gamma \lambda) < 1$. These findings are not due to shifts in cross-sectional inequality; if instead $\text{Var}(y_T) = 1$ (i.e. changes in $\rho_k$ and $\rho_l$ are offset by changes in the variance of $u_t$) we still have that $\Delta \beta_T < 0$ and $\Delta \beta_{T+1} > 0$.} Intuitively, mobility initially increases because the endowment for which returns increase from low levels are less prevalent among high-income parents than the endowment for which returns decrease from high levels. But the endowment for which returns increase 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 authors have shown 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. The key assumption underlying this result is that individuals have a comparative advantage in certain skills or endowments and that these comparative advantages are partially transmitted within families.

Based on our last two cases we can formulate a more general conclusion. 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 importance of parental income relative to own skills diminishes the prospects of offspring from 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 often generate such relative gains and losses,
Figure 1: Comparative Transitional Dynamics: Numerical Examples

(a) A change in the heritability of, or returns to, endowments

(b) A declining impact of parental income and increasing returns to skills

(c) A swap in prices

Note: Numerical examples of trends in the intergenerational elasticity: (a) in generation $T$ the heritability of endowments $\lambda$ decreases from $\lambda_1 = 0.6$ to $\lambda_2 = 0.5$ (assuming $\rho = 0.7$ and $\gamma = 0$) or the returns to endowments $\rho$ increase from $\rho_1 = 0.7$ to $\rho_2 = 0.8$ (assuming $\lambda = 0.6$); (b) in generation $T$ the impact of parental income $\gamma$ declines from $\gamma_1 = 0.4$ to $\gamma_2 = 0.2$ while the returns to endowments $\rho$ increase from $\rho_1 = 0.5$ to $\rho_2 = 0.7$ (assuming $\lambda = 0.6$); and (c) 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 $\lambda = 0.6$).
generating transitional mobility in the generation in which they occur – times of change tend to be times of high mobility. The underlying logic can be extended to other contexts. 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. Our analysis suggests that such transitional gains diminish as the economic environment stabilizes.

4 Joint Dynamics of Mobility and Inequality

We noted in the previous section that the transitional dynamics of the IGE depend also on shifts in the variance of income across generations. We now consider such shifts in cross-sectional inequality, and their interrelation with intergenerational mobility more explicitly.

4.1 Transitional Dynamics in Cross-Sectional Inequality

The close steady-state relationship between income inequality between families in the same generation and across generations was emphasized already by Becker and Tomes (1979), and is central to many studies in the literature (e.g. Solon, 2002; Davies et al., 2005; Hassler et al., 2007). A key argument here is that also the non-steady state dynamics of inequality and mobility are intertwined. First, note that the existence of intergenerational transmission channels implies that a change in cross-sectional inequality may propagate across multiple generations. For example, in response to a permanent shift in the variance of market luck from $\sigma_1^2$ to $\sigma_2^2$ in generation $T$, the variance of income will initially shift by the same amount. But if $\gamma \neq 0$ it will also continue to shift in future generations, according to

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

Intuitively, if individual characteristics are linked across generations, then the inequality in these characteristics will also be linked. Such transitional dynamics in cross-sectional inequality affect aggregate measures of mobility, and affect different measures differently. A shift in 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 does shift in generation $T + 1$,

---

13For example, assume that the vector $e_t$ 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 was high in the 19th century as of exceptional geographic mobility. Our result illustrate that not only internal migration itself may matter (that depends on who migrates, see Feigenbaum, 2015), but also one of its potential causes: changes in local conditions incentivize internal migration but, by affecting parents and their (non-migrating) children differently, also increase intergenerational mobility.
according to (see Appendix A.3)

\[
\Delta \beta_{T+1} = \beta_{T+1} - \beta_T = -\frac{\rho^2 \lambda}{1 - \gamma \lambda} \frac{\sigma_2^2 - \sigma_1^2}{1 + \sigma_2^2 - \sigma_1^2}. \tag{25}
\]

Intuitively, if the role of market luck or endowments that are not heritable increases \((\sigma_2^2 > \sigma_1^2)\), then income differences among parents will eventually become less crucial for children’s incomes and the IGE declines. The IGE does therefore not respond in the first affected, but only in the second (if \(\rho > 0\)) and subsequent (if \(\gamma > 0\)) generations, providing an illustration of Proposition 1. Other mobility measures, such as the intergenerational or sibling correlation, shift already in generation \(T\), illustrating Proposition 4.

### 4.2 The Great Gatsby Curve

In recent years, the relationship between inequality and mobility has seen renewed interest, following the documentation of growing skill premia from around 1980 in the US and other OECD countries on the one hand, and the observation of a negative correlation between inequality and mobility within (Chetty et al., 2014a, Mora et al., 2016) and across countries (often dubbed the “Great Gatsby Curve”, see Corak, 2013) on the other. One puzzling observation in this context is that, despite the statistical link between inequality and mobility, and the theoretical prediction from standard models that rising skill premia decrease intergenerational mobility (Solon, 2004), its level in the U.S. seems to have remained fairly constant in recent decades (e.g. Lee and Solon, 2009; Chetty et al., 2014b).

The observation that past events may still affect contemporaneous mobility trends (Proposition 1) provides one potential explanation, and we present here two other arguments that have received little attention. First, in a model with multiple skills, higher skill premia do not necessarily lead to lower steady-state mobility. Second, the relationship between inequality and mobility in transition may deviate strongly from their steady-state relationship. To illustrate these arguments, consider the following example:

**Case 4. TRANSITIONAL DYNAMICS IN THE “GREAT GATSBY CURVE”**

Assume that in a model with multiple endowments, the returns to one endowment increase, while the returns to other endowments remain unchanged.

Assume that children inherit two uncorrelated endowments \(k\) and \(l\) from their parents. Figure 2 plots, for two different economic environments (“countries”), the transition paths of the

---

14 Krueger (2012) 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.
IGE, the intergenerational correlation, and the variance of income.\textsuperscript{15} While otherwise characterized by the same environment, the returns to endowment $k$ and thus inequality are higher in country B. An increase in the returns to endowment $l$ increases inequality in both countries, but the steady-state shift in mobility differs. In country A, higher returns lead indeed to higher intergenerational persistence. In contrast, the same structural change \textit{decreases} steady-state persistence in country B, in which the returns to endowment $k$ are higher. Such differential pattern can also arise when the heritability of an endowment varies between countries.

More generally, in a setting with multiple skills, an increase in the returns to a single endowment decreases steady-state mobility only if its heritability is high compared to other determinants of income. This observation modifies the prediction from standard models with a single skill, in which increasing returns have a clear negative effect (Solon, 2004). It also suggests that the same type of structural change can have opposing effects if countries differ in other aspects. For example, rising skill premia are more likely to increase mobility in environments in which the role of market luck is small.

Moreover, the transition path that countries take through the Gatsby diagram is not a straight line. While the path of inequality is monotonic, the paths of the two mobility measures are both non-monotonic and of opposing signs. In some cases, the shift in the first affected generation and in steady state have different signs (“strong monotonicity”, see Proposition 2). Even if the steady-state response to increased returns is in line with the static “Great Gatsby Curve”, the transition path may not be. The Curve may thus also reflect – or be attenuated by – transitional dynamics. An improved understanding of such dynamics is thus useful not only for the interpretation of mobility trends, but also for the discourse on their relationship to cross-sectional inequality.

For example, Figure 2 illustrates that the effect of changing returns on \textit{steady-state} intergenerational mobility may not become evident before both the parent and child generations have experienced those higher returns.\textsuperscript{16} It also shows that different measures of mobility show different transitional dynamics (see Proposition 4). The latter argument may help to explain why Levine and Mazumder (2007) find a sharp increase in \textit{sibling} correlations since 1980, while there is less evidence of an increase in \textit{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.4).

\textsuperscript{15}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.

\textsuperscript{16}Recent studies do not observe such cohorts; for example, the last cohort observed in Lee and Solon (2009) were born in 1975, and skill premia did not yet increase in their parents’ early careers. However, Figure 2 shows that its initial shift should \textit{overstate} the steady-state response in the IGE. Its observed stability is thus still remarkable. The figure also illustrates that the lack of an increase in correlation coefficients is less surprising.
Figure 2: An Increase in the Returns to a Single Endowment

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 $\gamma = 0.2$, $\lambda_k = 0.3$, $\lambda_I = 0.8$, $\text{Var}(\epsilon_t) = 0.5$, and $\rho_I = 0.4$ (country A) or $\rho_I = 0.8$ (country B). In generation $T$, the returns to endowment $k$ increase from $\rho_{k,1} = 0.3$ to $\rho_{k,2} = 0.8$ in both countries. See Appendix A.4 for a corresponding numerical illustration of the dynamics of the sibling correlation.

5 Empirical Illustration: A Compulsory School Reform

Our theoretical analysis suggests that a single structural change can generate trends in intergenerational mobility across multiple generations. We now examine if such long-lasting dynamics can be observed empirically, an objective that is demanding in terms of both data coverage and identification. We consider the Swedish compulsory school reform, first studied by Meghir and Palme (2005) and carefully outlined in Holmlund (2007). Gradually implemented across municipalities from the late 1940s, the reform raised compulsory schooling from seven (or eight) to nine years and postponed tracking decisions (see Appendix A.8 for further 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 evidence indicate that they did raise mobility in the directly affected generation (Holmlund, 2008; Pekkarinen et al., 2009). Second, we have access to an unusually rich data set, covering 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.\footnote{A number of studies exploit this characteristic to assess the reform impact on individual outcomes in di-}
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.9.

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.\footnote{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.}\footnote{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.7). 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.} 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.\footnote{We thank an anonymous referee for this suggestion.}

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 appears larger and the subsequent rebound occurs somewhat earlier in their study (in line with our expectations from Appendix A.7).

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.\footnote{We thank an anonymous referee for this suggestion.} In a first step consider, for each cohort $c$ and municipality $m$, the regression model

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

(26)

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 $\beta_{cm}$ a measure of intergenerational persistence (e.g. the IGE). Our interest centers on the second-step model

$$\beta_{cm} = \alpha'_{1} D_{c} + \alpha'_{2} D_{m} + \gamma R_{cm} + v_{cm},$$

(27)
which allows for mobility differences across cohorts and municipalities (captured by indicator vectors $D_c$ and $D_m$). The indicator $R_{cm}$ equals one if the reform was in place for cohort $c$ in municipality $m$, and $\gamma$ captures the reform effect.

We estimate this reform effect in both the first affected and the subsequent generation.\(^\text{21}\) 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 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, this strategy is still susceptible to differences in area-specific trends. Moreover, the reform indicator is measured with error. We address both issues and provide further robustness tests in Appendix A.10.

Panel A of Table 1 reports estimates of the reform effect $\gamma$ on the intergenerational coefficient in years of schooling and log income (the IGE).\(^\text{22}\) Upon introduction of the reform,

\(^{21}\)As the dependent variable in equation (27) is estimated, its sampling distribution needs to be taken into account to obtain standard errors and efficient estimates of $\gamma$ (see Hanushek, 1974). In practice, we estimate both steps at once, pooling across cohorts and municipalities, and interacting the intercept and regressor of equation (26) with each of the regressors in the second-stage equation.

\(^{22}\)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.
### Table 1: Reform Effect on Educational and Income Mobility

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

Note: The table reports estimates of $\gamma$ in equation (27) 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, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The degree persistence of inequalities in both schooling and income decreased by about ten percent. In line with Holmlund (2008), we conclude 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 cohorts born in the late 1960s, the first cohorts in which some fathers had attended reformed schools. Our estimates in Table 1 suggest that the persistence in both schooling and income indeed increased in response to the reform in the previous generation. The estimates are robust to a number of sensitivity tests (see Appendix A.10).

For two reasons the estimates of $\gamma$ are larger for the second than for the first generation. First, the timing of fertility (see Section 2.3): 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.

(see Hertz, 2008).
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 1) 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: recent mobility trends can indeed be caused by events that occurred in previous generations (Proposition 1). In our data we can track mobility measures only up to the 1972 cohort, but the share of fathers subject to the reform continues to rise long after that (see Figure 3). Unless dominated by other events, mobility may thus transition for several decades after our records end (see Proposition 5). 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 results. First, changes in the economic environment affect mobility not only in the directly affected but also in subsequent generations; policy or institutional reforms 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 in fact 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 the effect of a Swedish compulsory schooling reform on parent-child mobility across multiple generations. As expected (Holmlund, 2008), the reform increased income and educational mobility in directly affected cohorts. 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. Due to variation in the timing of fertility, these transitional dynamics are more complex and long-lasting than the dynamics considered in our theoretical discussion.

Our model is of course highly 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.

With these drawbacks in mind, we noted examples in which a consideration of transitional dynamics may be fruitful. We offered a new perspective on the observations that, despite rising returns to skill, U.S. mobility appears to have remained fairly stable in recent decades (Lee and Solon, 2009, Chetty et al., 2014b), or that both geographic and intergenerational mobility were comparatively high in the late 19th century (Long and Ferrie, 2013). On the former, we showed that rising skill premia shift intergenerational measures over at least two generations, suggesting that the overall shift may not yet be fully visible in current data. Other measures respond more quickly, potentially explaining why sibling correlations in earnings did increase (Levine and Mazumder, 2007). On the latter, we noted that factors that trigger geographic mobility may also generate temporary gains in intergenerational mobility.

In Europe, the literature has found a substantial increase in mobility among cohorts born in the early or mid 20th century (Pekkala and Lucas, 2007; Bratberg et al., 2007; Lefranc, 2011), while mobility appears to have declined in more recent cohorts in the U.K. (Blanden et al., 2004). An interesting finding in this context is that a shift to a more meritocratic environment, in which the importance of own skills relative to parental status rises, can generate a non-monotonic response – a mobility gain in the first affected generation, followed by a long-lasting negative trend.

Promising avenues for future empirical research include the observation that different causes of mobility shifts or different assumptions regarding the transmission framework (such as multidimensional endowments) could be distinguished by their divergent dynamic implications; that the conditioning of mobility measures on parental age at birth may provide an indication of the effect of past events on current mobility trends; or that the covariance between income, skills and endowments in the parent generation plays a central role for the evolution of income mobility over generations.
References


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 \(I_{1,t-1}, \ldots, I_{J,t-1}\) 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 \(K \times 1\) vector of inherited endowments \(e_t\), and chance \(u_{j,t}\),

\[
h_{j,t} = \gamma_j \log I_{j,t-1} + \theta_j' e_t + u_{j,t} \quad \forall j \in 1, \ldots, J, \tag{A.2}
\]

where \(\gamma_j\) and elements of the vector \(\theta_j\) 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

\[
e_k,t = \lambda_k e_k,t-1 + v_k,t \quad \forall k \in 1, \ldots, K, \tag{A.3}
\]

where \(v_k,t\) is a white-noise error term and the heritability coefficient \(\lambda_k\) 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

\[
\log Y_t = \begin{cases} 
\delta' h_t + u_{y,t} & \text{with probability } p \\
\delta' E[h_t|Y_{t-1}] + u_{y,t} & \text{with probability } 1 - p
\end{cases} \tag{A.4}
\]

With probability \(p\) employers observe human capital of workers and pay them their marginal product \(\delta' h_t\) 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 \( k \in 1, \ldots, K \).

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

\[
U_{t-1} = (1 - \alpha) \log C_{t-1} + \alpha E[\log Y_t|Y_{t-1}, I_{t-1}, e_t],
\]

(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

\[
\mathcal{L}(C_{t-1}, I_{t-1}, \mu) = (1 - \alpha) \log C_{t-1} + \alpha \delta'(p E[h_t|Y_{t-1}, I_{t-1}, e_t] + (1 - p) E[h_t|Y_{t-1}]) + \mu ((1 - \tau) Y_{t-1} - C_{t-1} - 1' I_{t-1})
\]

The first-order conditions require that

\[
\frac{\partial \mathcal{L}}{\partial C_{t-1}} = \frac{1 - \alpha}{C_{t-1}} - \mu = 0,
\]

\[
\frac{\partial \mathcal{L}}{\partial I_{j,t-1}} = \frac{\alpha (1 - p) \delta_j \gamma_j}{I_{j,t-1}} - \mu = 0 \ \forall j \in 1, \ldots, J,
\]

\[
\frac{\partial \mathcal{L}}{\partial \mu} = (1 - \tau) Y_{t-1} - C_{t-1} - 1' I_{t-1} = 0.
\]

Optimal investments,

\[
I_{j,t-1} = \frac{\alpha p \delta_j \gamma_j}{(1 - \alpha) + \sum_{t=1}^{J} \alpha p \delta_j \gamma_j} (1 - \tau) Y_{t-1} \ \forall j \in 1, \ldots, J,
\]

(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

\[
\begin{pmatrix}
  y_t \\
  e_t
\end{pmatrix} =
\begin{pmatrix}
  \gamma_t & \rho_t' \Delta_t \\
  0 & \Lambda_t
\end{pmatrix}
\begin{pmatrix}
  y_{t-1} \\
  e_{t-1}
\end{pmatrix} +
\begin{pmatrix}
  \sigma_t u_t + \rho_t' \Phi_t v_t \\
  v_t
\end{pmatrix},
\]

(A.7)

which we may shorten to

\[
x_t = A_t x_{t-1} + w_t.
\]

(A.8)
Let subscripts 1, 2 index parameter values before and after a structural shock occurs in generation $T$. The stability condition $\lim_{s \to \infty} A_s^2 = 0$ is then satisfied by assuming that $\gamma_t$ and all eigenvalues of $\Lambda_2$ are non-negative and below one. For example, if $\Lambda_2$ is diagonal and elements of the endowment vector $e_t$ are uncorrelated then the diagonal elements of $\Lambda_2$ are required to be strictly between zero and one. Normalization of the variances of $y_t$ and elements of $e_t$ 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 $S_t$ and $W_t$, such that

$$S_t = A_t S_{t-1} A_t' + W_t.$$  

Denote by $\gamma$, $\rho$, and $\Lambda$ the steady-state parameter values before a structural change occurs in generation $t = T$. Note that in steady state $S_t = S_{t-1} = S$, normalize all diagonal elements of $S$ to one, and solve for the elements of $W_t$. For example, if $\Lambda_t$ is diagonal then $\text{Var}(e_{jt}) = 1 \forall j$ iff $\text{Var}(v_{jt}) = 1 - \lambda_j^2 \forall j$; 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

From equation (7), the transition path of the IGE depends in the first affected generation $T$ only on the transition path of the covariance of parent and offspring income,

$$\text{Cov}(y_t, y_{t-1}) = \gamma_t \text{Var}(y_{t-1}) + \rho_t \Lambda_t \text{Cov}(e_{t-1}, y_{t-1}),$$  \hspace{1cm} (A.9)

but in later generations also on the covariances and variances of income and endowments,

$$\text{Cov}(e_t, y_t) = \gamma_t \Lambda_t \text{Cov}(e_{t-1}, y_{t-1}) + \text{Var}(e_t) \rho_t \hspace{1cm} (A.10)$$

$$\text{Var}(y_t) = \gamma_t^2 \text{Var}(y_{t-1}) + \rho_t^2 \text{Var}(e_t) \rho_t + \sigma_t^2 + 2 \gamma_t \rho_t \Lambda_t \text{Cov}(e_{t-1}, y_{t-1}) \hspace{1cm} (A.11)$$

$$\text{Var}(e_t) = \Lambda_t \text{Var}(e_{t-1}) \Lambda_t + \Phi_t \Phi_t. \hspace{1cm} (A.12)$$

We here describe the transition paths for a number of structural changes, including those discussed in the main text. The shift of the IGE from the second generation onwards in response to a change in $\xi_{t<T} = \xi_1$ to $\xi_{t\geq T} = \xi_2$ can have a complicated expression. Using

$$\Delta \beta_{T+1} = \frac{\Delta \text{Cov}(y_{T+1}, y_T) - \beta_T \Delta \text{Var}(y_T)}{1 + \Delta \text{Var}(y_T)}$$  \hspace{1cm} (A.13)

we thus decompose it here in the transition path of its underlying moments $\Delta \text{Cov}(e_T, y_T)$, $\Delta \text{Var}(y_T)$, and $\Delta \text{Var}(e_T)$, illustrating the various contributing forces. In the main text we focus frequently on the compensated response in which $\Delta \text{Var}(y_T) = \Delta \text{Var}(e_T) = 0$.

---

\(^{23}\)Conlisk (1974b) derives stability conditions in a random coefficients model with repeated shocks.
Table A.1: Transitional Dynamics in Components of the IGE

<table>
<thead>
<tr>
<th></th>
<th>$\Delta\text{Cov}(y_T, y_{T-1})$</th>
<th>$\Delta\text{Cov}(y_{T+1}, y_T)$</th>
<th>$\Delta\text{Var}(y_T)$</th>
<th>$\Delta\text{Cov}(e_T, y_T)$</th>
<th>$\Delta\text{Var}(e_T)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scalar Model:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1 \to \sigma_2$</td>
<td>0</td>
<td>0</td>
<td>$\sigma_2^2 - \sigma_1^2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\rho_1 \to \rho_2$</td>
<td>$\frac{(\rho_2 - \rho_1)\lambda\rho_1}{1 - \gamma\lambda}$</td>
<td>$\gamma\Delta\text{Var}(y_T)$ + $\frac{2\gamma(\rho_2 - \rho_1)\lambda\rho_1}{1 - \gamma\lambda}$</td>
<td>$\rho_2 - \rho_1$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\lambda_1 \to \lambda_2$</td>
<td>$\frac{(\lambda_2 - \lambda_1)\rho_2}{1 - \gamma\lambda}$</td>
<td>$\gamma\Delta\text{Var}(y_T)$ + $\rho_2\lambda\Delta\text{Cov}(e_T, y_T)$ + $2\gamma(\lambda_2 - \lambda_1)\rho_2^2/(1 - \gamma\lambda)$</td>
<td>$\gamma(\lambda_2 - \lambda_1)\rho/(1 - \gamma\lambda)$</td>
<td>$\lambda_2^2 - \lambda_1^2$</td>
<td></td>
</tr>
<tr>
<td>$\gamma_1 \to \gamma_2$</td>
<td>$\gamma_2 - \gamma_1$</td>
<td>$\rho_2\Delta\text{Cov}(e_T, y_T)$ + $2(\gamma_2 - \gamma_1)\lambda\rho_2^2/(1 - \gamma\lambda)$</td>
<td>$\frac{(\gamma_2 - \gamma_1)\lambda\rho}{1 - \gamma\lambda}$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\gamma_1 \to \gamma_2$ and $\rho_1 \to \rho_2$</td>
<td>$\frac{\gamma_1(\rho_2 - \rho_1)\lambda\rho_1}{1 - \gamma\lambda}$</td>
<td>$\rho_2\Delta\text{Cov}(e_T, y_T)$ + $2(\gamma_2 - \gamma_1)\lambda\rho_2^2/(1 - \gamma\lambda)$</td>
<td>$\frac{(\gamma_2 - \gamma_1)\lambda\rho}{1 - \gamma\lambda}$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Multiplicity Model:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1 \to \rho_2$</td>
<td>$(\rho_2 - \rho_1)\Lambda(I - \gamma\Lambda)^{-1}\rho_1$</td>
<td>$\gamma\Delta\text{Var}(y_T)$ + $\rho_2\Delta\text{Cov}(e_T, y_T)$ + $2\gamma(\rho_2 - \rho_1)\Lambda(I - \gamma\Lambda)^{-1}\rho_1$</td>
<td>$\rho_2 - \rho_1\text{Var}(e_{T-1})$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\Lambda_1 \to \Lambda_2$</td>
<td>$\rho'(\Lambda_2 - \Lambda_1)(I - \gamma\Lambda_1)^{-1}\rho_1$</td>
<td>$\gamma\Delta\text{Var}(y_T)$ + $\rho'\Lambda_2\Delta\text{Cov}(e_T, y_T)$ + $2\gamma\rho'(\Lambda_2 - \Lambda_1)(I - \gamma\Lambda_1)^{-1}\rho_1$</td>
<td>$\gamma(\Lambda_2 - \Lambda_1)(I - \gamma\Lambda_1)^{-1}\rho_1$</td>
<td>$(\Lambda_2 - \Lambda_1)(\Lambda_2 + \Lambda_1)$</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table reports the change in intergenerational and cross-sectional moments (columns (i)-(v)) in the first two generations after a specific structural change (left column) occurs in generation $T$. In the main analysis we focus on the compensated response, in which $\Delta\text{Var}(y_T) = \Delta\text{Var}(e_T) = 0$, such that $\Delta\beta_{T+1} = \Delta\text{Cov}(y_T, y_{T-1})$. 


A.4 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 $\rho$, as in Case 1 and Case 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

\[
y_{tij} = \gamma_t y_{t-1,j} + \rho_t e_{tij} + u_{tij} \tag{A.14}
\]
\[
e_{tij} = \lambda_t e_{t-1,j} + \phi_t v_{tij}. \tag{A.15}
\]

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. $u_{tij} = c_{tij} + \tilde{u}_{tij}$. 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 $\text{Var}(y_{tij}) = \text{Var}(a_{tij}) + \text{Var}(b_{tij})$, where $\text{Var}(a_{tij})$ denotes the variance between families whereas $\text{Var}(b_{tij})$ denotes the variance across individuals within families. Specifically, in our model (suppressing the $i$ and $j$ subscripts)

\[
r_t^S = \frac{\text{Var}(a_t)}{\text{Var}(a_t) + \text{Var}(b_t)} = \frac{\text{Var}(\gamma_t y_{t-1} + \rho_t \lambda_t e_{t-1} + c_t)}{\text{Var}(\gamma_t y_{t-1} + \rho_t \lambda_t e_{t-1} + c_t + \rho_t \phi_t v_t + \tilde{u}_t)}, \tag{A.16}
\]

Focus first on the compensated response, such that $\text{Var}(y_t)$ 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

\[
\Delta r_T^S = r_T^S - r_{T-1}^S = (\rho_2^2 - \rho_1^2) \lambda^2 + \frac{2\gamma \lambda \rho_1 (\rho_2 - \rho_1)}{1 - \gamma \lambda}, \tag{A.17}
\]

and in the second generation according to

\[
\Delta r_{T+1}^S = 2\gamma \rho_2 \lambda \text{Cov}(e_T, y_T) = 2\gamma \lambda \rho_2 (\rho_2 - \rho_1), \tag{A.18}
\]

In a simple meritocratic economy ($\gamma = 0$), the sibling correlation only shifts in the first affected generation and is unaffected thereafter. In contrast, equation (16) 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 $\gamma$, which is
likely to be small, while the second-generation shift of the IGE does not.\footnote{The empirical literature rarely finds \textit{bivariate 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.} The sibling correlation responds therefore more immediately (supporting Proposition 4).

We finally consider the dynamics of the sibling correlation in the model with multiple skills and non-constant variances, as in Case 4 in Section 4. Figure A.1, based on the same parametrization as “Country A” in Figure 2, confirms that the IGE and the sibling correlation tend to exhibit quite different transition paths to their new steady states also in this more general model. As for the intergenerational measures, its transition path 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 $\gamma$) 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.16)) and may thus decrease if the endowment’s heritability is not similarly high.
A.5 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 $h_t$ 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 $\gamma$) 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 $\gamma_y$ and $\gamma_h$, 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.\(^{25}\) Those estimates may reflect to various degrees not only genetic inheritance but also correlated environmental factors; we capture both in the heritability parameter $\lambda$ (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 $\rho$ 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.\(^{26}\) 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

\(^{25}\)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. (2010).

\(^{26}\)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.
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 $\gamma_h$ (and its contribution to offspring income in $\gamma$). In any case, values of $\rho$ in the range of 0.6-0.8 should be at least roughly consistent with the empirical evidence.\footnote{In the initial steady state we standardize $\text{Var}(y) = \text{Var}(e) = 1$, such that $R^2 = 0.4$ translates into $\rho \approx 0.63$.}

These parameter ranges are consistent with recent estimates of the intergenerational income elasticity $\beta$ 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,

$$
\beta = \gamma + \frac{\rho^2 \lambda}{1 - \gamma \lambda} \quad \gamma = \frac{\beta \lambda + 1 \pm \sqrt{\beta^2 \lambda^2 - 2 \beta \lambda + 4 \lambda^2 \rho^2 + 1}}{2 \lambda}
$$

$$
\rho = \sqrt{\frac{(\beta - \gamma)(1 - \gamma \lambda)}{\lambda}} \quad \lambda = \frac{\beta - \gamma}{\beta \gamma + \rho^2 - \gamma^2},
$$

and plug in the discussed values on the right-hand sides to impute parameter ranges that are consistent with our reading of the empirical literature. Specifically we rule out too high values of $\lambda$ and $\rho$ as they cause $\gamma$ 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.6 Correlated endowments

We revisit Case 3 under the assumption that $\Lambda_t$ is not diagonal, such that elements of the endowment vector $e_t$ are potentially correlated. Suppose that at generation $T$ the returns to human capital change from $\rho_1$ to $\rho_2$ 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

$$
\beta_{T-1} = \text{Cov}(y_{T-1}, y_{T-2}) = \gamma + \rho_1' \text{Cov}(e_{T-1}, y_{T-2}) = \gamma + \rho_1' \Gamma \rho_1
$$

(A.20)
where
\[
\Gamma = \sum_{l=1}^{\infty} \gamma^{l-1} \text{Cov}(e_{T-1}, e_{T-1-l})
\]  
(A.21)

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 \( \Gamma \) 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
\[
\beta_T = \gamma + \rho_1' \Gamma \rho_1
\]  
(A.22)
\[
\beta_{t \to \infty} = \gamma + \rho_2' \Gamma \rho_2.
\]  
(A.23)

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 \( \text{Cov}(e_{T1}, e_{Tj}) \) \( \forall j > 1 \) 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
\[
2\beta_T = 2(\gamma + \rho_2' \Gamma \rho_1)
\]  
\[
= \gamma + \rho_1' \Gamma \rho_1 + (\rho_2' - \rho_1') \Gamma \rho_1 + \gamma + \rho_2' \Gamma \rho_2
\]  
\[
= \beta_{T-1} + \beta_{t \to \infty} + (\rho_2' - \rho_1') \Gamma \rho_1 - \rho_2' \Gamma (\rho_2 - \rho_1)
\]  
\[
= \beta_{T-1} + \beta_{t \to \infty} - (\rho_2' - \rho_1') \Gamma (\rho_2 - \rho_1),
\]  
(A.24)

where we expanded and subtracted \( \rho_1' \) and \( \rho_2 \), substituted equations (A.20) and (A.23), and finally took the transpose and used the symmetry of \( \Gamma \) 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 \( \Gamma_S \) and \( \Lambda_S \) the minors of \( \Gamma \) and \( \Lambda \) that are formed by deleting each row and column that correspond to an element in \( S \). The quadratic form \( (\rho_2' - \rho_1') \Gamma (\rho_2 - \rho_1) \) is greater than zero for \( \rho_2 \neq \rho_1 \) if \( \Gamma_S \) is positive definite. A sufficient condition for \( \Gamma_S \) to be positive definite is diagonality of the heritability matrix \( \Lambda_S \), with positive diagonal elements. More generally, the matrix \( \Gamma_S \) is positive definite if the respective minors of the cross-covariances \( \text{Cov}(e_{T-1}, e_{T-j}) \) \( \forall j > 1 \) 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
Price changes then increase intergenerational mobility temporarily ($\beta_T$ is below both the previous steady state $\beta_{T-1}$ and the new steady state $\beta_{t \rightarrow \infty}$) as long as the steady-state elasticity shifts not too strongly, specifically iff

$$|\beta_{t \rightarrow \infty} - \beta_{T-1}| < (\rho_2' - \rho_1') \Lambda (I - \gamma \Lambda)^{-1} (\rho_2 - \rho_1).$$

(A.25)

### A.7 Compulsory Schooling in the Intergenerational Model

To predict the impact of a compulsory schooling policy on educational and income mobility, first include constants $\alpha_y$ and $\alpha_h$ 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 schooling and thus schooling decisions. 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

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

(A.26)

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

$$\beta_{inc,t} = \frac{\text{Cov}(y_t, y_{t-1})}{\text{Var}(y_{t-1})} \quad \text{and} \quad \beta_{edu,t} = \frac{\text{Cov}(h_t, h_{t-1})}{\text{Var}(h_{t-1})}. \quad \text{(A.27)}$$

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.2 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

---

28 Spillover effects on educational attainment of individuals not directly affected by the reform were found to be small in Holmlund (2007).
Figure A.2: Raising the compulsory schooling level

(a) Intergenerational educational coefficient $\beta_{edu}$

(b) Intergenerational income elasticity $\beta_{inc}$

Note: Income and educational mobility trends in numerical example, with $x = 9$, $\alpha_y = 9$, $\gamma_y = 0$, $\delta = 0.2$ (dashed line: $\delta = 0.18$), $\alpha_h = 10$, $\gamma_h = 1$, $\theta = 2$, $\lambda = 0.6$, and $(u_y, u_h, v)$ normally distributed with variances $(0.1, 2.75, 0.64)$.

and second moments for income $y_t$ and schooling $h_t$ that are similar to the observed moments in the Swedish data.

Panel A plots the response of the intergenerational educational coefficient $\beta_{edu}$. In offspring generation $T$ the reform compresses the variance of schooling strongly, which decreases the numerator of $\beta_{edu}$ – 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 $\beta_{edu}$ 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 $\delta$), 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 $\beta_{inc}$ tends to be similar but differs in some aspects (Panel B in Figure A.2). 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.8 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.9 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. 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 cova-

29Educational 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.

A.13
Table A.2: Sample Statistics by Birth Cohort

<table>
<thead>
<tr>
<th>Source data</th>
<th>Intergenerational samples</th>
</tr>
</thead>
<tbody>
<tr>
<td># obs.</td>
<td># obs. with non-missing reform shares</td>
</tr>
<tr>
<td>(offspring)</td>
<td>(educ.) (inc.) (offspring) (fathers)</td>
</tr>
<tr>
<td>1943</td>
<td>42,138 0.04 17,211 15,008 11,059 0.04</td>
</tr>
<tr>
<td>1944</td>
<td>44,715 0.06 18,425 16,179 14,016 0.06</td>
</tr>
<tr>
<td>1945</td>
<td>44,682 0.06 18,604 16,441 15,984 0.07</td>
</tr>
<tr>
<td>1946</td>
<td>44,299 0.11 19,124 17,101 16,800 0.11</td>
</tr>
<tr>
<td>1947</td>
<td>43,288 0.18 19,078 17,103 16,775 0.18</td>
</tr>
<tr>
<td>1948</td>
<td>42,527 0.31 19,063 17,192 16,881 0.31</td>
</tr>
<tr>
<td>1949</td>
<td>40,628 0.39 19,449 16,768 16,424 0.40</td>
</tr>
<tr>
<td>1950</td>
<td>38,854 0.53 19,421 17,657 17,288 0.54</td>
</tr>
<tr>
<td>1951</td>
<td>36,951 0.56 18,644 17,016 16,693 0.57</td>
</tr>
<tr>
<td>1952</td>
<td>37,031 0.69 19,102 17,442 17,085 0.70</td>
</tr>
<tr>
<td>1953</td>
<td>37,537 0.79 19,452 17,904 17,565 0.80</td>
</tr>
<tr>
<td>1954</td>
<td>35,668 0.86 18,453 16,955 16,589 0.87</td>
</tr>
<tr>
<td>1955</td>
<td>36,440 0.95 19,122 17,569 17,179 0.96</td>
</tr>
<tr>
<td>1956</td>
<td>36,666 1.00 20,942 19,217 18,714 1.00</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1965</td>
<td>42,909 1.00 28,447 26,762 24,657 1.00</td>
</tr>
<tr>
<td>1966</td>
<td>43,050 1.00 29,043 27,415 25,166 1.00</td>
</tr>
<tr>
<td>1967</td>
<td>42,686 1.00 28,897 27,366 25,177 1.00</td>
</tr>
<tr>
<td>1968</td>
<td>54,105 1.00 33,526 32,524 30,124 1.00</td>
</tr>
<tr>
<td>1969</td>
<td>52,317 1.00 32,157 31,315 28,924 1.00</td>
</tr>
<tr>
<td>1970</td>
<td>53,908 1.00 32,508 31,788 29,195 1.00</td>
</tr>
<tr>
<td>1971</td>
<td>56,493 1.00 33,251 32,539 29,783 1.00</td>
</tr>
<tr>
<td>1972</td>
<td>57,035 1.00 33,081 32,409 29,472 1.00</td>
</tr>
</tbody>
</table>

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

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 we use income ranks by birth year.

While we present evidence on mobility in father-child pairs, the consideration of max-
imum 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). The gradual implementation of the reform affected cohorts born between 1938 and 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).

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.

The reform had a direct impact on educational attainment, which can be also measured with high precision over long time intervals. Figure A.3 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...
Figure A.3: Mean and Variance of Years of Schooling over Cohorts

(A) Mean:

(B) Variance:

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

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.4 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. But its impact shrinks over cohorts, and becomes indistinguishable from zero after the 1951 cohort. The reason becomes clear from Figure A.3. 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.10 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 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.
Table A.3: Robustness Tests

<table>
<thead>
<tr>
<th></th>
<th>baseline</th>
<th>with siblings</th>
<th>fathers below 30 dropped</th>
<th>pre-reform max.</th>
<th>parental controls</th>
<th>cohort time trends</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st gen.</td>
<td>-0.0371***</td>
<td>-0.0393***</td>
<td>-0.0408***</td>
<td>-0.0434***</td>
<td>-0.0357***</td>
<td>-0.0387***</td>
</tr>
<tr>
<td></td>
<td>(0.0072)</td>
<td>(0.0054)</td>
<td>(0.0089)</td>
<td>(0.0083)</td>
<td>(0.0064)</td>
<td>(0.0073)</td>
</tr>
<tr>
<td>2nd gen.</td>
<td>0.0655***</td>
<td>0.0651***</td>
<td>0.0655***</td>
<td>0.0710***</td>
<td>0.0307***</td>
<td>0.0655***</td>
</tr>
<tr>
<td></td>
<td>(0.0128)</td>
<td>(0.0122)</td>
<td>(0.0128)</td>
<td>(0.0139)</td>
<td>(0.0093)</td>
<td>(0.0126)</td>
</tr>
<tr>
<td>Income:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st gen.</td>
<td>-0.0196*</td>
<td>-0.0078</td>
<td>-0.0181</td>
<td>-0.0195*</td>
<td>-0.0210**</td>
<td>-0.0233**</td>
</tr>
<tr>
<td></td>
<td>(0.0100)</td>
<td>(0.0068)</td>
<td>(0.0115)</td>
<td>(0.0118)</td>
<td>(0.0088)</td>
<td>(0.0095)</td>
</tr>
<tr>
<td>2nd gen.</td>
<td>0.0410*</td>
<td>0.0148</td>
<td>0.0410*</td>
<td>0.0492**</td>
<td>0.0344**</td>
<td>0.0418**</td>
</tr>
<tr>
<td></td>
<td>(0.0216)</td>
<td>(0.0165)</td>
<td>(0.0216)</td>
<td>(0.0238)</td>
<td>(0.0155)</td>
<td>(0.0212)</td>
</tr>
</tbody>
</table>

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 offspring born in the cohort preceding 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.

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 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.4 – 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.5: Placebo Test: Second Generation

![Figure A.5: Placebo Test: Second Generation](image)

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.5. 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.4 – 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).