Mobility Across Multiple Generations: The Iterated Regression Fallacy

Jan Stuhler[*†]

11th June 2014

Abstract

Empirical evidence on the degree of *long-run* mobility across multiple generations is scarce. Predictions are instead routinely derived by exponentiation of intergenerational measures, implying high long-run mobility even when intergenerational mobility is low. Such extrapolations however presume that regression implies iterated regression, a statistical fallacy whose prevalence I briefly discuss. I then examine how elements of the transmission process affect the relation between intergenerational and multigenerational mobility. Considering direct and indirect transmission, the multiplicity of skills, and the role of grandparents I conclude that long-run mobility will likely be lower, possibly much lower, than predictions from intergenerational evidence suggest. I support these theoretical predictions with evidence from Swedish registries that cover three generations.

[*]University College London, Department of Economics and Centre for Research and Analysis of Migration (email: j.stuhler@ucl.ac.uk)

[†]I am grateful for support from the German National Academic Foundation and the Centre for Research and Analysis of Migration at University College London. I thank Marco Alfano, Manudeep Bhuller, Anders Björklund, Steven Durlauf, Christian Dustmann, Martin Dribe, Simon Gorlach, Cristina Gualdani, Toru Kitagawa, Claus Thustrup Kreiner, Mikael Lindahl, Magne Mogstad, Martin Nybom, Anna Raute, Kjell Salvanes, Uta Schönberg, Gary Solon, Yanos Zylberberg, and seminar participants at University College London for valuable comments. An earlier version of this paper was published as IZA Discussion Paper No. 7072.

Introduction

A vast empirical literature has estimated the degree of intergenerational persistence in socio-economic characteristics between parents and their children. There exists however much less evidence on the degree of *long-run* mobility across *multiple* generations, such as between grandparents and their grandchildren. In its absence we rely on extrapolations from parent-child correlations. For example, Hertz (2006) reports an intergenerational income elasticity of 0.47 for the United States and proceeds to note:

> *"To understand what these statistics mean, consider a rich and a poor family in the United States [...] and ask how much of the difference in the parents' incomes would be transmitted, on average, to their grandchildren. In the United States this would be (0.47)^2 or 22 percent;"*

This procedure – extrapolation by exponentiation – shapes our interpretation of the intergenerational evidence as it is common in policy reports, standard textbooks (Borjas, 2009) and specialised survey articles (Piketty, 2000).

This interpretation matters, as the persistence of economic status across generations is a central aspect in sociological, economic and political theory. Erikson and Goldthorpe (1992) note that competing political theories contain strong and opposing hypotheses about its extent in industrialised societies. Piketty (2000) observes that conflicting views feature also prominently in economic writings. Measuring multigenerational persistence may help to discriminate between competing schools of thoughts, but it also matters on a practical level. We may for example wonder if specific social policies mask inequalities between families only temporarily or if they have lasting effects on their relative fortunes.

But conflicting views about the degree of long-run mobility persist because we lack direct empirical evidence. Our knowledge about intergenerational mobility on the other hand has advanced greatly in the last two decades. The finding that income mobility is much lower than previously believed, and particularly low in countries with high levels of cross-sectional inequality in which it is more consequential (e.g. Corak, 2013), has been received with some concern. But the standard extrapolation procedure provides ammunition for a contrarian standpoint that disputes the significance of those findings, as it implies high long-run mobility even when parent-child mobility is low (see for example Mankiw, 2006).

Its prevalence may seem puzzling, as Hodge (1966) already notes that mobility may not be well described by a first-oder Markov process[1]. It can perhaps be explained by three factors. First, no comprehensive study exists on how intergenerational and multigenerational mobility *should* relate. In its absence, the iteration of intergenerational measures may be a pragmatic response. Second, this procedure appears prominently in influential studies in the literature. Section IV of Becker and Tomes (1979) draw attention to special *theoretical* cases that vindicate the iteration of intergenerational measures, quoting the old proverb *"from shirtsleeves to shirtsleeves in four generations*" to illustrate its implications. Becker and Tomes (1986) go further, applying procedure and proverb (*"... in three generations*") to estimates from the *empirical* literature. Their striking conclusion is that differences between families and the prevalence of poverty tend to disappear within few generations.[2] Finally, the idea that regression implies *iterated* regression appears quite natural in a linear regression context. Indeed, such belief turns out to be a common statistical fallacy, arising frequently also in other economic literatures and disciplines.[3]

In this note I present various simple models of intergenerational transmission to illustrate why iteration-based procedures are unlikely to approximate the true relation between intergenerational and multigenerational mobility. Starting from a baseline model I discuss the role of indirect transmission and market luck; the multiplicity of skills; the role of grandparents; and finally the causal effect of parental income. This discussion leads to a specific hypothesis: various properties of the intergenerational transmission process imply that long-run mobility will likely be *lower*, possibly much lower, than the standard extrapolation procedure implies.

I illustrate those arguments using data on three generations from Swedish registers, but my main objective is to provide a theoretical complement to the recent wave of recent empirical studies on the subject. Lindahl et al. (2014) exploit survey data on the parents, children and grandchildren of a Swedish population, and find that multigenerational persistence in income and education is severely understated by iterated parent-child estimates. Longitudinal data as used in this and other forthcoming studies (Dribe and Helgertz, 2013; Boserup et al., 2013) provide an exceptional but rare opportunity to study multigenerational persistence. Other researchers thus rely on novel methods

[1]The early empirical literature focuses on occupational mobility and is mostly concerned with the (related but distinct) question if grandparents have a direct causal influence on their grandchildren; exemplary is Warren and Hauser (1997), who also summarises earlier studies.

[2]This conclusion seems conflicting with the observation that some groups, such as black Americans, experience persistent economic disadvantage. Becker and Tomes (1986) interpret this observation not as a evidence against the extrapolation procedure from intergenerational equations as such, but instead argue that those equations could differ for Blacks (e.g. p. 28).

[3]For example, Bernard and Durlauf (1996) examine tests for the *convergence hypothesis* of the neoclassical growth model, showing that a negative slope coefficient in a cross-country regression of growth rates on initial levels of output does not suggest that poorer economies tend to fully catch up to richer ones.

to exploit repeated cross-sections instead: Long and Ferrie (2013) link individuals in British and U.S. censuses; Collado et al. (2013) exploit socioeconomic bias in the distribution of surnames in two Spanish regions; Clark (2013) relies on the informative content in rare surnames; and Olivetti et al. (2014) on information in first names.

In contrast, little theoretical work exists on the topic. Zylberberg (2013) studies the inheritance of *careers*, and shows that income persistence will decay less than geometrically if mobility is high within but low between distinct blocks of careers. Solon (2013) extends the classic Becker-Tomes framework to study multigenerational persistence, which leads to a variant of the simultaneous equation system that I am considering here. He pays particular attention on grandparent coefficients in multigenerational regressions, showing that its signs are ambiguous if grandparents have independent causal effects on their grandchildren. I consider various other elements of the transmission process, which together lead to a specific hypothesis on the relationship between inter- and multigenerational persistence.

The Iterated Regression Fallacy

I consider the iteration procedure in some detail, as it is common not only in the intergenerational literature. The degree to which differences in socio-economic outcomes between *parents* remain among their children is often measured by the slope coefficient in a linear regression of outcome y in offspring generation t of family i on parental outcome in generation $t-1$,

$$y_{it} = \alpha + \beta_{-1} y_{it-1} + \varepsilon_{it}. \tag{1}$$

If y measures log lifetime income then β_{-1} captures the *intergenerational income elasticity*, which measures the percentage differential in expected offspring income with respect to a percentage differential in parental income; a high elasticity represents low mobility. For simplicity assume stationarity, such that β_{-1} is constant over t. The arguments apply likewise in a non-stationary environment.

How does this parent-child coefficient compare with the coefficient across three or more generations, e.g. between *grandparents* and their grandchildren? The idea that the latter equals the square of the former, so that persistence declines geometrically, may appear as a natural consequence of regression: if β_{-1} captures to what degree deviations from the mean tend to be passed from parents to children then surely $(\beta_{-1})^2$ measures their expected extent after being passed twice from parents to children? Formally, one may believe that equation (1) can be used to rewrite the grandparent-grandchild elasticity β_{-2} as

$$\beta_{-2} \equiv \frac{Cov(y_{it}, y_{it-2})}{Var(y_{it-2})} = \frac{Cov(\beta_{-1} y_{it-1} + \varepsilon_{it}, y_{it-2})}{Var(y_{it-2})} = (\beta_{-1})^2. \quad (2)$$

The error lies in the last step. While ε_{it} is by construction uncorrelated to y_{it-1}, it is not necessarily uncorrelated with grandparental outcome y_{it-2}. The interpretation of equation (1) itself may be the source of confusion; it has no structural interpretation, nor does it represent an AR(1) or Markov process. Instead it captures a simple statistical relationship: $\beta_{-1} y_{it-1}$ is the best linear approximation (in a MMSE sense) to the conditional expectation $E[y_{it}|y_{it-1}]$.

The belief that regression toward the mean between two observations implies iterated regression between multiple observations appears to be a classic fallacy. Francis Galton himself fell fault of it (Bulmer, 2003). As noted in the introduction, iteration-based extrapolations are common in the intergenerational and other economic literatures, as well as other disciplines: Nesselroade et al. (1980) discuss their prevalence in developmental psychology under the caption *"expectation fallacy"*.[4] I use the term *"iterated regression fallacy"* here to relate to the consecutive nature of intergenerational transmission and to other classic regression fallacies.[5]

But is the extrapolation error from an iteration of regression coefficients quantitatively important? I provide a brief empirical application using Swedish population and education registers, covering a 35 percent random sample of the Swedish cohorts born between 1932 and 1967 and their biological parents and children. Educational attainment (converted into years of schooling) for offspring, their parents and their grandparents can be observed in about 150,000 cases.[6]

The first row of Table 1 reports coefficient estimates from child-father ($\hat{\beta}_{-1} = 0.238$) and father-grandfather ($\hat{\beta}_{-1} = 0.406$) regressions. The predicted child-grandfather coefficient based on the iteration of intergenerational measures is $0.238 \times 0.406 = 0.096$ (second row), but the estimate from an actual child-grandfather regression ($\hat{\beta}_{-2} = 0.137$, third row) is more than 40 percent higher. The extrapolation error is even larger in child-mother-grandmother coefficients (second panel). Educational differences are not very persistent in Sweden, but the iteration of intergenerational coefficients sub-

[4] The fallacy can be viewed as an incorrect application of the law of iterated expectations: extrapolation by exponentiation would be reasonable if $E[y_{it}|y_{it-2}] = E[E[y_{it}|y_{it-1}]|y_{it-2}]$, which however only holds if y_{it} follows a Markov process.

[5] Such as the belief that regression towards the mean implies convergence to the mean, see Friedman (1992); or the failure to account for regression to the mean in comparisons over time (Jerrim and Vignoles, 2012, discuss a recent example).

[6] Schooling is not observed for cohorts before 1911 and becomes increasingly right-censored in cohorts after 1972. I thus restrict my sample to individuals born 1966-1972 and their parents and grandparents.

stantially understates long-run persistence.[7]

Models of Inter- and Multigenerational Transmission

The iterated regression fallacy will become more evident in the light of a theoretical model. Consider a simplified one-parent one-offspring family structure for which income and intergenerational transmission are governed by

$$y_{it} = \rho e_{it} + u_{it} \tag{3}$$

$$e_{it} = \lambda e_{it-1} + v_{it}, \tag{4}$$

such that income y_{it} depends on human capital e_{it} (according to *returns* ρ), which is partially inherited within families (according to *heritability* λ). I use the term heritability in a wide sense, representing not only genetic but also other causal pathways of transmission from parents to children (e.g. parental upbringing). The noise terms u_{it} and v_{it} represent market and endowment luck, and are assumed to be uncorrelated with each other and past values. To simplify the presentation assume throughout that variables are measured as trendless indices with mean zero and variance one, such that slope parameters can be interpreted as correlations; and further that those indices measure favourable traits that are not negatively correlated within families, such that all parameters are non-negative. The parameter ρ then measures the fraction of income that is explained by inheritable own characteristics, as opposed to factors or events outside of individual control; for example, $\rho = 1$ implies that income differences are fully explained by own characteristics. The i subscript is dropped in the subsequent analysis.

Given equations (3) and (4), and the assumption that all variances are unity, the intergenerational elasticity equals

$$\begin{aligned} \beta_{-1} &= Cov(y_t, y_{t-1}) \\ &= \rho^2 \lambda, \end{aligned} \tag{5}$$

and the elasticity across three generations instead equals

$$\begin{aligned} \beta_{-2} &= Cov(y_t, y_{t-2}) \\ &= \rho^2 \lambda^2. \end{aligned} \tag{6}$$

[7] These findings are consistent with Lindahl et al. (2014), who provide more comprehensive evidence.

The extrapolation error from exponentiation of the parent-child elasticity equals

$$\begin{aligned}\Delta &= (\beta_{-1})^2 - \beta_{-2} \\ &= (\rho^2 - 1)\rho^2\lambda^2,\end{aligned} \tag{7}$$

which is negative if $0 < \rho < 1$ and $0 < \lambda < 1$, that is as long as income is not perfectly determined by human capital, and human capital is not perfectly inherited within families. The extrapolation error Δ will be large when ρ is small relative to the degree of heritability captured by λ.

Indirect Transmission

This simple model illustrates that multigenerational coefficients cannot be recovered by the iteration of parent-child coefficients. Why do such extrapolations *overstate* mobility here? The result stems from the interplay of causal mechanisms between and within generations. Both the imperfect inheritability of traits between ($\lambda < 1$) and the imperfect determination of incomes by those traits within generations ($\rho < 1$) decrease the intergenerational persistence of income. But regression beyond two generations depends only on the heritability parameter: persistence equals $\beta_{-2} = \beta_{-1}\lambda$ across three generations, $\beta_{-3} = \beta_{-1}\lambda^2$ across four generations, and so on. The intuition is simple: traits are inherited multiple times, but they are only once transformed into income for each generation.

The underlying assumptions should be uncontroversial. We know that at least part of the intergenerational transmission of income occurs via indirect mechanisms, for example through genetic inheritance or parental upbringing. And we do not expect individuals with equivalent levels of human capital to have exactly equal incomes, perhaps because workers trade income for non-pecuniary aspects, or because factors outside of individual control drive a wedge between skill and income.[8] The extrapolation error can be substantial even if the role of such external factors or "market luck" is modest. For example, exponentiation of an intergenerational elasticity of 0.5 implies $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.25, 0.125\}$ and thus rapid regression to the mean. But if $\rho = 0.8$ (market luck explains about a third of the cross-sectional variance in income) then $\lambda \approx 0.78$ and $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.39, 0.31\}$, implying substantial long-run persistence of economic status within families.

The gap between extrapolated and actual long-run mobility will tend to rise if the

[8]This wedge may be sizeable – earnings regressions only explain a fraction of the variation in the dependent variable, even when the list of regressors is large (e.g. Zax and Rees, 2002); monozygotic twins have substantially different earnings even while their genetic and early family background are similar; and various economic literatures show that events outside of individual control (such as occupation-, region-, or firm-specific demand shocks) affect incomes.

transmission of income is less direct. Assume that human capital is not directly transmitted within families either, but that parents bequeath certain traits a_t (according to heritability π), which in turn affect human capital e_t (according to transferability μ),[9] and thus

$$y_t = \rho e_t + u_t \tag{8}$$
$$e_t = \mu a_t + v_t \tag{9}$$
$$a_t = \pi a_{t-1} + w_t. \tag{10}$$

The parent-child and grandparent-child elasticities are then equal to $\beta_{-1} = \rho^2\mu^2\pi$ and $\beta_{-2} = \rho^2\mu^2\pi^2$. Consider parameterizations that yield the same parent-child elasticity as the two-layers model, which requires $\lambda = \mu^2\pi$. The extrapolation error then equals $\Delta = (\rho^2 - \frac{1}{\mu^2})\rho^2\lambda^2$. Comparison to equation (7) reveals that this error will be larger in the three- than in the two-layers model iff $\mu < 1$.

Various implications follow. First, long-run income mobility will be smaller the more intergenerational mobility is attributable to market luck instead of low heritability of traits. Second, policies and institutions may mask inequality only temporarily. For example, a track-school system that separates children by ability may increase the degree to which differences in child ability lead to differences in human capital (an increase of μ in the three-layers model) and thus decrease intergenerational mobility. But long-run mobility may be less affected if the heritability of those abilities remains unchanged. Third, the degree to which cross-country differences in parent-child mobility extend to the long run depends on if those differences are due to variation in the heritability or the the transferability of endowments.[10] Clark (2013) pushes this idea further, arguing that long-run mobility is instead closer to a universal constant across countries and time.

Identification from Multigenerational Data

I focus here on the implications of different causal processes for the extent of multigenerational mobility. But that relationship is interesting also the opposite way, as it may help to identify features of the underlying causal process that are otherwise difficult to capture. For example, the heritability of ability might not be directly estimable, as at best we can hope to observe a noisy proxy for certain types of traits. Multigenerational

[9] As my main intention is to capture the idea that income transmission may occur rather indirectly I abstain from specific interpretations for each layer (e.g. the lowest layer may be thought to represent genetic transmission, as in Conlisk, 1974).

[10] For example, Nordic countries will be characterised by exceptional long-run mobility if their high levels of intergenerational mobility are caused by policies and institutions that decrease the heritability of traits, less so if they are due to policies that interfere with the formation of market prices for those traits.

data offers potentially an indirect route to identification. In the three-layers model, the slope coefficient from a regression of offspring on parent human capital equals $\beta_{-1} = \mu^2\pi$ while the slope coefficient from a child-grandparent regression equals $\beta_{-2} = \mu^2\pi^2$. The ratio β_{-2}/β_{-1} identifies thus π, while $({\beta_{-1}}^2/\beta_{-2})^{1/2}$ identifies μ, from data on educational attainment alone.

Empirical implementation of distributional models is not straightforward if the environment cannot be assumed to be in steady state (see Atkinson and Jenkins, 1984). Indeed, the slope coefficient in a parent-child regression of years of schooling changes considerably over time in Sweden (see Table 1), predominantly due to variation in the cross-sectional variance. To abstract at least from this variation consider instead the correlation coefficient r, which is considerably more stable. Dividing the child-grandfather correlation ($\hat{r}_{-2} = 0.156$) by the average of the child-father ($\hat{r}_{-1} = 0.323$) and father-grandfather ($\hat{r}_{-1} = 0.340$) correlations yields $\hat{\pi} = 0.471$, and thus $\hat{\mu} = 0.839$.

These estimates can in turn be used to extrapolate beyond three generations. Simple iteration of the intergenerational correlation yields

$$\{r_{-1}, r_{-2}, r_{-3}, r_{-4}\} = \{0.332, 0.110, 0.037, 0.012\},$$

but the model-based procedure implies

$$\{r_{-1}, r_{-2}, r_{-3}, r_{-4}\} = \{0.332, 0.156, 0.074, 0.035\}.$$

This prediction is still flawed if the true causal process is not be well captured by eqs. (8) to (10). But in contrast to the iteration of bivariate coefficients it has a conceptual justification and is consistent with data over three instead of two generations. Its validity can be tested if more than three generations are observed.

An Additional Factor

A second fundamental reason why multigenerational persistence of economic status may decay less rapidly in the long than in the short run relates to the multiplicity of the transmission process. Introduce a second factor into our starting model,

$$y_t = \rho_1 e_{1t} + \rho_2 e_{2t} + u_t \tag{11}$$

$$e_{1t} = \lambda_1 e_{1,t-1} + v_{1t} \tag{12}$$

$$e_{2t} = \lambda_2 e_{2,t-1} + v_{2t}, \tag{13}$$

assuming that two traits are inherited from parents according to heritability parameters λ_1 and λ_2. For simplicity also assume that the endowment luck terms v_{1t} and v_{2t} are uncorrelated, such that $Cov(e_{1t}, e_{2t}) = 0 \; \forall \; t$. Assume further that both traits affect incomes, such that $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$. The parent-child elasticity then equals

$$\beta_{-1} = \rho_1^2 \lambda_1 + \rho_2^2 \lambda_2, \tag{14}$$

and the grandparent-grandchild elasticity equals

$$\beta_{-2} = \rho_1^2 \lambda_1^2 + \rho_2^2 \lambda_2^2. \tag{15}$$

The extrapolation error equals

$$\Delta = (\rho_1^2 - 1)\rho_1^2 \lambda_1^2 + (\rho_2^2 - 1)\rho_2^2 \lambda_2^2 + 2\rho_1^2 \rho_2^2 \lambda_1 \lambda_2. \tag{16}$$

Assume for a moment that incomes are indeed *perfectly* determined by individual traits, such that $\rho_1^2 + \rho_2^2 = 1$ and $Var(u_t) = 0$. Equation (16) can then be written as

$$\Delta = \rho_1^2 (\rho_1^2 - 1)(\lambda_1 - \lambda_2)^2. \tag{17}$$

This expression is negative for $\lambda_1 \neq \lambda_2$. In contrast to the previous models, exponentiated parent-child elasticities understate multigenerational persistence even when human capital determines incomes perfectly, as long as those traits that constitute human capital are not all equally strong inherited within families.[11] This result can be understood as the application of Jensen's inequality: the square of the average heritability across traits is smaller than the average of the square of those heritabilities. Inequality between families declines therefore more slowly if intergenerational income persistence stems from multiple causal pathways.

Highly inheritable traits explain an increasing share of the long-run persistence in income. In particular, multigenerational elasticities will never converge to zero if any characteristic is perfectly transmitted. For example, physical traits such as skin colour may be highly persistent in multi-ethnic societies if interracial marriage is rare, and may lead to persistent disadvantage of families if groups are discriminated on the labour market. The observation of high intergenerational mobility can therefore be consistent with substantial long-run persistence of economic status, provided that the multiplicity of traits is taken into account.

For the analysis of long-run mobility it is thus essential to look beyond scalar models, even if those models have proved to be useful for other questions in the literature. This applies in particular to the framework presented in Becker and Tomes (1979), which underlies much of the theoretical work in the literature. It contains only a scalar

[11] For example, Anger (2011) and Grönqvist et al. (2010) study if inheritance is stronger in cognitive than in non-cognitive abilites .

measure of human capital and does not capture implications from the existence of multiple transmission mechanisms. Moreover, human capital is often assumed to determine incomes perfectly in extended versions of this model, such as Solon (2013). I find that both the imperfect relation between skills and income and the multiplicity of those skills have important implications for long-run persistence.[12]

An Additional Generation

A question that has received much attention in multigenerational studies is if grandparents have a direct causal influence on their grandchildren (e.g. Warren and Hauser, 1997; Mare, 2011; Long and Ferrie, 2013). Such higher-order effects are often presented as a potential explanation why multigenerational persistence might decline more slowly than at a geometric rate. But the previous sections illustrated that other properties of the transmission process lead to the same implication. From the observation that $(\beta_{-1})^2 < \beta_{-2}$ we can therefore not conclude that intergenerational transmission has a memory of more than one generation.

The intuition that such higher-order effects raise long-run persistence is of course correct. To see this assume that offspring human capital depends on both parents and grandparents, such that equation (4) becomes

$$e_t = \lambda_{-1} e_{t-1} + \lambda_{-2} e_{t-2} + v_t, \tag{18}$$

with $\lambda_{-2} > 0$. Assuming stationarity the parent-child elasticity equals

$$\beta_{-1} = \rho^2 \left(\frac{\lambda_{-1}}{1 - \lambda_{-2}} \right), \tag{19}$$

Consider parameterizations that yield the same intergenerational elasticity as the previous model, such that $\lambda = \lambda_{-1}/(1 - \lambda_{-2})$. The grandparent-grandchild elasticity,

$$\beta_{-2} = \rho^2 \lambda^2 + \rho^2 \lambda_{-2}(1 - \lambda^2), \tag{20}$$

is then greater than the respective elasticity in the baseline model (assuming $\rho > 0$ and $\lambda < 1$). This simple example does not illustrate the various ways how grandparents may influence their grandchildren, but it illustrates that such influence strengthens

[12] These findings support arguments made by Goldberger (1989), who notes that an explicit consideration of utility maximisation behaviour of parents (as in Becker and Tomes, 1979) to motivate "mechanical" transmission equations may provide little additional implications but distract from the assumed properties of those equations. The Becker and Tomes model leads to transmission equations that are simplified compared to earlier models in the literature, which did contain noise terms to capture market luck and multiple inheritance mechanisms (e.g. Conlisk, 1969).

multigenerational relative to intergenerational persistence.[13]

How can we distinguish such higher-order from other causal mechanisms? One potential strategy is to find quasi-exogenous variation in the exposure to certain family members in a careful research design (Adermon, 2013). But simpler methods may be sufficient to bound their magnitude. Note that given eq. (18), the slope coefficient in linear regression of child on grandparent human capital equals

$$\beta_{-2,e} = Cov(e_t, e_{t-2}) = \lambda_{-2} + \lambda_{-1} Cov(e_{t-1}, e_{t-2}). \tag{21}$$

The coefficient in a regression of child on grandparent characteristics may have a positive coefficient either because grandparents have a direct effect on grandchildren ($\lambda_{-2} \neq 0$) – or because grandparent and parent characteristics are correlated, and children are affected by the latter. These two channels are in my simple model identified by the coefficients in a regression of child on parent *and* grandparent human capital. But in practice we cannot be sure that all relevant parent characteristics are included; we do not know if a positive coefficient on grandparents reflects direct causal effects or an omitted variable bias.

However, we can illustrate the potential magnitude of this bias by conditioning on exceedingly many characteristics of the parent generation. Column (1) of Table 2 reports estimates from a regression of offspring on fathers' and (paternal) grandfathers' years of schooling. The estimated coefficient on grandfathers' schooling is sizeable and statistically significant. However, already the inclusion of mothers' years of schooling reduces this estimate by nearly one half (colum 2); a large fraction of the grandparent coefficient reflects that fathers with highly-educated grandfathers tend to have highly-educated partners, and the latter have a more direct relation with child outcomes. Controlling for parental income (column 3), allowing for a more flexible functional form by including schooling levels as indicator variables (column 4), or including schooling for all grandparents (column 5) reduces the grandfather coefficient to a precisely estimated zero.

For a subset of the fathers we observe detailed information on cognitive and non-cognitive ability from military enlistment tests. This subsample is small and quite peculiar, as test scores are observed only for the youngest parents in my sample. It is nevertheless interesting that their inclusion pushes the grandfather coefficient below zero (column 6).

These results do not suggest that grandparents have no direct effects on their grandchildren.[14] They however suggest that a consideration of parent-child transmission pro-

[13] Mare (2011) and Solon (2013) suggest various channels via which grandparents may affect their grandchildren.

[14] First, we are considering only distributional, not mean effects. Second, one might expect a *negative*

cesses may often be sufficient even if our objective is to understand multigenerational persistence.[15] The observed deceleration of regression to the mean beyond two generations is more plausibly explained by the multiplicity and indirectness of parent-child transmission processes.

Parental Investment

All previous results point to "excess persistence", to the conclusion that extrapolated intergenerational elasticities understate long-run persistence. But we can certainly think of circumstances in which the opposite holds, for which I will give one example.

Assume that parental income or economic status have a causal effect on offspring; for example indirectly through parental investments in offspring human capital, or more directly through reputation or networking effects on the labour market. Consider the first case, such that equations (3) and (4) change into

$$y_t = \rho e_t + u_t \tag{22}$$

$$e_t = \theta y_{t-1} + \eta e_{t-1} + v_t. \tag{23}$$

The parent-child and grandparent-grandchild elasticities then equal

$$\beta_{-1} = \rho\theta + \rho^2\eta$$
$$\beta_{-2} = (\rho\eta + \rho^2\theta)(\rho\eta + \theta).$$

Consider again parameterizations that yield the same level of β_{-1}, which requires $\eta < \lambda$ (assuming $\rho > 0$ and $\theta > 0$). The extrapolation error,

$$\Delta = (\rho^2 - 1)\eta\beta_{-1}, \tag{24}$$

is smaller than the error in our first model (which equals $(\rho^2 - 1)\lambda\beta_{-1}$), but it will still be negative. Our previous findings still hold when parental income affects offspring human capital.

Now instead assume that parental income has a direct effect on offspring income that is independent of offspring characteristics, such that equations (3) and (4) change

grandparent coefficient in the absence of direct grandparental effects, as explained in Solon (2013). Finally, grandparents may play a more important role in other populations or in other outcomes, such as wealth.

[15] Warren and Hauser (1997), based on a smaller sample but more comprehensive analysis, come to a similar conclusion.

into

$$y_t = \phi y_{t-1} + \tau e_t + u_t \tag{25}$$
$$e_t = \lambda e_{t-1} + v_t. \tag{26}$$

The parent-child and grandparent-grandchild elasticities then equal

$$\beta_{-1} = \phi + \frac{\tau^2 \lambda}{1 - \phi\lambda}$$
$$\beta_{-2} = \phi^2 + \frac{\tau^2 \lambda}{1 - \phi\lambda}(\phi + \lambda),$$

The extrapolation error equals

$$\Delta = \left(\frac{\tau^2 \lambda}{1 - \phi\lambda}\right)^2 + (\phi - \lambda)\frac{\tau^2 \lambda}{1 - \phi\lambda}. \tag{27}$$

which in contrast to the previous examples may be positive, in particular if $\phi > \lambda$. In this model, income is affected by both parental income and ability, but offspring ability is affected exclusively by parental ability. While parent-child persistence may be strongly affected by the direct effect of income ϕ, long-run persistence will be dominated by the heritability of ability λ. If the former is larger than the latter we have a system in which multigenerational persistence is weaker than exponentiation of β_{-1} implies.

We may expect that the causal effect of parental income is small (Björklund and Jäntti, 2009), and that at least part of it is indirect, for example through parental investments in child human capital. The case $\phi > \lambda$ seems thus of less practical relevance. Still, the model has interesting implications for cross-country differences in short- and long-run mobility.[16] The direct effect of parental income captured by ϕ will tend to be larger if credit constraints are more important. Eq. (27) then implies that extrapolation from parent-child correlations understates long-run persistence more in those countries in which credit constraints play less of a role.

These models illustrate that different beliefs about causal pathways of transmission are consistent with different expectations about long-run mobility. The belief that children from affluent families tend to fare better mainly because inherited traits and parental investment raise their productive abilities is consistent with the expectation that long-run mobility is lower than the iteration of intergenerational coefficients implies. But some authors emphasise the importance of genetic inheritance and parental investment while challenging the significance of low intergenerational mobility estimates on

[16] I thank an anonymous referee for the following observations.

the grounds that they nevertheless imply high long-run mobility (Mankiw, 2006). The opposite argument applies if one believes that income persistence stems mainly from mechanisms that are unrelated to individual productivity. If income persistence is *only* due to the direct influence of parental income then its decline over generations is indeed geometrically, and even low levels of intergenerational mobility would imply rapid multigenerational regression to the mean.

Conclusions

For lack of direct evidence, predictions on the degree of long-run mobility across multiple generations are routinely derived by extrapolation from intergenerational evidence. But an iteration of parent-child correlations implies high long-run mobility even when intergenerational mobility is low – suggesting that the extensive literature measuring such correlations is of little consequences for the distribution of socio-economic characteristics beyond two generations.

In this note I studied various elements of the intergenerational transmission process that lead to a different conclusion: the persistence of socio-economic differences in status is likely to be higher, perhaps much higher than the iteration of parent-child correlations implies.

The idea to iterate slope coefficients appears quite natural in a linear regression context, and this *iterated regression fallacy* is widespread not only in the intergenerational but also in other economic literatures and disciplines. I first illustrated why iteration-based extrapolation procedures are unlikely to provide a good approximation of multigenerational persistence. I provided empirical support for this argument using data on educational attainment across three generations from Swedish registries. Regression to the mean is comparatively strong in Sweden, but its rate slows indeed substantially after two generations.

I considered then how various elements of the transmission process affect the relation between inter- and multigenerational mobility. I discussed the role of direct and and indirect pathways of transmission; of the multiplicity of skills; of higher-order causal effects; and the role of parental income. Regression to the mean slows over generations if factors that are orthogonal to individual characteristics explain some fraction of the variation in socio-economic outcomes. The multiplicity of skills also matters, as highly inheritable traits explain an increasing share of socio-economic persistence across generations. Moreover, multiplicity provides a simple explanation why groups can suffer from persistent economic disadvantage even when parent-child mobility is high.

The recent literature has been particularly concerned with the question if grandparents have a direct causal influence on their grandchildren. I argued that other properties of the transmission process are more important for understanding long-run persistence in socio-economic outcomes. I used the Swedish data for a simple illustration: while all coefficients in a regression of child on parent and grandparent schooling tend to be positive, the coefficient on grandparent schooling vanishes quickly and may even turn negative when we include a wider set of parental controls.

Questions on mobility across multiple generations are closely related to questions on the causal pathways of transmission. This note focused on the implications that different causal channels have for the extent of multigenerational mobility, but the relation is interesting both ways. I illustrated how a comparison of intergenerational and three-generation coefficients may help to identify features of the underlying causal process and lead to better extrapolations from the available intergenerational evidence.

References

ADERMON, A. (2013): "Educational Mobility Across Three Generations," Ph.D. thesis, Uppsala University.

ANGER, S. (2011): "The Intergenerational Transmission of Cognitive and Non-Cognitive Skills During Adolescence and Young Adulthood," IZA Discussion Papers 5749, Institute for the Study of Labor (IZA).

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

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

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

BERNARD, A. B., AND S. N. DURLAUF (1996): "Interpreting Tests of the Convergence Hypothesis," *Journal of Econometrics*, 71(1-2), 161–173.

BJÖRKLUND, A., AND M. JÄNTTI (2009): "Intergenerational Income Mobility and the Role of Family Background," in *Oxford Handbook of Economic Inequality*, ed. by B. N. Salverda Wiemer, and T. M. Smeeding. Oxford University Press.

BORJAS, G. J. (2009): *Labor Economics*. New York: Irwin/McGraw-Hill, 5th edn.

BOSERUP, S. H., W. KOPCZUK, AND C. T. KREINER (2013): "Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations," Discussion paper, University of Copenhagen.

BULMER, M. (2003): *Francis Galton: Pioneer of Heredity and Biometry*. Johns Hopkins University Press.

CLARK, G. (2013): "What is the True Rate of Social Mobility? Evidence What is the True Rate of Social Mobility? Evidence from the Information Content of Surnames," Unpublished working paper, presented at the AEA Annual Meeting, San Diego.

COLLADO, M. D., I. ORTUÑO-ORTIN, AND A. ROMEU (2013): "Long-run Intergenerational Social Mobility and the Distribution of Surnames," Discussion paper, Unpublished working paper.

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

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

CORAK, M. (2013): "Inequality from Generation to Generation: The United States in Comparison," in *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ed. by R. Rycroft. ABC-CLIO.

DRIBE, M., AND J. HELGERTZ (2013): "Long-term social mobility in Sweden: A three-generation approach," Discussion paper, mimeo.

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

FRIEDMAN, M. (1992): "Do Old Fallacies Ever Die?," *Journal of Economic Literature*, 30(4), 2129–2132.

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

GRÖNQVIST, E., B. ÖCKERT, AND J. VLACHOS (2010): "The Intergenerational Transmission of Cognitive and Non-cognitive Abilities," IFAU Working Paper 2010:12, Institute for Evaluation of Labour Market and Education Policy.

HERTZ, T. (2006): "Understanding Mobility in America," in *Report For the Center for American Progress*.

HODGE, R. (1966): "Occupational Mobility as a Probability Process," *Demography*, 3(1), 19–34.

JERRIM, J., AND A. VIGNOLES (2012): "Social Mobility, Regression to the Mean and the Cognitive Development of High Ability Children From Disadvantaged Homes," *Journal of the Royal Statistical Society: Series A*.

LINDAHL, M., M. PALME, S. SANDGREN MASSIH, AND A. SJÖGREN (2014): "Dynastic Capital and Long-term Intergenerational Persistence of Earnings and Educational Attainment," *Journal of Human Resources*, forthcoming.

LONG, J., AND J. FERRIE (2013): "Grandfathers Matter(ed): Occupational Mobility Across Three Generations in the U.S. and Britain, 1850-1910," Unpublished working paper, presented at the AEA Annual Meeting, San Diego.

MANKIW, G. (2006): "Rich Dad, Poor Dad," in *Greg Mankiw's Blog, published on April 27, 2006*. http://gregmankiw.blogspot.co.uk/2006/04/rich-dad-poor-dad.html.

MARE, R. (2011): "A Multigenerational View of Inequality," *Demography*, 48(1), 1–23.

NESSELROADE, J. R., S. M. STIGLER, AND P. B. BALTES (1980): "Regression Toward the Mean and the Study of Change," *Psychological Bulletin*, 88(3).

OLIVETTI, C., D. PASERMAN, AND L. SALISBURY (2014): "Intergenerational Mobility Across Three Generations in the 19th Century: Evidence from the US Census," Discussion paper, Unpublished working paper.

PIKETTY, T. (2000): "Theories of Persistent Inequality and Intergenerational Mobility," in *Handbook of Income Distribution*, ed. by A. B. Atkinson, and F. Bourguignon, vol. 1, pp. 429 – 476. Elsevier.

SOLON, G. (2013): "Theoretical Models of Inequality Transmission across Multiple Generations," NBER Working Papers 18790, National Bureau of Economic Research, Inc.

WARREN, J. R., AND R. M. HAUSER (1997): "Social Stratification across Three Generations: New Evidence from the Wisconsin Longitudinal Study," *American Sociological Review*, 62(4), pp. 561–572.

ZAX, J., AND D. REES (2002): "IQ, Academic Performance, Environment, and Earnings," *Review of Economics and Statistics*, 84(4), 600–616.

ZYLBERBERG, Y. (2013): "Dynastic Income Mobility," Discussion paper, CREi.

References

ADERMON, A. (2013): "Educational Mobility Across Three Generations," Ph.D. thesis, Uppsala University.

ANGER, S. (2011): "The Intergenerational Transmission of Cognitive and Non-Cognitive Skills During Adolescence and Young Adulthood," IZA Discussion Papers 5749, Institute for the Study of Labor (IZA).

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

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

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

BERNARD, A. B., AND S. N. DURLAUF (1996): "Interpreting Tests of the Convergence Hypothesis," *Journal of Econometrics*, 71(1-2), 161–173.

BJÖRKLUND, A., AND M. JÄNTTI (2009): "Intergenerational Income Mobility and the Role of Family Background," in *Oxford Handbook of Economic Inequality*, ed. by B. N. Salverda Wiemer, and T. M. Smeeding. Oxford University Press.

BORJAS, G. J. (2009): *Labor Economics*. New York: Irwin/McGraw-Hill, 5th edn.

BOSERUP, S. H., W. KOPCZUK, AND C. T. KREINER (2013): "Intergenerational Wealth Mobility: Evidence from Danish Wealth Records of Three Generations," Discussion paper, University of Copenhagen.

BULMER, M. (2003): *Francis Galton: Pioneer of Heredity and Biometry*. Johns Hopkins University Press.

CLARK, G. (2013): "What is the True Rate of Social Mobility? Evidence What is the True Rate of Social Mobility? Evidence from the Information Content of Surnames," Unpublished working paper, presented at the AEA Annual Meeting, San Diego.

COLLADO, M. D., I. ORTUÑO-ORTIN, AND A. ROMEU (2013): "Long-run Intergenerational Social Mobility and the Distribution of Surnames," Discussion paper, Unpublished working paper.

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

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

CORAK, M. (2013): "Inequality from Generation to Generation: The United States in Comparison," in *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ed. by R. Rycroft. ABC-CLIO.

DRIBE, M., AND J. HELGERTZ (2013): "Long-term social mobility in Sweden: A three-generation approach," Discussion paper, mimeo.

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

FRIEDMAN, M. (1992): "Do Old Fallacies Ever Die?," *Journal of Economic Literature*, 30(4), 2129–2132.

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

GRÖNQVIST, E., B. ÖCKERT, AND J. VLACHOS (2010): "The Intergenerational Transmission of Cognitive and Non-cognitive Abilities," IFAU Working Paper 2010:12, Institute for Evaluation of Labour Market and Education Policy.

HERTZ, T. (2006): "Understanding Mobility in America," in *Report For the Center for American Progress*.

HODGE, R. (1966): "Occupational Mobility as a Probability Process," *Demography*, 3(1), 19–34.

JERRIM, J., AND A. VIGNOLES (2012): "Social Mobility, Regression to the Mean and the Cognitive Development of High Ability Children From Disadvantaged Homes," *Journal of the Royal Statistical Society: Series A*.

LINDAHL, M., M. PALME, S. SANDGREN MASSIH, AND A. SJÖGREN (2014): "Dynastic Capital and Long-term Intergenerational Persistence of Earnings and Educational Attainment," *Journal of Human Resources*, forthcoming.

LONG, J., AND J. FERRIE (2013): "Grandfathers Matter(ed): Occupational Mobility Across Three Generations in the U.S. and Britain, 1850-1910," Unpublished working paper, presented at the AEA Annual Meeting, San Diego.

MANKIW, G. (2006): "Rich Dad, Poor Dad," in *Greg Mankiw's Blog, published on April 27, 2006*. http://gregmankiw.blogspot.co.uk/2006/04/rich-dad-poor-dad.html.

MARE, R. (2011): "A Multigenerational View of Inequality," *Demography*, 48(1), 1–23.

NESSELROADE, J. R., S. M. STIGLER, AND P. B. BALTES (1980): "Regression Toward the Mean and the Study of Change," *Psychological Bulletin*, 88(3).

OLIVETTI, C., D. PASERMAN, AND L. SALISBURY (2014): "Intergenerational Mobility Across Three Generations in the 19th Century: Evidence from the US Census," Discussion paper, Unpublished working paper.

PIKETTY, T. (2000): "Theories of Persistent Inequality and Intergenerational Mobility," in *Handbook of Income Distribution*, ed. by A. B. Atkinson, and F. Bourguignon, vol. 1, pp. 429 – 476. Elsevier.

SOLON, G. (2013): "Theoretical Models of Inequality Transmission across Multiple Generations," NBER Working Papers 18790, National Bureau of Economic Research, Inc.

WARREN, J. R., AND R. M. HAUSER (1997): "Social Stratification across Three Generations: New Evidence from the Wisconsin Longitudinal Study," *American Sociological Review*, 62(4), pp. 561–572.

ZAX, J., AND D. REES (2002): "IQ, Academic Performance, Environment, and Earnings," *Review of Economics and Statistics*, 84(4), 600–616.

ZYLBERBERG, Y. (2013): "Dynastic Income Mobility," Discussion paper, CREi.

Tables and Figures

Table 1: Intergenerational and Multigenerational Persistence in Educational Attainment

	Child	Father	Grandfather
two generations	0.238***		0.406***
	(0.002)		(0.004)
three gen. (prediction)		0.096***	
		(0.001)	
three gen. (actual)		0.137***	
		(0.003)	
	Child	Mother	Grandmother
two generations	0.267***		0.301***
	(0.002)		(0.005)
three gen. (prediction)		0.080***	
		(0.002)	
three gen. (actual)		0.152***	
		(0.004)	

Notes: Slope coefficients from separate regressions of years of schooling of offspring on years of schooling of family member in older generation. N=145,590 observations for panel A (fathers/grandfathers), N=156,847 for panel B (mothers/grandmothers). Standard errors (in parentheses) are clustered by family.

Table 2: The Grandparent Coefficient in Educational Persistence

	Years of schooling - Child					
	(1)	(2)	(3)	(4)	(5)	(6)
Parents:						
schooling father	0.222***	0.159***	0.135***	saturated	saturated	saturated
	(0.0025)	(0.0026)	(0.0027)			
schooling mother		0.182***	0.169***	saturated	saturated	saturated
		(0.00289)	(0.0029)			
income father			0.546***	0.461***	0.407***	0.191***
			(0.0171)	(0.0169)	(0.0245)	(0.068)
income mother			-0.0176	-0.0021	0.0298*	0.120**
			(0.0095)	(0.0094)	(0.0152)	(0.055)
ability	-	-	-	-	-	x
Grandparents:						
schooling grandfather	**0.0456*** **	**0.0259*** **	**0.0183*** **	**0.0083** **	**0.0029**	**-0.0132**
(paternal)	(0.0031)	(0.0030)	(0.0030)	(0.0030)	(0.0047)	(0.0164)
schooling grandmother					0.0069	
(paternal)					(0.0060)	
schooling grandfather					0.0061	
(maternal)					(0.0046)	
schooling grandmother					0.0069	
(maternal)					(0.0059)	
# obs.	104,904	104,904	104,904	104,904	47,797	2,789

Notes: Slope coefficients from separate regressions of years of schooling of offspring on characteristics of parents and grandparents. Standard errors (in parentheses) are clustered by family.