Does Opportunity Skip Generations?

Reassessing Evidence from Sibling and Cousin Correlations

Ian Lundberg*

Department of Sociology and Office of Population Research

Princeton University

Last updated: 6 January 2020

Keywords: social mobility, multigenerational mobility, demography, quantitative methods, family

*Direct correspondence to Ian Lundberg, 227 Wallace Hall, Princeton, NJ 08544; ilundberg@princeton.edu. This is a preprint of an article forthcoming in *Demography*. Once available, this will be updated with a link to the final authenticated version. Replication code is available on the Harvard Dataverse: <https://doi.org/10.7910/DVN/HUZ1CD>. A previous version was presented at the 2018 Annual Meeting of the Population Association of America and at a workshop on sibling models in intergenerational transmission research held at Oxford University in January 2019. For useful comments on earlier versions, I thank Brandon Stewart, Dalton Conley, Sara McLanahan, Matthew Salganik, Michael Hout, Robert Mare, Per Engzell, and students in the Joint Degree Program in Social Policy at Princeton University. All errors are my own. Research reported in this publication was supported by The Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number P2CHD047879. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and the National Science Foundation under award number 1157698.

Title

Does Opportunity Skip Generations? Reassessing Evidence from Sibling and Cousin Correlations

Abstract

Sibling (cousin) correlations are empirically straightforward: they capture the degree to which siblings' (cousins') socioeconomic outcomes are similar. At face value, these quantities seem to summarize something about how families constrain opportunity. Their meaning, however, is complicated. One empirical set of sibling and cousin correlations can be generated from a multitude of distinct theoretical processes. I illustrate this problem in the context of multigenerational mobility: the relationship between the incomes of an ancestor and a descendant separated by several generations in a family. When cousins' outcomes are similar (an empirical fact), prior authors have favored a particular theoretical interpretation: that extended kin affect life chances through pathways not involving the parents of the focal individual. I show that this evidence is consistent with alternative theories of latent transmission (measurement error) or dynamic transmission (a parent-to-child transmission process that changes over generations). Theoretical assumptions are required to lend meaning to a point estimate. Further, I show that point estimates alone may be misleading because they can be highly uncertain. To facilitate uncertainty estimation for the key test statistic, I develop a Bayesian procedure to estimate sibling and cousin correlations. I conclude by outlining how future research might use sibling and cousin correlations as effective descriptive quantities while remaining cognizant that these quantities could arise from a variety of distinct theoretical processes.

(220 words)

1 Introduction

Sibling and cousin correlations are a standard tool in the study of stratification and mobility. Suppose we construct a scatterplot with each pair of siblings represented by a point, with one sibling's income on the x -axis and the other's income on the y -axis. The correlation between the two is often taken as an omnibus measure of the degree to which family of origin constrains life chances (Solon et al., 1991). Noting that cousins share extended kin (e.g. grandparents) but come from different nuclear families, recent research has used sibling and cousin correlations in combination to assess the degree to which life chances vary across nuclear and extended families (Jæger, 2012; Hällsten, 2014; Knigge, 2016).

The problem with sibling and cousin correlations is that they are difficult to interpret. Siblings can be similar because parents strongly influence life chances, because they are raised in the same neighborhood, because they influence each other's outcomes, or for numerous other possible reasons. Many theoretical stories are consistent with a given set of empirical estimates. The same is true of cousin correlations. While it is tempting to interpret a large cousin correlation as evidence that extended families shape life chances, cousins' outcomes would also be similar under a parent-to-child transmission process in which transmission is strong from grandparents to parents and from parents to offspring. Echoing Stuhler's (2012) argument that parent-offspring correlations can arise from a variety of causal processes, the present paper shows that sibling and cousin correlations can also arise from a variety of causal processes. Similar to the way biologists interpret similarities of traits in light of path models that summarize the generative process (Otto et al., 1995), a sibling correlation in social science also takes its meaning only under a theory about how that correlation came to be.

As a reference point, I begin with the Becker and Tomes (1979) model of income mobility. This model predicts that the cousin correlation will equal the squared sibling correlation. When empirical evidence contradicts this model, I show that the evidence is consistent with at least three distinct theoretical interpretations. Estimates would deviate from the Becker-Tomes prediction if the process is dynamic (changing across generations), latent (outcomes measured with error), or

second-order (involving a direct role for grandparents). While some prior authors have favored the third interpretation, there is no particular reason to do so.

Second, I apply a Bayesian estimation procedure to produce new estimates of sibling and cousin correlations in permanent log family income in the U.S. The advantage of this procedure is that it facilitates reporting of uncertainty about the key test statistic and implied parameters of theoretical models. When estimates seem to contradict a theoretical model, it becomes essential to assess our degree of confidence by incorporating uncertainty.

The paper concludes with a discussion of the role sibling and cousin correlations may continue to play in applied research. I highlight how these quantities remain useful in a descriptive sense and may serve as a springboard for new theories, even if they cannot provide definitive evidence for or against a particular theory under assumptions we can credibly defend.

2 Sibling and cousin correlations: A variance decomposition

Sibling correlations capture the expected degree to which an outcome tends to be similar for a pair of siblings. Although these quantities are standard in stratification and mobility (Jencks et al., 1972; Mazumder, 2008), this section briefly introduces them for the general reader. Suppose an outcome μ_i for person i is decomposed into three components.

$$\underbrace{\mu_i}_{\text{Socioeconomic outcome}} = \underbrace{\delta_i}_{\text{Component unique to this person}} + \underbrace{\eta_{\mathcal{S}(i)}}_{\text{Component shared within sibling set}} + \underbrace{\lambda_{\mathcal{C}(i)}}_{\text{Component shared within cousin set}} \quad (1)$$

In the notation above, $\mathcal{S}(i)$ indexes the sibling set to which i belongs: all siblings of i have the same value of $\mathcal{S}(i)$. Likewise, $\mathcal{C}(i)$ indexes the cousin set to which i belongs: all siblings and cousins of i have the same value of $\mathcal{C}(i)$. Denote the variance of each component by $\{\sigma_{\text{Person}}^2, \sigma_{\text{SiblingSet}}^2, \sigma_{\text{CousinSet}}^2\}$.

The relationship of Eq. 1 to sibling and cousin correlations relies critically on a set of independence assumptions that components are independent across units and across levels. To make this concrete, sibling correlations assume that the person-specific component is independent across

individuals ($\delta_i \perp\!\!\!\perp \delta_j \quad \forall \quad i, j$). This assumption would be violated if siblings differentiate from each other, thereby making δ_i and δ_j negatively correlated for a pair of siblings i and j . In the extreme, this would occur in a dystopian world in which every amount that sibling 1 sacrifices helps sibling 2 to succeed (Conley, 2008). This dystopian world may seem implausible, but in the actual world siblings do in fact differentiate from each other to some degree (Conley, 2004). The use of variance decompositions to study sibling correlations requires us to set aside this possibility. Likewise, estimators for sibling and cousin correlations typically assume independence across levels, for instance ruling out that the average attainment of one's sibling or cousin set is related to the variance of attainment for individuals within that set.

Armed with the independence assumption, the correlation between μ for a pair of siblings i and j is a function of the variance at each level (Mazumder, 2008).

$$\rho_{\text{Sibling}} = \text{Cor}(\mu_i^{\text{Person}}, \mu_j^{\text{Person}}) \quad (2)$$

$$= \frac{\text{Cov}(\mu_i^{\text{Person}}, \mu_j^{\text{Person}})}{\left(\text{V}(\mu_i^{\text{Person}}) \text{V}(\mu_j^{\text{Person}}) \right)^{\frac{1}{2}}} \quad \text{by definition of correlation} \quad (3)$$

$$= \frac{\text{Cov}(\delta_i + \eta_{\mathcal{S}(i)} + \lambda_{\mathcal{C}(i)}, \delta_j + \eta_{\mathcal{S}(j)} + \lambda_{\mathcal{C}(j)})}{\left(\text{V}(\mu_i^{\text{Person}}) \text{V}(\mu_j^{\text{Person}}) \right)^{\frac{1}{2}}} \quad \text{by Eq. 1} \quad (4)$$

Expand, dropping cross-level terms by independence:

$$= \frac{\overbrace{\text{Cov}(\delta_i, \delta_j)}^{=0 \text{ by independence}} + \overbrace{\text{Cov}(\eta_{\mathcal{S}(i)}, \eta_{\mathcal{S}(j)})}^{=\sigma_{\text{SiblingSet}}^2 \text{ since } \mathcal{S}(i)=\mathcal{S}(j) \text{ for siblings}} + \overbrace{\text{Cov}(\lambda_{\mathcal{C}(i)}, \lambda_{\mathcal{C}(j)})}^{=\sigma_{\text{CousinSet}}^2 \text{ since } \mathcal{C}(i)=\mathcal{C}(j) \text{ for siblings}}}{\left(\underbrace{\text{V}(\mu_i^{\text{Person}})}_{=\sigma_{\text{Person}}^2 + \sigma_{\text{SiblingSet}}^2 + \sigma_{\text{CousinSet}}^2} \underbrace{\text{V}(\mu_j^{\text{Person}})}_{=\sigma_{\text{Person}}^2 + \sigma_{\text{SiblingSet}}^2 + \sigma_{\text{CousinSet}}^2} \right)^{\frac{1}{2}}} \quad (5)$$

$$= \frac{\overbrace{\sigma_{\text{SiblingSet}}^2 + \sigma_{\text{CousinSet}}^2}^{\text{Variance between sibling sets}}}{\underbrace{\sigma_{\text{Person}}^2 + \sigma_{\text{SiblingSet}}^2 + \sigma_{\text{CousinSet}}^2}_{\text{Total variance}}} \quad (6)$$

Under the independence assumption, the proportion of the variance that is between sibling

sets equals the sibling correlation. Analogously, the proportion of the variance between cousin sets equals the cousin correlation. For this reason, even descriptive use of sibling and cousin correlations as measures of similarity within families relies on the independence assumption. This paper’s primary focus is the *interpretation* of sibling and cousin correlations in light of theoretical models of social mobility. The next section formalizes theoretical models of multigenerational mobility using parameters (e.g. β) linking outcomes across multiple generations. Sibling and cousin correlations provide a tool to infer these parameters given data on the outcome generation alone. As I will argue, inferring multigenerational parameters from sibling and cousin correlations requires conceptual arguments for a particular theoretical model. It is often difficult to make a compelling case for one theory over another, which makes inference about multigenerational processes from sibling and cousin correlations difficult.

3 Deviation from Becker-Tomes: Three distinct interpretations

A deep literature has built on the formal model of Becker and Tomes (1979), in which the expected income Y_g of someone in generation g given the incomes of all ancestors $\{Y_{g'}\}_{g' < g}$ is a linear function β of the parent income Y_{g-1} .

$$E(Y_g \mid \{Y_{g'}\}_{g' < g}) = \beta Y_{g-1} \quad \forall g \quad (7)$$

This model forms the basis of the famous claim that families can go “from shirtsleeves to shirtsleeves in three generations” (Becker and Tomes, 1986, p. S28), rising from poverty to luxury and back again. A unit increase in generation g income produces a β increase in the expected income of generation $g + 1$, but only a β^k increase k generations in the future. If we conceptualize each income measure Y_g as standardized to mean 0 and variance 1, the coefficient β can be interpreted as the correlation between the incomes of parents and offspring within a family. Fig 1 plots the rapid decay across generations that results for the case when $\beta = 0.5$. Despite a moderately high 0.5 correlation between the incomes of parents and offspring, a great-great-grandparent’s income is

only correlated 0.06 with offspring income. In short, families have abundant opportunity to move away from their origins over the course of a few generations.

Figure 1. Multigenerational permanent income correlations decay rapidly under the Becker-Tomes model. The attainment of a distant ancestor is only weakly associated with one’s own attainment even if β is large ($\beta = 0.5$ depicted).

The prediction of substantial mobility over multiple generations holds only insofar as the assumptions of the Becker and Tomes (1979) model are correct. Three assumptions are noteworthy (Fig 2). First, the assumed process is *observed rather than latent*: observed parent income Y_{g-1} is sufficient to block all statistical dependence between earlier ancestor incomes $\{Y_{g'}\}_{g' < g}$ and offspring income Y_g . Even if this holds for true income, income reported with error $\tilde{Y}_{g-1}$ would not be sufficient to block this dependence (Ferrie et al., 2016). This is the sense in which the process is observed. Second, the assumed process is *static rather than dynamic*: the process that generates incomes in generation g is the same as the process that generates incomes in generation $g - 1$. This would be violated if the process of mobility changes over time. If the ability to buy advantages for one’s offspring differs across generations, a single β cannot represent this relationship in all generations. Third, the assumed process is *first-order rather than second-order*: once the incomes of parents are known, the incomes of grandparents provide no additional information about the life chances of offspring. This would not be true, for instance, if grandparents directly affect life chances net of parents (Mare, 2011; Pfeffer, 2014).

Statistical evidence that contradicts the Becker-Tomes model may mean that any of these

A) Becker-Tomes

B) Dynamic

C) Latent

D) Second-order

Figure 2. Theoretical models of multigenerational mobility. Each $\bullet$ represents a variable in the transmission process. Figure depicts the outcomes of a pair of siblings and one of their cousins. Panel A is the Becker-Tomes model. Panel B allows the parent-offspring transmission coefficient to be dynamic across generations. Panel C maintains a static transmission parameter but adds a measurement step with reliability η . Panel D assumes a static process with perfect measurement but allows a direct grandparent-offspring link. Evidence against the Becker-Tomes model (A) can be interpreted as evidence for any of the three alternatives (B–D), or others not depicted.

assumptions does not hold. Sibling and cousin correlations offer one source of evidence against the model. In particular, the model predicts a sibling correlation of β^2 and a cousin correlation of β^4 (online appendix, part Part 1). Without directly estimating β , one can assess evidence against this model by a test statistic τ : the difference between the cousin correlation and the squared sibling correlation,

$$\tau = \rho_{\text{Cousin}} - \rho_{\text{Sibling}}^2 \quad (8)$$

where $\tau = 0$ under the model of Becker and Tomes (1979). The following subsections show how this prediction does not hold under violations of each of the assumptions that the process is static, observed, and first-order.

3.1 Interpretation 1: From a static to a dynamic process

Suppose the process is dynamic, so that the intergenerational income elasticity β_g is different in each generation g .

$$\mathbb{E}(Y_g \mid \{Y_{g'}\}_{g' < g}) = \underbrace{\beta_g}_{\text{Dynamic across generations } g} Y_{g-1} \quad \forall g \quad (9)$$

In this case, a unit increase in parent income increases offspring income in generation g by β_g . A unit increase in grandparent income increases parent income in generation $g - 1$ by a different value β_{g-1} . Combining these, a unit increase in grandparent income increases offspring income by $\beta_{g-1}\beta_g$. This could be greater or less than β_g^2 (the prediction of the static model) depending on whether the grandparent-parent transmission β_{g-1} is greater or less than the parent-offspring transmission β_g .

In the analogous intragenerational case, each sibling's income is correlated β_g with parent income, producing a correlation of β_g^2 between siblings (online appendix, part Part 1). Cousin's incomes are correlated $\beta_{g-1}\beta_g$ with the income of a shared grandparent, producing a cousin correlation of $\beta_{g-1}^2\beta_g^2$. Notably, this cousin correlation can be larger or smaller than the squared sibling

correlation. Given an estimated sibling correlation, the Becker-Tomes model is consistent with only one cousin correlation whereas a dynamic model is consistent with any cousin correlation.

$$\rho_{\text{Sibling}} = \beta_g^2 \quad (10)$$

$$\rho_{\text{Cousin}} = \beta_g^2 \beta_{g-1}^2 \quad (11)$$

$$\rho_{\text{Sibling}}^2 < \rho_{\text{Cousin}} \text{ if } \beta_g < \beta_{g-1} \quad (12)$$

$$\rho_{\text{Sibling}}^2 = \rho_{\text{Cousin}} \text{ if } \beta_g = \beta_{g-1} \quad (13)$$

$$\rho_{\text{Sibling}}^2 > \rho_{\text{Cousin}} \text{ if } \beta_g > \beta_{g-1} \quad (14)$$

3.2 Interpretation 2: From an observed to a latent process

Measurement error has formed the core of numerous debates in stratification research. As early as the 1960s, it was known that interpretation of correlations in socioeconomic status depends on the “assumptions one is willing to defend” about the relationship between measured variables and true values (Siegel and Hodge, 1968). For instance, while Blau and Duncan (1967) found substantial returns to education net of family background variables, Bowles (1972, p. S222) called this conclusion “seriously misleading” due to reporting error in family background variables that downwardly bias their association with the outcome and produce an overstatement of the returns to education. In studies of mobility over three generations, it is well-known that credible estimates require measurement models for socioeconomic outcomes (Warren and Hauser, 1997). Central conclusions in the field of mobility often hinge on assumptions about measurement.

Sibling and cousin correlations are no exception to this rule. As first noted in the context of sibling correlations (Solon et al., 1991), intragenerational correlations in measured attainment $\tilde{Y}$ will be less than correlations in true attainment Y to the extent to which attainment is reported with classical measurement error. This occurs because measurement error makes siblings (and cousins) look more different than they actually are.¹ Formally, suppose that reported income $\tilde{Y}$

¹ The measurement error that biases sibling and cousin correlations is error in the outcome

(standardized to mean 0 and variance 1) is a noisy indicator of true income Y (also standardized to mean 0 and variance 1), with reliability factor η linking the two. Suppose that reported income is uncorrelated across individuals given actual income.

$$\mathbf{E}(\tilde{Y}_g \mid \{Y_{g'}\}_{g' \leq g}) = \underbrace{\eta Y_g}_{\substack{Y_g \text{ is latent} \\ \text{and measured} \\ \text{with error}}} \quad (15)$$

$$\text{Cov}(\tilde{Y}_{gi}, \tilde{Y}_{gj} \mid Y_{gi}, Y_{gj}) = 0 \quad (16)$$

Pairing this measurement model with the Becker-Tomes transmission model (Eq. 33) yields the following sibling and cousin correlations.

$$\rho_{\text{Sibling}} = \eta^2 \beta^2 \quad (17)$$

$$\rho_{\text{Cousin}} = \eta^2 \beta^4 \quad (18)$$

$$\rho_{\text{Sibling}}^2 \leq \rho_{\text{Cousin}} \quad (19)$$

In short, measurement error depresses the sibling and cousin correlations by a factor of $\eta^2 \leq 1$. Counterintuitively, the typical practice of squaring the sibling correlation also squares the contribution of measurement error, thereby driving the squared sibling correlation down to a greater degree than the cousin correlation, which is not squared. Because the central test statistic for the Becker-Tomes model involves a comparison of the squared sibling correlation and the cousin correlation, conclusions can be especially sensitive to the presence of measurement error. Even a shift from annual to permanent income could be sufficient to alter conclusions about multi-generational persistence.

variable. This makes the problem distinct from econometric work on the use of proxies for latent predictor variables (e.g. Lubotsky and Wittenberg 2006).

3.3 Interpretation 3: From a first-order to a second-order process

Finally, evidence against the Becker-Tomes model can be interpreted in light of a second-order process under which the attainment of grandparents and offspring are directly linked through a path that does not involve the attainment of parents.²

$$\mathbf{E}(Y_g \mid \{Y_{g'}\}_{g' < g}) = \beta_{\text{Parent}} Y_{g-1} + \underbrace{\beta_{\text{Grandparent}} Y_{g-2}}_{\text{Second-order process adds this direct grandparent contribution}} \quad \forall g \quad (20)$$

Numerous theories suggest that $\beta_{\text{Grandparent}} \neq 0$. Many children today see their grandparents frequently (Dunifon et al., 2018), sometimes living in doubled-up multigenerational households (Taylor et al., 2010; Mykyta and Macartney, 2011; Pilkauskas and Martinson, 2014; Pilkauskas et al., 2014). Prior research has documented substantial multigenerational associations in education (Zeng and Xie, 2014; Sheppard and Monden, 2018), occupation (Chan and Boliver, 2013), and wealth (Pfeffer and Killewald, 2017; Adermon et al., 2018). Extended kin may compensate for the shortcomings of parents, providing resources that parents cannot (Prix and Pfeffer, 2017; Erola et al., 2018), or may serve as a cultural reference point that propels offspring toward academic achievement and attainment (Hertel and Groh-Samberg, 2014; Hällsten and Pfeffer, 2017). Although some evidence casts doubt on a direct role for grandparents in the attainment process

² I focus on a theoretical model that assumes only one income variable in each generation, even though a given person has two sets of grandparents, four sets of great-grandparents, etc. This simplification is similar to the assumption in many demographic studies of a one-sex population. In the setting of multigenerational mobility, Becker and Tomes (1979, p. 1183) motivate it as an assumption of perfect assortative mating on social origins; under this assumption, all of my grandparents have the same family incomes and can be represented by a single variable. Extending the Becker-Tomes model to allow some independence between social origins is beyond the scope of this paper but may be an important topic for future work.

(Warren and Hauser, 1997), much of the literature suggests that one exists in at least some contexts. For a variety of reasons, we might expect grandparent income to be informative about offspring income net of parent income.

Correlations in a second-order process are mathematically complicated. To build intuition for this complexity, consider the case in which parents and grandparents both directly cause the incomes of offspring. The parent-offspring correlation then includes a causal component as well as components that arise through confounding (e.g. both are influenced by the grandparents and previous ancestors). The equations below show the sibling and cousin correlations generated by a second-order process (derivations in the online appendix, part Part 1).

$$\rho_{\text{Sibling}} = \underbrace{\beta_{\text{Parent}}^2}_{\text{Direct path through shared parent}} + \underbrace{\beta_{\text{Grandparent}}^2}_{\text{Direct path through shared grandparent}} + \underbrace{\frac{2\beta_{\text{Parent}}^2\beta_{\text{Grandparent}}}{1 - \beta_{\text{Grandparent}}}}_{\text{Other paths through several ancestors}} \quad (21)$$

$$\rho_{\text{Cousin}} = \underbrace{\beta_{\text{Parent}}^2\rho_{\text{Sibling}}}_{\text{Path through parents who are siblings}} + \underbrace{\beta_{\text{Grandparent}}^2}_{\text{Direct path through shared grandparent}} + \underbrace{\frac{2\beta_{\text{Parent}}^2\beta_{\text{Grandparent}}}{1 - \beta_{\text{Grandparent}}}}_{\text{Other paths through several ancestors}} \quad (22)$$

For the present purpose, it is sufficient to note that the cousin correlation need not equal the squared sibling correlation in the formulas above.

3.4 Three interpretations: What we learn and do not learn from evidence

The sections above introduced the Becker-Tomes model as well as three theoretical extensions to processes that are dynamic, latent, or second-order. The Becker-Tomes model has only one unknown (β), but sibling and cousin correlations provide two sources of empirical information. This makes the Becker-Tomes model overidentified: the two sources of empirical evidence

can point toward different estimates of β and thus cast doubt on the validity of the model.³ Because the extensions have more unknown parameters, they are consistent with a wider range of possible sibling and cousin correlations. Thus, the data cannot in general cast doubt on these alternatives.⁴ Conclusions that favor one interpretation over another require some other reason, such as additional data or a conceptual argument, to justify the preferred interpretation over the other candidates.

3.5 Re-interpreting published evidence

Table 1 presents a few published sibling and cousin correlations. In the text, I highlight one estimate from each paper. Every author provides compelling evidence that casts doubt on the Becker-Tomes model, but the language they use to discuss this evidence often favors a second-order process in which grandparents directly influence offspring attainment. Yet, the evidence is often also consistent with a dynamic or latent process.

Analyzing educational attainment in the Wisconsin Longitudinal Study, Jæger (2012, Table 2) estimates that 14% of the variance in years of education is between extended families, and 26% of the variance is between nuclear families within extended families. These estimates indicate a sibling correlation of $0.14 + 0.26 = 0.40$, which squares to predict a cousin correlation under the Becker-Tomes model of 0.16. The estimated cousin correlation is 0.14. This evidence is consistent with several theories. The process could be dynamic, with grandparent-parent income transmission of $\beta_{g-1} = 0.58$ and parent-offspring transmission of $\beta_g = 0.64$. It could also arise from a second-order model in which parent income raises offspring income by $\beta_{\text{Parent}} = 0.68$ and grandparent

³ The Becker-Tomes model is over-identified if we assume that $\beta > 0$; since β only enters formulas in squared form, the data would be equally consistent with $\beta < 0$ of equal magnitude.

⁴ Empirical evidence can sometimes cast doubt on the latent transmission model because one subspace of parameters (cousin correlations less than the squared sibling correlation) is impossible in this model.

Author	Context	Outcome	Sibling correlation	Squared sibling correlation	Cousin correlation	Parameters consistent with estimates under three alternative models		
						Dynamic	Latent	Second-order
Jæger (2012)	Wisconsin	Years of education	0.41	0.17	0.14	$\{\beta_g = 0.64, \beta_{g-1} = 0.58\}$	None	$\{\beta_{\text{Parent}} = 0.68, \beta_{\text{Grandparent}} = -0.06\}$
		Log odds of college completion	0.56	0.31	0.21	$\{\beta_g = 0.75, \beta_{g-1} = 0.61\}$	None	$\{\beta_{\text{Parent}} = 0.89, \beta_{\text{Grandparent}} = -0.22\}$
	NLSY-CYA ^a	PIAT math score	0.38	0.14	0.23	$\{\beta_g = 0.62, \beta_{g-1} = 0.78\}$	$\{\beta = 0.78, \eta = 0.79\}$	$\{\beta_{\text{Parent}} = 0.49, \beta_{\text{Grandparent}} = 0.18\}$
		PIAT reading score	0.37	0.14	0.19	$\{\beta_g = 0.61, \beta_{g-1} = 0.72\}$	$\{\beta = 0.72, \eta = 0.85\}$	$\{\beta_{\text{Parent}} = 0.53, \beta_{\text{Grandparent}} = 0.11\}$
		Log odds of high school completion	0.37	0.14	0.14	$\{\beta_g = 0.61, \beta_{g-1} = 0.62\}$	$\{\beta = 0.62, \eta = 0.99\}$	$\{\beta_{\text{Parent}} = 0.60, \beta_{\text{Grandparent}} = 0.01\}$
		Years of education	0.37	0.14	0.26	$\{\beta_g = 0.61, \beta_{g-1} = 0.84\}$	$\{\beta = 0.84, \eta = 0.73\}$	$\{\beta_{\text{Parent}} = 0.42, \beta_{\text{Grandparent}} = 0.26\}$
		Log odds of college completion	0.45	0.20	0.29	$\{\beta_g = 0.67, \beta_{g-1} = 0.80\}$	$\{\beta = 0.80, \eta = 0.84\}$	$\{\beta_{\text{Parent}} = 0.54, \beta_{\text{Grandparent}} = 0.18\}$
Hällsten (2014)	Sweden	Years of education	0.39 ^b	0.15	0.15	$\{\beta_g = 0.62, \beta_{g-1} = 0.62\}$	None	$\{\beta_{\text{Parent}} = 0.63, \beta_{\text{Grandparent}} = 0.00\}$
		Occupational prestige	0.29	0.08	0.11	$\{\beta_g = 0.54, \beta_{g-1} = 0.62\}$	$\{\beta = 0.62, \eta = 0.87\}$	$\{\beta_{\text{Parent}} = 0.50, \beta_{\text{Grandparent}} = 0.06\}$
		GPA (9th grade)	0.51 ^b	0.26	0.19	$\{\beta_g = 0.71, \beta_{g-1} = 0.61\}$	None	$\{\beta_{\text{Parent}} = 0.81, \beta_{\text{Grandparent}} = -0.14\}$
		Cognitive ability	0.47 ^b	0.22	0.16	$\{\beta_g = 0.69, \beta_{g-1} = 0.58\}$	None	$\{\beta_{\text{Parent}} = 0.76, \beta_{\text{Grandparent}} = -0.13\}$
Knigge (2016)	Netherlands	Occupation status	0.50	0.25	0.32	$\{\beta_g = 0.71, \beta_{g-1} = 0.80\}$	$\{\beta = 0.80, \eta = 0.88\}$	$\{\beta_{\text{Parent}} = 0.60, \beta_{\text{Grandparent}} = 0.14\}$
Pfeffer et al. (2016)	U.S.	Wealth	0.34	0.12	0.19	$\{\beta_g = 0.58, \beta_{g-1} = 0.75\}$	$\{\beta = 0.75, \eta = 0.78\}$	$\{\beta_{\text{Parent}} = 0.48, \beta_{\text{Grandparent}} = 0.16\}$

Note: ^a The original author notes that the NLSY-CYA sample was young and highly selected at the time of analysis, so results should be interpreted with caution. ^b With the exception of occupational prestige, Hällsten (2014) does not directly estimate sibling correlations but cites those reported by Björklund and Jäntti (2012).

Table 1. Published estimates of cousin correlations provide evidence against the Becker-Tomes model if the cousin correlation deviates from the squared sibling correlation. Deviations can arise under a dynamic model (parameter values change across generations), a latent model (outcome is measured with error), or a second-order model (direct grandparent effects).

income directly reduces offspring income by a small amount $\beta_{\text{Grandparent}} = -0.06$. Jaeger (2012:913) concludes, “a two-generation Markov process [parent-to-child transmission] does not represent the total effect of family background on educational success.” The evidence does indeed cast doubt on a two-generation Markov process defined as the Becker-Tomes model, but it does not imply that the extended family plays any direct role in the attainment of offspring.

Hällsten (2014) references an estimate by Björklund and Jäntti (2012) that sibling’s years of education are correlated 0.39 in Sweden. Squaring this yields a Becker-Tomes prediction that the cousin correlation would be 0.15. This is almost exactly the same as the 0.15 cousin correlation that Hällsten (2014, Table 2) estimates with Swedish register data. This estimate is consistent with a static, first-order Markov process with intergenerational transmission by $\beta = 0.62$. Nevertheless, on the basis of the inability to explain away the cousin correlation with measured covariates, Hällsten (2014:31) concludes, “unless *unobserved* characteristics of parents account for *all* of the 1st and 2nd cousin correlations, the estimated and adjusted correlations are clearly incompatible with a Markov process [parent-to-child transmission],” (emphasis in original). The robustness of the cousin correlation to controls for parental characteristics is an intriguing result; this represents one example where new external information may inform multigenerational theories beyond what we can learn from a sibling and cousin correlation alone, in a way not formalized in this paper. The purpose of including this estimate in the present paper is simply to highlight that the sibling and cousin correlations alone are insufficient to provide evidence against a process of parent-to-child transmission.

Knigge (2016) studies occupational status attainment in the Netherlands in the late 19th and early 20th century, estimating a sibling correlation of 0.50. Squaring the sibling correlation yields a predicted cousin correlation of 0.25, which is slightly smaller than the estimated cousin correlation of 0.32, as the author notes (p. 1233). The conclusion that “these results are not congruent with the Markovian model” (Knigge 2016:1233) is thus true only if “the Markovian model” strictly means the static, observed, first-order Becker-Tomes model. The evidence is in fact consistent with a dynamic process (coefficient 0.80 from grandparents to parents and 0.71 from parents to offspring)

or a latent process ($\beta = 0.80$ and measurement reliability $\eta = 0.88$). Finally, it is also consistent with a view that the Markovian model is wrong and that the parent effect ($\beta_{\text{Parent}} = 0.60$) combines with a direct grandparent effect ($\beta_{\text{Grandparent}} = 0.14$) to produce offspring income. It is not clear a priori which of these theories is most conceptually plausible.

Finally, using PSID data from the U.S., Pfeffer et al. (2016) estimate a sibling correlation in log net worth of 0.34, which implies a cousin correlation of $0.34^2 = 0.12$. The authors estimate a cousin correlation of 0.19, which is slightly higher than expected and warrants the authors' conclusion that "19% of individuals' wealth attainment can be traced to the common origins of cousins (i.e. grandparent environments), reflecting concentration of family wealth within lineages beyond just two generations." The paper makes no claims about whether this multigenerational persistence violates parent-to-child transmission; it is included here only because readers may suspect that wealth transmission in the U.S. is a context in which multigenerational transmission is especially strong. The estimate is consistent with any of the three models of this paper: dynamic transmission ($\beta_{g-1} = 0.75$, $\beta_g = 0.58$), latent transmission ($\beta = 0.75$, $\eta = 0.48$), or a second-order process ($\beta_{\text{Parent}} = 0.48$, $\beta_{\text{Grandparent}} = 0.16$).

Each of the papers discussed above presents additional evidence based on intergenerational correlations to support its claims, and the purpose of this review is not to discredit the contribution of these authors. Their estimates have brought a novel framework to bear on an old question, yielding an important result: cousins' outcomes are in many settings more similar than we would expect under the Becker-Tomes model. However, the estimates do not necessarily imply the particular alternative model in which cousin similarities arise because of direct multigenerational effects. These estimates are instead consistent with a variety of alternative models, including processes that operate from parent to child but are dynamic or latent.

4 Empirical tools to estimate uncertainty and incorporate repeated measures

The review above comments solely on whether published *point* estimates diverge from the prediction of the Becker-Tomes model. This omits an entirely separate reason why estimates may be different from expected: estimation uncertainty. This section develops a Bayesian estimation procedure for sibling and cousin correlations that quantifies uncertainty about the test statistic (the cousin correlation minus the squared sibling correlation). By incorporating repeated measures, the estimator also addresses some concerns that arise from classical measurement error and life-cycle bias, each of which could produce a situation of latent transmission if left unaddressed.

4.1 Proposed model

Suppose we obtain data on the log family income Y_{it} experienced by each person i at various ages t from 25 to 45. I standardize income to mean 0 and standard deviation 0.5 for analysis,⁵ calculated by pooling all persons and years. Denote the population-average income at age t by β_t and assume each person's income is a normal draw around a person-specific mean $\mu_i^{\text{Person}} + \beta_t$ with variance $\sigma_{\text{Observation}}^2$.

$$Y_{it} \sim \mathcal{N}\left(\underbrace{\mu_i^{\text{Person}}}_{\text{Permanent income}} + \underbrace{\beta_t}_{\text{Population mean at age } t}, \underbrace{\sigma_{\text{Observation}}^2}_{\text{Observation-level variance}}\right) \quad (23)$$

The person-specific intercept μ_i^{Person} allows that each person's incomes may be shifted up or down by a fixed amount from the population average. The population mean β_t at each age t adjusts for the life-cycle bias problem that some people are observed at ages when incomes tend to be lower or higher. The observation-level variance $\sigma_{\text{Observation}}^2$ subsumes actual income volatility

⁵ I use 0.5 to follow the recommendations of Gelman et al. (2008) to place continuous variables on a scale that works well with the default priors they propose.

and classical measurement error. To examine sibling and cousin correlations in permanent income μ_i^{Person} , I decompose the variance of this quantity into three components as formalized in Eq. 1: a person-specific component δ_i , a sibling set component $\eta_{S(i)}$, and a cousin set component $\lambda_{C(i)}$. I assume a normal prior on the values of each component.

$$\delta_1, \dots, \delta_{n_{\text{Person}}} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{Person}}^2) \quad (24)$$

$$\eta_1, \dots, \eta_{n_{\text{SiblingSet}}} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{SiblingSet}}^2) \quad (25)$$

$$\lambda_1, \dots, \lambda_{n_{\text{CousinSet}}} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{CousinSet}}^2) \quad (26)$$

The central task for estimation of sibling and cousin correlations is to estimate the values of the variance components $\{\sigma_{\text{Person}}^2, \sigma_{\text{SiblingSet}}^2, \sigma_{\text{CousinSet}}^2\}$, because these imply particular values of sibling and cousin correlations (see Eq. 2–6). I place priors on the variance components and the other unknown parameters to target the search for a solution to the range of reasonable values and provide the basis for fully Bayesian inference.

$$\vec{\beta} \sim \mathcal{N}(\vec{0}, \mathbf{I}) \quad (27)$$

$$\{\sigma_{\text{Observation}}, \sigma_{\text{Person}}, \sigma_{\text{SiblingSet}}, \sigma_{\text{CousinSet}}\} \stackrel{\text{iid}}{\sim} \text{Half-Cauchy}(0, 1) \quad (28)$$

The normal prior on the age trajectory shrinks all coefficients toward zero similar to ridge regression, but in this setting the prior is vastly overwhelmed by the data because the population-mean income at each age can be estimated with considerable precision (Fig A4 in the online appendix). The half-Cauchy prior on the variance components is the absolute value of a Cauchy distribution, with scale parameter corresponding to the median of the distribution. It is very diffuse (Fig A3 in the online appendix) and is a standard choice of a weakly-informative prior for variance parameters (Gelman et al., 2008).

4.2 Data: Panel Study of Income Dynamics

I fit the model to U.S. data from the Panel Study of Income Dynamics (PSID). The PSID selected a probability sample of 1968 households and interviewed descendants from those households annually through 1999 and biannually thereafter through 2015, producing a sample of individuals nested in family lineages. I restrict to the Survey Research Center (SRC) sample, for which all households in the contiguous U.S. in 1968 had an equal probability of selection (for a discussion of other samples, see the online appendix, part Part 2). The SRC sample initially selected 3,000 households, of whom 755 produced descendants two generations later who were interviewed at ages 25-45. This reduction in the number of families reflects the combined influence of survey nonresponse, childlessness, delayed fertility, and survey attrition (see the online appendix, part Part 2). The analytic sample includes 11,730 observations on 2,363 respondents nested in 1,260 sibling sets in 755 cousin sets. A total of 1,854 respondents have at least one cousin in the sample and 1,431 respondents have at least one sibling in the sample. The outcome variable of interest is log total family income, adjusted to 2014 dollars using the Consumer Price Index. To reduce the influence of very low family incomes when logged, I bottom-code inflation-adjusted family incomes at \$5,000.

4.3 Posterior sampling

I estimate the model using the `rstan` package in R (Stan Development Team, 2017) which simulates from the posterior distribution using Hamiltonian Monte Carlo in Stan. I simulate 10,000 burn-in draws and 10,000 posterior samples for each of four chains executed sequentially from independent starting points, which produce a pooled set of 40,000 posterior draws.⁶ Point estimates reported are posterior means. Credible intervals report the 5th and 95th percentiles of posterior draws. Trace plots for key parameters of interest are provided in Fig A5 in the online appendix.

⁶ This procedure took several hours on a Windows cluster computer with 512 GB of RAM and an Intel Xeon CPU E7-4850 v3 @2.20 GHz processor.

Figure 3. The cousin correlation in permanent log family income is slightly greater than expected under the Becker-Tomes model. This empirical result does not point to any particular preferred alternative interpretation without additional assumptions. Error bars denote 95% posterior credible intervals.

4.4 Results

Fig. 3 presents results. The estimated sibling correlation is 0.38 (95% CI: 0.32–0.44). Squaring all posterior draws to predict the cousin correlation under the Becker-Tomes model yields a posterior mean estimate of 0.14 (95% CI: 0.10–0.20). The observed cousin correlation is slightly larger at 0.24 (95% CI: 0.17,0.31). The key test statistic subtracts the squared sibling correlation from the cousin correlation. For this test statistic, the estimate is 0.10 (95% CI: 0.02,0.17). The cousin correlation therefore provides some evidence against the Becker-Tomes model for transmission of permanent income. Estimation uncertainty clarifies the strength of this evidence.

Rejection of the Becker-Tomes model (Fig 3) leaves open the question of which alternative is most theoretically plausible. The results are consistent with at least three alternatives. The transmission process may be dynamic with grandparent-parent transmission $\beta_{g-1} = 0.80$ (95% CI: 0.67, 0.91) and parent-offspring transmission $\beta_g = 0.62$ (95% CI: 0.56, 0.66).⁷ The use of permanent income removes the possibility that this result could be explained by classical measurement error

⁷ The slightly higher coefficient required for grandparent-parent transmission is consistent with

at the observation level, but two alternative latent processes are possible. One is that individuals consistently over- or under-report their incomes in all waves of the survey, thereby producing classical measurement error at the person level with reliability η . A second is that the variable transmitted across generations is actually a latent variable distinct from income, and this variable is only correlated η with permanent income. The observed data are consistent with either scenario with $\beta = 0.80$ (95% CI: 0.67, 0.91) and measurement reliability $\eta = 0.78$ (95% CI: 0.64, 0.94). Finally, results are consistent with the interpretation favored by some prior researchers: a second-order process. In this case, parents affect offspring incomes by $\beta_{\text{Parent}} = 0.47$ (95% CI: 0.30, 0.61) and grandparents directly affect offspring incomes by $\beta_{\text{Grandparent}} = 0.21$ (95% CI: 0.04, 0.38). Assessing the relative merits of these three theories would require conceptual argument; the data do not point toward any one over the others.

4.5 Comparison to prior approaches

This section briefly reviews what has been gained by the Bayesian statistical model of this paper relative to other approaches. Under some assumptions, the model solves three key problems: uncertainty, life cycle bias, and measurement error.

Quantification of uncertainty is the primary advantage of the Bayesian approach. The key test statistic is a complex function of the variance components of a model, involving division, squaring, and subtraction. Posterior samples make uncertainty straightforward even in this setting: pass each posterior sample through the transformation and then summarize the distribution of posterior samples that result. It is possible to arrive at similar uncertainty estimates using frequentist methods, but doing so is mathematically complicated. Standard software packages (e.g. `nlme`, Pinheiro et al. 2019) estimate the frequentist variance of a transformation of the variance of

a supplemental result. Following the a procedure analogous to the main model, I estimate the parent generation sibling correlation to be 0.51 (95% CI: 0.46, 0.55), slightly higher than the estimate of 0.38 (95% CI: 0.32–0.44) in the offspring generation.

random intercepts at each level of the model, but mathematical derivations are required to convert these to estimates for the variance of sibling and cousin correlations (online appendix, part Part 3). It is therefore understandable that prior research has generally omitted uncertainty intervals on key quantities. A Bayesian model provides one solution to this problem.

The model also conducts an adjustment for age that addresses some concerns about life-cycle bias: the tendency of substantive conclusions in mobility research to be sensitive to the age at which income is measured (Haider and Solon, 2006). The problem arises because incomes are not stable over the life course and because not all individuals report incomes at exactly the same ages. Sibling correlations might appear large simply because the age at which incomes are reported is more similar within than between sibling sets. This source of similarity is a nuisance; theories of sibling similarities in multigenerational research are typically not interested in similarities that arise due to the age of measurement. The proposed model corrects for this problem to the degree that every individual's income trajectory is an intercept shift μ_i^{Person} away from the population-average income trajectory $\vec{\beta}$. If we observe incomes at ages 25–30 for one pair of siblings and ages 35–40 for another pair of siblings, the population-average trajectory upwardly adjusts the estimate of μ_{Person}^i in the former case and downwardly adjusts μ_{Person}^i in the latter case.

The model does not correct for life-cycle bias if income trajectories follow different patterns across individuals. If, for instance, individuals with higher long-term earnings also experience steeper earnings profiles as in the example of Haider and Solon (2006), then a homogeneous population-average trajectory will not correct for this problem. Nonetheless, the problem is not likely to be severe in this example because the reporting ages do not vary dramatically within the analytic sample: across individuals, the interquartile range of the mean age of reporting is 27.0 to 31.4. Similarity in reporting ages across individuals means that, even if the age adjustment is incorrect, results are likely to approximate the sibling and cousin correlation around the ages at which incomes are reported in the sample. The model is especially promising if we benchmark against the most standard approach of averaging income over whatever ages are observed for each individual within a specified age window (e.g. Björklund and Jäntti 2012). Incorporating the age

trajectory directly into the model is an improvement over this benchmark because it explicitly adjusts for age differences across individuals in the sample.

Net of life-cycle bias, the model addresses one type of measurement error by estimating the observation-level residual variance $\sigma_{\text{Observation}}^2$ and omitting this component from calculation of sibling and cousin correlations. The information to estimate $\sigma_{\text{Observation}}^2$ comes from the fact that each person is observed an average of 5 times, with an interquartile range from 2 to 7. If only one observation per person were available, then sibling and cousin correlations would be unable to distinguish the person-specific permanent variance component σ_{Person}^2 from the observation-level variance component $\sigma_{\text{Observation}}^2$. Pulling these apart avoids a problem where random observation-level reporting error would drive down sibling and cousin correlations (Solon et al., 1991). The model does not resolve problems of non-classical measurement error. If individuals tend to over- or under-report their incomes in the same direction over time in a way independent of all other individuals, this would inflate σ_{Person}^2 and drive down sibling and cousin correlations. This might happen, for instance, if respondents seek to make themselves look good on the survey by over-reporting income whenever they are asked. The tendency to over-report may be correlated among siblings, for instance if their parents instilled an especially high value on presenting oneself as materially successful. This would inflate $\sigma_{\text{SiblingSet}}^2$, thereby increasing sibling correlations and reducing cousin correlations. Finally, cousin correlations would be higher if a tendency to over-report is correlated among cousins, thereby inflating $\sigma_{\text{CousinSet}}^2$. Addressing these types of non-classical measurement error would likely require new forms of data about how individuals respond to surveys and whether response errors are correlated within families. I leave these tasks to future research. For the present paper, a focus on permanent income only resolves measurement error if this error is classical at the level of observations.

5 Discussion

Sibling and cousin correlations in socioeconomic outcomes serve a useful descriptive purpose, summarizing the degree to which outcomes are similar for those born into the same family. Knowledge of these quantities and their variation across outcomes, over time, and under various institutional arrangements can spur new theories about how these correlations come to be. Sibling and cousin correlations serve as descriptive facts which force us to look at the world in a new light and develop hypotheses about the relationship between family origins and socioeconomic opportunity.

Distinguishing between various theories, however, is difficult because the range of conceptual processes that could produce sibling and cousin correlations involves far more parameters than the amount of data available. This is in some sense unsurprising: data in one generation can tell us about a process occurring in past generations only under incredibly strong assumptions. To interpret a variance decomposition in terms of sibling and cousin correlations requires assumptions about independence across units. To further interpret those correlations as evidence of a dynamic, latent, or second-order process requires a conceptual argument against the other interpretations equally consistent with the observed data. To conclude that extended kin shape life chances requires assumptions that outcomes have been measured precisely and that the transmission process has unfolded with the same parameter values over the past two generations. A credible defense of these assumptions is difficult to build. For this reason, caution is warranted when using sibling and cousin correlations to draw multigenerational inferences.

Beyond sibling and cousin correlations, the claims of this paper relate to a growing body of evidence that various indicators of socioeconomic attainment persist over three or more generations to a greater degree than the Becker-Tomes model would predict (Jæger, 2012; Chan and Boliver, 2013; Hertel and Groh-Samberg, 2014; Pfeffer, 2014; Wightman and Danziger, 2014; Huang et al., 2015; Lindahl et al., 2015; Pfeffer and Killewald, 2015; Knigge, 2016; Olivetti et al., 2016; Ziefle, 2016; Hällsten and Pfeffer, 2017; Pfeffer and Killewald, 2017; Liu, 2018; Anderson et al., 2018; Colagrossi et al., 2019). This body of work clearly demonstrates that decades of past research

have overstated societal openness by leaving complexities out of the model. Already, others have pointed out that evidence of grandparent-offspring associations conditional on parent characteristics can arise from various processes (Stuhler, 2012; Braun and Stuhler, 2016; Ferrie et al., 2016; Breen, 2018; Solon, 2018). I join these authors in arguing that evidence from multigenerational and intragenerational correlations remains unable to pinpoint the precise complexity that should be added to our model. The interpretation favored by some authors (a direct role for grandparents) is only one candidate among many possibilities that could make society more rigid than Becker and Tomes (1979) predict. One model that is wrong opens the door to numerous new theories that might be more consistent with the observed data.

Proper interpretation of evidence is especially important in the domain of multigenerational mobility because of the theoretical positions at stake. As an extreme example, Clark (2014) notes in many societies that certain surnames tend to have high average attainment across many generations and concludes that genetic factors must be responsible for this high degree of persistence. Torche and Corvalan (2016) thoroughly critique this argument by clarifying that the persistence of group-level attainment (averaged over all those with a given surname) is distinct from the persistence of individual-level attainment, so that the former does not imply the latter. This level of clarity about evidence is obviously essential when a researcher claims that genetics underly mobility processes. I argue that researchers must take a similarly critical view when assessing the evidence on any theory of multigenerational transmission. Sociologists and demographers may be drawn to theories in which socioeconomic positions persist to a high degree because individuals are embedded in complex family networks that constrain opportunity socially from many directions. To support such a strong and complex theory, however, requires an accurate statement of the evidence in support of this theory. Researchers must devote at least as much thought to more mundane alternatives, such as measurement error, which could produce the empirical finding.

To apply this lens critically to the present paper, I cannot rule out direct effects of extended kin on life chances. The actual process by which income is transmitted over generations is likely to be far more complicated than the three simple models considered in this paper. All I am able to

credibly conclude is that the estimated sibling and cousin correlations could have been generated by a dynamic, latent, or second-order process, or something else I have not considered. With confidence, I can conclude that the evidence does not point cleanly toward any one of these as the correct interpretation. Sibling and cousin correlations provide a useful point of reference, but in the case of multigenerational mobility the answers they provide are far from definitive.

REFERENCES

- Adermon, A., M. Lindahl, and D. Waldenström 2018. Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations. *The Economic Journal*, 128(612):F482–F513.
- Anderson, L. R., P. Sheppard, and C. W. Monden 2018. Grandparent effects on educational outcomes: A systematic review. *Sociological Science*, 5:114–142.
- Becker, G. S. and N. Tomes 1979. An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, 87(6):1153–1189.
- Becker, G. S. and N. Tomes 1986. Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(3, Part 2):S1–S39.
- Björklund, A. and M. Jäntti 2012. How important is family background for labor-economic outcomes? *Labour Economics*, 19(4):465–474.
- Blau, P. M. and O. D. Duncan 1967. *The American Occupational Structure*. New York: Wiley.
- Bowles, S. 1972. Schooling and inequality from generation to generation. *Journal of Political Economy*, 80(3, Part 2):S219–S251.
- Braun, S. T. and J. Stuhler 2016. The transmission of inequality across multiple generations: Testing recent theories with evidence from Germany. *The Economic Journal*, 128:576–611.
- Breen, R. 2018. Some methodological problems in the study of multigenerational mobility. *European Sociological Review*, 34(6):603–611.
- Chan, T. W. and V. Boliver 2013. The grandparents effect in social mobility: Evidence from British birth cohort studies. *American Sociological Review*.
- Clark, G. 2014. *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton: Princeton University Press.
- Colagrossi, M., B. d’Hombres, and S. V. Schnepf 2019. Like (grand)parent, like child? Multigenerational mobility across the EU. *IZA Discussion Paper No. 12302*.
- Conley, D. 2004. *The Pecking Order: Which Siblings Succeed and Why*. New York: Pantheon.

- Conley, D. 2008. Bringing sibling differences in: Enlarging our understanding of the transmission of advantage in families. In *Social Class: How Does it Work?*, A. Lareau and D. Conley, eds., Pp. 179–200. New York: Russell Sage Foundation.
- Dunifon, R. E., C. E. Near, and K. M. Ziol-Guest 2018. Backup parents, playmates, friends: Grandparents' time with grandchildren. *Journal of Marriage and Family*, 80(3):752–767.
- Erola, J., E. Kilpi-Jakonen, I. Prix, and H. Lehti 2018. Resource compensation from the extended family: Grandparents, aunts, and uncles in Finland and the United States. *European Sociological Review*, 34(4):348–364.
- Ferrie, J., C. Massey, and J. Rothbaum 2016. Do grandparents and great-grandparents matter? multigenerational mobility in the U.S., 1910-2013. *National Bureau of Economic Research*.
- Gelman, A., A. Jakulin, M. G. Pittau, and Y.-S. Su 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, Pp. 1360–1383.
- Haider, S. and G. Solon 2006. Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4):1308–1320.
- Hällsten, M. 2014. Inequality across three and four generations in egalitarian sweden: 1st and 2nd cousin correlations in socio-economic outcomes. *Research in Social Stratification and Mobility*, 35:19–33.
- Hällsten, M. and F. T. Pfeffer 2017. Grand advantage: Family wealth and grandchildren's educational achievement in Sweden. *American Sociological Review*, 82(2):328–360.
- Hertel, F. R. and O. Groh-Samberg 2014. Class mobility across three generations in the US and Germany. *Research in Social Stratification and Mobility*, 35:35–52.
- Huang, J. Y., A. R. Gavin, T. S. Richardson, A. Rowhani-Rahbar, D. S. Siscovick, and D. A. Enquobahrie 2015. Are early-life socioeconomic conditions directly related to birth outcomes? Grandmaternal education, grandchild birth weight, and associated bias analyses. *American Journal of Epidemiology*, 182(7):568–578.

- Jæger, M. M. 2012. The extended family and children's educational success. *American Sociological Review*, 77(6):903–922.
- Jencks, C., M. Smith, H. Acland, M. Jo Bane, D. Cohen, H. Gintis, B. Heyns, and S. Michelson 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. Basic Books.
- Knigge, A. 2016. Beyond the parental generation: The influence of grandfathers and great-grandfathers on status attainment. *Demography*, 53(4):1219–1244.
- Lindahl, M., M. Palme, S. S. Massih, and A. Sjögren 2015. Long-term intergenerational persistence of human capital: An empirical analysis of four generations. *Journal of Human Resources*, 50(1):1–33.
- Liu, H. 2018. Social and genetic pathways in multigenerational transmission of educational attainment. *American Sociological Review*, 83(2):278–304.
- Lubotsky, D. and M. Wittenberg 2006. Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics*, 88(3):549–562.
- Mare, R. D. 2011. A multigenerational view of inequality. *Demography*, 48(1):1–23.
- Mazumder, B. 2008. Sibling similarities and economic inequality in the US. *Journal of Population Economics*, 21(3):685–701.
- Mykyta, L. and S. Macartney 2011. The effects of recession on household composition: “Doubling up” and economic well-being. *U.S. Census Bureau SEHSD Working Paper Number 2011-4*.
- Olivetti, C., M. D. Paserman, and L. Salisbury 2016. Three-generation mobility in the United States, 1850-1940: The role of maternal and paternal grandparents. *National Bureau of Economic Research*.
- Otto, S. P., F. B. Christiansen, and M. W. Feldman 1995. Genetic and cultural inheritance of continuous traits. *Morrison Institute for Population and Resource Studies Paper Number*, 64.
- Pfeffer, F. T. 2014. Multigenerational approaches to social mobility: A multifaceted research agenda. *Research in Social Stratification and Mobility*, 35:1.

- Pfeffer, F. T. and A. Killewald 2015. How rigid is the wealth structure and why? Inter-and multi-generational associations in family wealth. *Population Studies Center Research Report*, Pp. 15–845.
- Pfeffer, F. T. and A. Killewald 2017. Generations of advantage. multigenerational correlations in family wealth. *Social Forces*, 96(4):1411–1442.
- Pfeffer, F. T., A. Killewald, and A. Siliunas 2016. The concentration of wealth within family lineages and intergenerational transfers. In *Conference on New Directions in Intergenerational Transfers and Time Use in Later Life*.
- Pilkauskas, N. V., I. Garfinkel, and S. S. McLanahan 2014. The prevalence and economic value of doubling up. *Demography*, 51(5):1667–1676.
- Pilkauskas, N. V. and M. L. Martinson 2014. Three-generation family households in early childhood: Comparisons between the United States, the United Kingdom, and Australia. *Demographic Research*, 30:1639.
- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team 2019. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-141.
- Prix, I. and F. T. Pfeffer 2017. Does Donald need Uncle Scrooge? Extended family wealth and children's educational attainment in the united states. *Social Inequality Across the Generations: The Role of Resource Compensation and Multiplication in Resource Accumulation*, Pp. 112–135.
- Sheppard, P. and C. Monden 2018. The additive advantage of having educated grandfathers for children's education: Evidence from a cross-national sample in Europe. *European Sociological Review*, 34(4):365–380.
- Siegel, P. M. and R. W. Hodge 1968. A causal approach to the study of measurement error. *Methodology in Social Research*. New York: McGraw-Hill, 28(59):35.
- Solon, G. 2018. What do we know so far about multigenerational mobility? *The Economic Journal*, 128(612):F340–F352.

- Solon, G., M. Corcoran, R. Gordon, and D. Laren 1991. A longitudinal analysis of sibling correlations in economic status. *The Journal of Human Resources*, Pp. 509–534.
- Stan Development Team 2017. *RStan: the R interface to Stan*. R package version 2.16.2.
- Stuhler, J. 2012. Mobility across multiple generations: The iterated regression fallacy. *IZA Discussion Paper No. 7072*.
- Taylor, P., J. Passel, R. Fry, R. Morin, W. Wang, G. Velasco, and D. Dockterman 2010. *The Return of the Multi-Generational Family Household*. Pew Research Center.
- Torche, F. and A. Corvalan 2016. Estimating intergenerational mobility with grouped data: A critique of clark’s the son also rises. *Sociological Methods and Research*.
- Warren, J. R. and R. M. Hauser 1997. Social stratification across three generations: New evidence from the Wisconsin Longitudinal Study. *American Sociological Review*, Pp. 561–572.
- Wightman, P. and S. Danziger 2014. Multi-generational income disadvantage and the educational attainment of young adults. *Research in Social Stratification and Mobility*, 35:53–69.
- Zeng, Z. and Y. Xie 2014. The effects of grandparents on children’s schooling: Evidence from rural china. *Demography*, 51(2):599–617.
- Ziefle, A. 2016. Persistent educational advantage across three generations: Empirical evidence for Germany. *Sociological Science*, 3:1077–1102.

Online appendix for “Does Opportunity Skip Generations”

Ian Lundberg

ilundberg@princeton.edu

Part 1 Sibling and cousin correlations implied by multigenerational models

This section develops the sibling and cousin correlations implied by various theories of multigenerational mobility. Three aspects of the formal models are worth clarifying before moving into any proofs. First, I assume all random variables are standardized to mean 0 and variance 1. This implies that every marginal covariance between two random variables equals the marginal correlation between those random variables. This simplifies some notation. Second, I follow the standard in the literature by assuming perfect assortative mating on social origins (Becker and Tomes, 1979, p. 1183). This simplifies the demographic problem that each individual has one set of parents, two sets of grandparents, four sets of great-grandparents, and so on. Under perfect sorting, income is the same for all ancestors within a single generation and each generation can be represented by a single variable. Third, every theory considered in this paper invokes a non-interference assumption that the incomes of any two individuals within a generation are independent given the incomes of their ancestors. This is consistent, for example, with the assumption in the empirical model that siblings do not affect each other.

$$Y_{gi} \perp\!\!\!\perp Y_{gj} \mid \{Y_{g'i}, Y_{g'j}\}_{g' < g} \quad (29)$$

Under these assumptions, the income correlation for any two individuals i and j is the covariance of their expected values given the past.

$$\text{Cor}(Y_i^t, Y_j^t) = \text{Cov}(Y_i^t, Y_j^t) \quad \text{since standardized} \quad (30)$$

$$= \underbrace{\text{E}(\text{Cov}(Y_{gi}, Y_{gj}) \mid \{Y_{g'i}, Y_{g'j}\}_{g' < g})}_{=0 \text{ under independence assumption}} \quad \text{by law of total covariance} \\ + \text{Cov}(\text{E}(Y_{gi} \mid \{Y_{g'i}\}_{g' < g}), \text{E}(Y_g \mid \{Y_{g'j}\}_{g' < g})) \quad (31)$$

$$= \text{Cov}(\text{E}(Y_{gi} \mid \{Y_{g'i}\}_{g' < g}), \text{E}(Y_g \mid \{Y_{g'j}\}_{g' < g})) \quad (32)$$

Eq. 32 becomes the basis for many of the proofs that follow.

Before moving to extensions, I briefly review the implications of the Becker-Tomes model for sibling and cousin correlations. The model (as written in the main text) is copied below.

$$\text{E}(Y_g \mid \{Y_{g'}\}_{g' < g}) = \beta Y_{g-1} \quad \forall g \quad (33)$$

This model yields the following formula for the permanent income correlation of two individuals i and j .

$$\text{Cor}(Y_i^t, Y_j^t) = \text{Cov}(\text{E}(Y_{gi} \mid \{Y_{g'i}\}_{g' < g}), \text{E}(Y_g \mid \{Y_{g'j}\}_{g' < g})) \quad \text{by Eq. 32} \quad (34)$$

$$= \text{Cov}(\beta Y_{g-1,i}, \beta Y_{g-1,j}) \quad (35)$$

$$= \beta^2 \text{Cov}(Y_{g-1,i}, Y_{g-1,j}) \quad (36)$$

$$= \beta^2 \text{Cor}(Y_{g-1,i}, Y_{g-1,j}) \quad \text{since variables standardized} \quad (37)$$

Siblings i and j have the same parent, so $\text{Cor}(Y_{g-1,i} = Y_{g-1,j}) = 1$. Cousins i and j have parents who are siblings, so $\text{Cor}(Y_{g-1,i}, Y_{g-1,j}) = \rho_{\text{Sibling}}$. Plugging these facts into Eq. 37 yields three

results.

$$\rho_{\text{Sibling}} = \beta^2 \quad (38)$$

$$\rho_{\text{Cousin}} = \beta^2 \rho_{\text{Sibling}} = \beta^4 \quad (39)$$

$$\rho_{\text{Cousin}} = \rho_{\text{Sibling}}^2 \quad (40)$$

Rearranging yields two formulas for β .

$$\beta = \pm \sqrt{\rho_{\text{Sibling}}} \quad (41)$$

$$\beta = \pm \rho_{\text{Cousin}}^{\frac{1}{4}} \quad (42)$$

As shown above, the sign of β is not identified by sibling and cousin correlations because β enters the formula for these correlations only in squared form. Because we can reasonably assume that $\beta > 0$ so that high parent incomes are associated with high offspring incomes, I say no more about this lack of identification. Because there are two empirical sources of information (sibling and cousin correlations) about one theoretical parameter (β), this parameter is overidentified by the Becker-Tomes model with data on sibling and cousin correlations. This allows us to assess evidence against the Becker-Tomes model.

1.1 Dynamic process

A dynamic process allows the intergenerational elasticity β_g to differ across generations.

$$\mathbb{E}(Y_g \mid \{Y_{g'}\}_{g' < g}) = \beta_g Y_{g-1} \quad \forall g \quad (43)$$

This model yields the following formula for the permanent income correlation of two individuals i and j .

$$\text{Cor}(Y_i^t, Y_j^t) = \text{Cov}(\mathbf{E}(Y_{gi} \mid \{Y_{g'i}\}_{g' < g}), \mathbf{E}(Y_g \mid \{Y_{g'j}\}_{g' < g})) \quad \text{by Eq. 32} \quad (44)$$

$$= \text{Cov}(\beta_g Y_{g-1,i}, \beta_g Y_{g-1,j}) \quad (45)$$

$$= \beta_g^2 \text{Cov}(Y_{g-1,i}, Y_{g-1,j}) \quad (46)$$

$$= \beta_g^2 \text{Cor}(Y_{g-1,i}, Y_{g-1,j}) \quad \text{since variables standardized} \quad (47)$$

Siblings i and j have the same parent, so $\text{Cor}(Y_{g-1,i} = Y_{g-1,j}) = 1$. Cousins i and j have parents who are siblings, so $\text{Cor}(Y_{g-1,i}, Y_{g-1,j}) = \rho_{g-1, \text{Sibling}}$. Plugging into Eq. 47 yields the following formulas.

$$\underbrace{\rho_{g, \text{Sibling}}}_{\text{Sibling correlation in generation } g} = \underbrace{\beta_g^2}_{\text{Squared transmission generation } g-1 \text{ to } g} \quad (48)$$

$$\underbrace{\rho_{g, \text{Cousin}}}_{\text{Cousin correlation in generation } g} = \beta_g^2 \rho_{g-1, \text{Sibling}} = \underbrace{\beta_g^2}_{\text{Squared transmission generation } g-1 \text{ to } g} \underbrace{\beta_{g-1}^2}_{\text{Squared transmission generation } g-2 \text{ to } g-1} \quad (49)$$

The cousin correlation can therefore be less than, equal to, or greater than the squared sibling correlation depending on the values of β_g and β_{g-1} .

Rearranging, a dynamic process implies the following parameter estimates for a given set of sibling and cousin correlations.

$$\beta_g = \pm \sqrt{\rho_{g, \text{Sibling}}} \quad (50)$$

$$\beta_{g-1} = \pm \sqrt{\frac{\rho_{g, \text{Cousin}}}{\rho_{g, \text{Sibling}}}} \quad (51)$$

As in the previous section, I assume the positive values for β_g and β_{g-1} . The parameters of the dynamic model are just-identified because there are two empirical estimates and two unknown

parameters. We can estimate the parameters from a given set of sibling and cousin correlations but cannot provide any evidence against the dynamic process.

1.2 Latent process

A latent process assumes the Becker-Tomes model is correct except that measured income $\tilde{Y}_g$ is an indicator of income Y_g with classical measurement error.

$$\mathbf{E}(\tilde{Y}_g \mid \{Y_{g'}\}_{g' \leq g}) = \eta Y_g \quad (52)$$

$$\text{Cov}(\tilde{Y}_{gi}, \tilde{Y}_{gj} \mid Y_{gi}, Y_{gj}) = 0 \quad (53)$$

This model yields the following formula for the measured income correlation of two individuals i and j .

$$\text{Cor}(\tilde{Y}_{gi}, \tilde{Y}_{gj}) = \text{Cov}(\tilde{Y}_{gi}, \tilde{Y}_{gj}) \quad \text{since standardized} \quad (54)$$

$$\begin{aligned} &= \underbrace{\mathbf{E}(\text{Cov}(\tilde{Y}_{gi}, \tilde{Y}_{gj} \mid Y_{gi}, Y_{gj}))}_{=0 \text{ by Eq. 53}} \\ &\quad + \text{Cov}(\mathbf{E}(\tilde{Y}_{gi} \mid Y_{gi}), \mathbf{E}(\tilde{Y}_{gj} \mid Y_{gj})) \end{aligned} \quad (55)$$

$$= \text{Cov}(\eta Y_{gi}, \eta Y_{gj}) \quad \text{by Eq. 52} \quad (56)$$

$$= \eta^2 \text{Cov}(Y_{gi}, Y_{gj}) \quad (57)$$

$$= \eta^2 \text{Cor}(Y_{gi}, Y_{gj}) \quad \text{since standardized} \quad (58)$$

Since the latent process does not modify the Becker-Tomes model for the transmission of actual income Y , I plug the results from the Becker-Tomes model into this equation to get three results.

$$\rho_{\text{Sibling}} = \eta^2 \beta^2 \quad (59)$$

$$\rho_{\text{Cousin}} = \eta^2 \beta^4 \quad (60)$$

$$\rho_{\text{Cousin}} \geq \rho_{\text{Sibling}}^2 \quad \text{since } \eta \leq 1 \quad (61)$$

Rearranging, a latent process implies the following parameter estimates for a given set of sibling and cousin correlations.

$$\eta = \sqrt{\frac{\rho_{\text{Sibling}}^2}{\rho_{\text{Cousin}}}} \quad (62)$$

$$\beta = \pm \sqrt{\frac{\rho_{\text{Cousin}}}{\rho_{\text{Sibling}}}} \quad (63)$$

Throughout the paper, I focus only on the positive root for β because it seems theoretically implausible that the intergenerational elasticity would be negative. η cannot be negative under the assumption of classical measurement error. The parameters of the latent model are just-identified because there are two empirical estimates and two unknown parameters. However, as noted previously this model can only generate sibling and cousin correlations with $\rho_{\text{Cousin}} \geq \rho_{\text{Sibling}}^2$. Evidence that contradicts this prediction can provide evidence against the latent process.

1.3 Second-order process

Finally, I consider the second-order process favored in the interpretations of some prior research.

$$\mathbf{E}(Y_g \mid \{Y_{g'}\}_{g' < g}) = \beta_{\text{Parent}} Y_{g-1} + \beta_{\text{Grandparent}} Y_{g-2} \quad \forall g \quad (64)$$

Before deriving the sibling and cousin correlations, it is necessary to derive the parent-offspring correlation under this model.

$$\rho_{\text{Intergenerational}} = \text{Cor}(Y_{gi}, Y_{g-1,i}) \quad (65)$$

$$= \text{Cov}(Y_{gi}, Y_{g-1,i}) \quad (66)$$

$$= \underbrace{\text{E}(\text{Cov}(Y_{gi}, Y_{g-1,i} \mid \{Y_{g'i}\}_{g' < g}))}_{=0 \text{ since } Y_{g-1,i} \text{ in conditioning set}} + \text{Cov}(\text{E}(Y_{gi} \mid \{Y_{g'i}\}_{g' < g}), \text{E}(Y_{gj}, Y_{g-1,i} \mid \{Y_{g'i}\}_{g' < g})) \quad (67)$$

$$= \text{Cov}(\beta_{\text{Parent}} Y_{g-1,i} + \beta_{\text{Grandparent}} Y_{g-2,i}, Y_{g-1,i}) \quad (68)$$

$$= \beta_{\text{Parent}} \text{Cov}(Y_{g-1,i}, Y_{g-1,i}) + \beta_{\text{Grandparent}} \text{Cov}(Y_{g-1,i}, Y_{g-2,i}) \quad (69)$$

$$= \beta_{\text{Parent}} + \beta_{\text{Grandparent}} \rho_{\text{Intergenerational}} \quad (70)$$

$$\rho_{\text{Intergenerational}} = \frac{\beta_{\text{Parent}}}{1 - \beta_{\text{Grandparent}}} \quad (71)$$

This model yields the following formula for the permanent income correlation of two individuals i and j .

$$\text{Cor}(Y_i^t, Y_j^t) = \underbrace{\text{Cov}(\text{E}(Y_{gi} \mid \{Y_{g'i}\}_{g' < g}), \text{E}(Y_{gj} \mid \{Y_{g'j}\}_{g' < g}))}_{\text{by Eq. 32}} \quad (72)$$

$$= \text{Cov}(\beta_{\text{Parent}} Y_{g-1,i} + \beta_{\text{Grandparent}} Y_{g-2,i}, \beta_{\text{Parent}} Y_{g-1,j} + \beta_{\text{Grandparent}} Y_{g-2,j}) \quad (73)$$

$$= \beta_{\text{Parent}}^2 \text{Cov}(Y_{g-1,i}, Y_{g-1,j}) + \beta_{\text{Grandparent}}^2 \text{Cov}(Y_{g-2,i}, Y_{g-2,j}) + \beta_{\text{Parent}} \beta_{\text{Grandparent}} (\text{Cov}(Y_{g-1,i}, Y_{g-2,j}) + \text{Cov}(Y_{g-1,j}, Y_{g-2,i})) \quad (74)$$

For siblings i and j , the parent incomes are equal and the grandparent incomes are equal. Further, $Y_{g-1,i}$ is the offspring income from $Y_{g-2,j}$, so this covariance equals the intergenerational

correlation $\rho_{\text{Intergenerational}}$.

$$\rho_{\text{Sibling}} = \text{Cor}(Y_i^t, Y_j^t) \quad (75)$$

$$\begin{aligned} &= \beta_{\text{Parent}}^2 \underbrace{\text{Cov}(Y_{g-1,i}, Y_{g-1,j})}_{=1} + \beta_{\text{Grandparent}}^2 \underbrace{\text{Cov}(Y_{g-2,i}, Y_{g-2,j})}_{=1} \\ &\quad + \beta_{\text{Parent}}\beta_{\text{Grandparent}} \underbrace{(\text{Cov}(Y_{g-1,i}, Y_{g-2,j}) + \text{Cov}(Y_{g-1,j}, Y_{g-2,i}))}_{=2\rho_{\text{Intergenerational}}} \end{aligned} \quad (76)$$

$$= \beta_{\text{Parent}}^2 + \beta_{\text{Grandparent}}^2 + \frac{2\beta_{\text{Parent}}^2\beta_{\text{Grandparent}}}{1 - \beta_{\text{Grandparent}}} \quad (77)$$

Cousins i and j have parents who are siblings and have the same grandparents. Because cousins' parents are children of the same grandparents, it is still the case that the $\text{Cov}(Y_{g-2}, Y_{g-1})$ terms are intergenerational correlations.

$$\rho_{\text{Cousin}} = \text{Cor}(Y_i^t, Y_j^t) \quad (78)$$

$$\begin{aligned} &= \beta_{\text{Parent}}^2 \underbrace{\text{Cov}(Y_{g-1,i}, Y_{g-1,j})}_{=1} + \beta_{\text{Grandparent}}^2 \underbrace{\text{Cov}(Y_{g-2,i}, Y_{g-2,j})}_{=1} \\ &\quad + \beta_{\text{Parent}}\beta_{\text{Grandparent}} \underbrace{(\text{Cov}(Y_{g-1,i}, Y_{g-2,j}) + \text{Cov}(Y_{g-1,j}, Y_{g-2,i}))}_{=2\rho_{\text{Intergenerational}}} \end{aligned} \quad (79)$$

$$= \beta_{\text{Parent}}^2\rho_{\text{Sibling}} + \beta_{\text{Grandparent}}^2 + \frac{2\beta_{\text{Parent}}^2\beta_{\text{Grandparent}}}{1 - \beta_{\text{Grandparent}}} \quad (80)$$

By rearranging, it is possible to solve for the parameter estimates $\{\beta_{\text{Parent}}, \beta_{\text{Grandparent}}\}$ that could generate an observed set of sibling and cousin correlations. The formula for β_{Parent} is analytically simple.

$$\beta_{\text{Parent}} = \pm \sqrt{\frac{\rho_{\text{Sibling}} - \rho_{\text{Cousin}}}{1 - \rho_{\text{Sibling}}}} \quad (81)$$

The formula for $\beta_{\text{Grandparent}}$ involves solving a cubic polynomial.

$$\rho_{\text{Sibling}} = \beta_{\text{Parent}}^2 + \beta_{\text{Grandparent}}^2 + \frac{2\beta_{\text{Parent}}^2\beta_{\text{Grandparent}}}{1 - \beta_{\text{Grandparent}}} \quad (82)$$

$$\begin{aligned} \rho_{\text{Sibling}} - \rho_{\text{Sibling}}\beta_{\text{Grandparent}} &= \beta_{\text{Parent}}^2 - \beta_{\text{Parent}}^2\beta_{\text{Grandparent}} \\ &\quad + \beta_{\text{Grandparent}}^2 - \beta_{\text{Grandparent}}^3 \\ &\quad + 2\beta_{\text{Parent}}^2\beta_{\text{Grandparent}} \end{aligned} \quad (83)$$

$$\begin{aligned} 0 &= -\beta_{\text{Grandparent}}^3 \\ &\quad + \beta_{\text{Grandparent}}^2 \\ &\quad + (\beta_{\text{Parent}}^2 + \rho_{\text{Sibling}})\beta_{\text{Grandparent}} \\ &\quad + (\beta_{\text{Parent}}^2 - \rho_{\text{Sibling}}) \end{aligned} \quad (84)$$

It is therefore possible to solve for β_{Parent} analytically and then solve for $\beta_{\text{Grandparent}}$ as the roots of the cubic polynomial above. I assume that $\beta_{\text{Parent}} > 0$ and report the root for $\beta_{\text{Grandparent}}$ which is closest to zero. Because the second-order process is consistent with any set of sibling and cousin correlations, estimated sibling and cousin correlations cannot provide evidence against the second-order process.

Part 2 PSID sample selection

The original 1968 PSID sample contains two groups: the Survey of Economic Opportunity (SEO) sample and the Survey Research Center (SRC) sample. The SRC sample is a multistage probability sample of households in the contiguous 48 states in which every household has equal probability of selection. The SEO sampled households residing in Standard Metropolitan Statistical Areas (SMSAs), with the exception of the South which included non-SMSAs, and is drawn from a sample used by the Census Bureau to study economic opportunity among low-income households. A substantial share of the SEO sample was dropped in 1997 to reduce costs. In the SRC sample, 89% of observations on grandchildren at ages 25-45 occur in 1997 or later, rendering the 1997 reduction in the SEO sample important for this particular application. Because the two samples are distinct and because it is much easier to assess the representativeness of the SRC sample, I focus on results for the SRC sample in the main text and do not weight any analyses. Fitting the model to the SEO sample yields a different set of results: sibling and cousin correlations are quite consistent with the Becker-Tomes model (Fig A1). However, I hesitate to draw firm conclusions from the SEO sample because of its unusual design and subsequent reduction in sample size.

In addition to the samples that began in 1968, the PSID added refresher samples of Latino and immigrant families in 1990 and 1997, respectively. I exclude these samples from all analysis because they have not been followed long enough to produce sizable samples of grandchildren descended from the sampled families.

Not all families from the initial 1968 SRC sample have descendants who contribute to the analytic sample, reflecting a combination of fertility processes and survey attrition. As shown in Table A1, the analytic sample includes grandchildren descended from only 755 of the 3,000 households originally sampled. Some households initially sampled in 1968 were childless or had children who had already moved out of the home before the initial interview, and therefore were not followed. A total of 1,754 families have descendants interviewed in the PSID. Among these, 1,226 families have at least one grandchild interviewed in the PSID. These reductions reflect a

Figure A1. SEO sample estimates of sibling and cousin correlations in log family income. Error bars denote 95% credible intervals. Analogous SRC sample estimates are provided in the main text, Figure 3.

combination of survey attrition and childlessness. There are 1,185 households for whom parent and grandparent information is complete, when the data are merged with the Family Identification Mapping System prospective three-generation file. Among these, 755 households have grandchildren observed at ages 25–45. These households are the analytic sample.

	SRC sample (main text)	SEO sample (appendix)
Sampled households	3,000	2,000
Interviewed in 1968	2,930	1,872
With descendants in the PSID	1,754	1,237
With grandchildren in the PSID	1,226	996
With complete lineage information	1,185	956
Analytic sample observed at ages 25-45:		
Cousin sets	755	504
Sibling sets	1,260	887
Persons	2,363	1,667
Observations	11,730	8,201

Table A1. PSID sample restrictions. The Survey Research Center (SRC) sample is an equal-probability sample of American households in 1968, whereas the Survey of Economic Opportunity (SEO) sample is a sample of low-income households selected through a more complex process. The main text reports estimates on the cross-sectional SRC sample (Fig 3). The online appendix reports estimates using the SEO sample (Fig A1).

Part 3 Frequentist estimates are similar

This paper advocates Bayesian estimation of sibling and cousin correlations to facilitate quantification of uncertainty. This section develops a frequentist approach to uncertainty using the delta method. Results are nearly identical to the Bayesian approach, offering an alternative for those unwilling to rely on the assumed prior distributions for inference. However, this alternative relies on a mathematical derivation that is complicated.

Use $\vec{\theta}$ to denote the log standard deviation of the four variance components.

$$\vec{\theta} = \begin{bmatrix} \log(\sigma_{\text{CousinSet}}) \\ \log(\sigma_{\text{SiblingSet}}) \\ \log(\sigma_{\text{Person}}) \\ \log(\sigma_{\text{Observation}}) \end{bmatrix} \quad (85)$$

The model proposed in this paper (minus the priors) can be fit by maximum likelihood in the `nlme` package in R (Pinheiro et al., 2019). From maximum likelihood theory, $\hat{\vec{\theta}}$ follows a sampling distribution which is asymptotically normal with mean $\vec{\theta}$ and variance-covariance matrix Σ , both of which are returned by the `nlme` package. To translate these estimates into a variance-covariance matrix for the sibling and cousin correlations, I apply the delta method by defining functions $f()$, $g()$, and $h()$ that convert the estimated parameters to sibling and cousin correlations and the test statistic t .

$$\rho_{\text{Cousin}} = f\left(\vec{\theta}\right) \quad (86)$$

$$= \frac{\exp(\theta_1)^2}{\exp(\theta_1)^2 + \exp(\theta_2)^2 + \exp(\theta_3)^2} \quad (87)$$

$$= \frac{\exp(2\theta_1)}{\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)} \quad (88)$$

$$\rho_{\text{Sibling}} = g\left(\vec{\theta}\right) \quad (89)$$

$$= \frac{\exp(\theta_1)^2 + \exp(\theta_2)^2}{\exp(\theta_1)^2 + \exp(\theta_2)^2 + \exp(\theta_3)^2} \quad (90)$$

$$= \frac{\exp(2\theta_1) + \exp(2\theta_2)}{\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)} \quad (91)$$

$$\tau = \rho_{\text{Cousin}} - \rho_{\text{Sibling}}^2 \quad (92)$$

$$= h(\vec{\theta}) \quad (93)$$

$$= f\left(\vec{\theta}\right) - \left(g\left(\vec{\theta}\right)\right)^2 \quad (94)$$

The delta method approximates these functions by a first-order Taylor expansion.

$$f\left(\hat{\vec{\theta}}\right) \approx f\left(\vec{\theta}\right) + \left[\nabla f\left(\vec{\theta}\right)\right]^T \left(\hat{\vec{\theta}} - \vec{\theta}\right) \quad (95)$$

$$g\left(\hat{\vec{\theta}}\right) \approx g\left(\vec{\theta}\right) + \left[\nabla g\left(\vec{\theta}\right)\right]^T \left(\hat{\vec{\theta}} - \vec{\theta}\right) \quad (96)$$

$$\left(g\left(\hat{\vec{\theta}}\right)\right)^2 \approx \left(g\left(\vec{\theta}\right)\right)^2 + 2g\left(\vec{\theta}\right) \left[\nabla g\left(\vec{\theta}\right)\right]^T \left(\hat{\vec{\theta}} - \vec{\theta}\right) \quad (97)$$

$$h\left(\hat{\vec{\theta}}\right) \approx h\left(\vec{\theta}\right) + \left[\nabla h\left(\vec{\theta}\right)\right]^T \left(\hat{\vec{\theta}} - \vec{\theta}\right) \quad (98)$$

This yields a formula for the variance of each estimate.

$$\mathbf{V}(\hat{\rho}_{\text{Cousin}}) = \mathbf{V}\left[f\left(\hat{\vec{\theta}}\right)\right] \approx \left[\nabla f\left(\vec{\theta}\right)\right]^T \underbrace{\mathbf{V}\left(\hat{\vec{\theta}}\right)}_{=\Sigma} \left[\nabla f\left(\vec{\theta}\right)\right] \quad (99)$$

$$\mathbf{V}(\hat{\rho}_{\text{Sibling}}) = \mathbf{V}\left[g\left(\hat{\vec{\theta}}\right)\right] \approx \left[\nabla g\left(\vec{\theta}\right)\right]^T \underbrace{\mathbf{V}\left(\hat{\vec{\theta}}\right)}_{=\Sigma} \left[\nabla g\left(\vec{\theta}\right)\right] \quad (100)$$

$$\mathbf{V}(\hat{\rho}_{\text{Sibling}}^2) = \mathbf{V}\left[\left(g\left(\hat{\vec{\theta}}\right)\right)^2\right] \approx \left(2g\left(\vec{\theta}\right)\left[\nabla g\left(\vec{\theta}\right)\right]\right)^T \underbrace{\mathbf{V}\left(\hat{\vec{\theta}}\right)}_{=\Sigma} \left(2g\left(\vec{\theta}\right)\left[\nabla g\left(\vec{\theta}\right)\right]\right) \quad (101)$$

$$\mathbf{V}(\hat{\tau}) = \mathbf{V}\left[h\left(\hat{\vec{\theta}}\right)\right] \approx \left[\nabla h\left(\vec{\theta}\right)\right]^T \underbrace{\mathbf{V}\left(\hat{\vec{\theta}}\right)}_{=\Sigma} \left[\nabla h\left(\vec{\theta}\right)\right] \quad (102)$$

The only remaining step is to calculate each gradient and plug in to produce an estimate.

For the cousin correlation function h ,

$$\frac{\partial f(\vec{\theta})}{\partial \theta_1} = \frac{2 \exp(2\theta_1) [\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)] - 2 \exp(2\theta_1) \exp(2\theta_1)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (103)$$

$$= \frac{2 \exp(2\theta_1 + 2\theta_2) + 2 \exp(2\theta_1 + 2\theta_3)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (104)$$

$$\frac{\partial f(\vec{\theta})}{\partial \theta_2} = \frac{-\exp(2\theta_1) 2 \exp(2\theta_2)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (105)$$

$$= \frac{-2 \exp(2\theta_1 + 2\theta_2)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (106)$$

$$\frac{\partial f(\vec{\theta})}{\partial \theta_3} = \frac{-2 \exp(2\theta_1 + 2\theta_3)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (107)$$

For the sibling correlation function g ,

$$\frac{\partial g(\vec{\theta})}{\partial \theta_1} = \frac{2 \exp(2\theta_1) [\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)] - 2 \exp(2\theta_1) [\exp(2\theta_1) + \exp(2\theta_2)]}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (108)$$

$$= \frac{2 \exp(2\theta_1 + 2\theta_3)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (109)$$

$$\frac{\partial g(\vec{\theta})}{\partial \theta_2} = \frac{2 \exp(2\theta_2 + 2\theta_3)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (110)$$

$$\frac{\partial g(\vec{\theta})}{\partial \theta_3} = \frac{-2 \exp(2\theta_3) [\exp(2\theta_1) + \exp(2\theta_2)]}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (111)$$

$$= \frac{-2 \exp(2\theta_1 + 2\theta_3) - 2 \exp(2\theta_2 + 2\theta_3)}{[\exp(2\theta_1) + \exp(2\theta_2) + \exp(2\theta_3)]^2} \quad (112)$$

For the test statistic function h , the formula is the same for each component $i = 1, \dots, 3$.

$$\frac{\partial h(\vec{\theta})}{\partial \theta_i} = \frac{\partial f(\vec{\theta})}{\partial \theta_i} - 2g(\vec{\theta}) \frac{\partial g(\vec{\theta})}{\partial \theta_i} \quad (113)$$

Plugging the estimated $\hat{\vec{\theta}}$ into these gradients and plugging the gradients into the variance formulas (Eq. 99–102) yields an analytical estimate of the variance of the cousin correlation, sibling correlation, and test statistic. Because $\hat{\vec{\theta}}$ is asymptotically normal under maximum likelihood theory and these transformed parameters are linear combinations of $\hat{\vec{\theta}}$ (to a first-order approximation), the transformed parameters are also asymptotically normal. This justifies the construction of 95% asymptotic confidence intervals by a normal approximation with the estimated variance. Fig A2 presents the Bayesian and frequentist point estimates with 95% intervals and shows that they are nearly identical.

Figure A2. Frequentist estimates following the procedure of Sec. Part 3 are nearly identical to the Bayesian estimates reported in the main text.

Part 4 Supplemental figures referenced in main text

Figure A3. Graphical depiction of half-cauchy prior distribution on the random intercepts at the person, sibling set, and cousin set level, as well as the observation-level standard deviation.

Figure A4. Estimated association between age and log family income

Figure A5. Trace plots showing Hamiltonian Monte Carlo sampling for the standard deviation of random intercepts at each level.