Received: 3 June 2016 | Revised: 6 March 2017
DOI: 10.1002/jae.2578

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Decomposing economic mobility transition matrices

Jeremiah Richey[1] | Alicia Rosburg[2]

[1]Office of the Comptroller of the Currency, Washington, DC, USA

[2]University of Northern Iowa, Cedar Falls, IA, USA

Correspondence
Jeremiah Richey, Office of the Comptroller of the Currency, 400 7th St SW, Washington, DC 20219, USA.
Email: Jeremiah.richey@occ.treas.gov

Summary

We present a decomposition method for transition matrices to identify forces driving the persistence of economic status across generations. The method decomposes differences between an estimated transition matrix and a benchmark transition matrix into portions attributable to differences in characteristics between individuals from different households (a composition effect) and portions attributable to differing returns to these characteristics (a structure effect). A detailed decomposition based on copula theory further decomposes the composition effect into portions attributable to specific characteristics and their interactions. To examine potential drivers of economic persistence in the USA, we apply the method to white males from the 1979 US National Longitudinal Survey of Youth. Depending on the transition matrix entry of interest, differing characteristics between sons from different households explain between 40% and 70% of observed income persistence, with differing returns for these characteristics explaining the remaining gap. Further, detailed decompositions reveal significant heterogeneity in the role played by specific characteristics (e.g., education) across the income distribution.

1 | INTRODUCTION

A topic of recurring interest among social scientists is the persistence of economic status between parent and offspring.[1] While great strides have been made to improve empirical estimates of income persistence (see Solon, 1999, for a review and Mazumder, 2005), for a recent reassessment), much remains to be understood about the drivers or determinants of this persistence (Black & Devereux, 2011). The existing literature that has sought to understand the driving mechanisms of this persistence has mainly focused on mean effects such as the intergenerational elasticity of income (IGE) (Björklund, Lindahl, & Plug, 2006; Blanden, Gregg, & Macmillan, 2007; Bowles & Gintis, 2002; Cardak, Johnston, & Martin, 2013; Lefgren, Lindquist, & Sims, 2012; Liu & Zeng, 2009; Mayer & Lopoo, 2008; Richey & Rosburg, 2017; Shea, 2000). While informative, summary measures such as the IGE may conceal interesting details regarding differences in economic mobility at different points across the distribution (Black & Devereux, 2011). Alternative measures such as transition matrices provide a more "complete" picture of intergenerational persistence by looking across the entire income distribution (Bhattacharya & Mazumder, 2011; Black & Devereux, 2011; Jäntti et al., 2006). The literature on transition matrices, however, has focused on point estimation rather than understanding the driving mechanisms behind the estimates.[2] A closer examination of the drivers underlying the transition matrix would improve

[1]Interest in intergenerational status persistence in general has a long history, especially in the field of sociology (see Blau & Duncan, 1967). While sociologists have been primarily concerned with mobility in social class positions (e.g., occupations), economists have focused more on income or wealth mobility (Erikson & Goldthorpe, 2002). Our discussion will focus solely on income mobility, which has a large recent history in the economics literature.
[2]One notable exception in the intergenerational mobility setting is Bhattacharya and Mazumder (2011). They propose a nonparametric approach to estimate *conditional* transition matrices and apply the approach to better understand black–white differences in mobility. However, the curse of dimensionality limits their analysis to the role of a single covariate at a time, and the authors explore the role of cognitive ability and education.

our understanding of income persistence and the ability to evaluate policies aimed at equality of opportunity. With this goal in mind, we propose a “decomposition” method for transition matrices and use the method to examine drivers of economic persistence in the USA.

More generally, transition matrices are a tool that can be used to understand complex links in the distribution of a variable over time. Prime examples include, but are not limited to, intragenerational mobility and intergenerational mobility, as well as credit risk ratings migrations. The method we propose explores the role of intermediate variables within these links. In particular, we focus on intergenerational economic mobility transition matrices as both an important application of the method and as a working example to ease exposition. Thus our contribution is twofold: we provide a method to understand the drivers of any transition matrix, and we add to the economic mobility literature by broadening our understanding of the drivers of economic persistence in the USA.

The method we propose is not a decomposition in the traditional sense but builds directly on the decomposition literature. Traditional decompositions seek to understand differences in a variable between two groups, such as differences in incomes between men and women. Our goal is to understand the link between a variable over time, such as incomes across generations. We frame this as trying to understand differences between an (actual) empirical transition matrix and a hypothetical transition matrix where there is no link between parental and children's incomes (i.e., an “independent” transition matrix). We decompose these differences into two components: (1) the portion attributable to differing characteristics (e.g., education, experience) between children from different households (referred to as the composition effect); and (2) the portion attributable to differing returns to characteristics—a wage structure—between children from different households (referred to as the structure effect). We then extend the basic decomposition using copula theory to perform a detailed decomposition on the composition effect Rothe (2015). The detailed decomposition attributes portions of the composition effect to specific variables (e.g., education, experience) and their interactions.

Our proposed method begins by recognizing that transition matrices are representations of multiple conditional distributions (e.g., distributions of children's incomes conditional on parental income grouping). Analogously, we construct a hypothetical independent matrix with multiple, but identical, “baseline” income distributions. Understanding the differences between the empirical and independent matrices therefore begins with the underlying distributions from which the matrices are constructed. Specifically, we analyze what the differences between the (actual) distributions underlying the empirical matrix and the “baseline” distribution underlying the independent matrix imply for the empirical–independent matrix difference. Thus our decomposition method is based on an array of counterfactual experiments that ask how outcomes for children from different households would change if we varied certain characteristics and/or wage structures. The counterfactual distributions are then recast into counterfactual transition matrices that are used to attribute portions of the empirical–independent differences to specific effects and variables.

To examine the drivers of economic persistence in the USA, we apply the decomposition method to the intergenerational income mobility of white males surveyed by the 1979 National Longitudinal Survey of Youth (NLSY79). We base our wage structure on an extended Mincer equation that includes education, experience, and cognitive and non-cognitive ability measures. While the general method we propose extends to any function of transition matrices, we focus our application on two summary indices of this matrix and the top and bottom rows of the matrix, which represent the transition probabilities for children from the poorest and wealthiest households. The summary indices indicate an even 50:50 split between the composition and structure effect in explaining the observed mobility gap. They also point to years of education and cognitive ability to be the main drivers of the composition effect, accounting for 20% and 15% of the overall observed gap, respectively. However, these summary indices hide underlying heterogeneity in the effects. Across transition probabilities for children from the poorest and wealthiest households, the composition effect explains between 38% and 72% of the observed mobility gap, with the structure effect explaining the remaining gap. Further, in examining why sons from the wealthiest homes are more likely to end up in the upper quartile, education differences between households explains 25% of the gap. But looking at why they are less likely to end up in the bottom quartile, education only explains about 16% of the gap. A similar story emerges when we look at sons from the poorest households. Differences in education explain 33% of why they are less likely to end up in the upper quartile but only 15% of why they are more likely to end up in the bottom quartile. Differences in measured cognitive ability plays a significant role in explaining why they seem to be stuck in the bottom quartile. While we are unable to attach causal claims to these findings, and thus encourage the reader to interpret these results with caution, they seem to indicate that policies that aim to equalize educational outcomes may have a greater effect on the right tail of the income distribution (i.e., who ends up in the upper quartile) than on the left tail (i.e., who ends up in bottom quartile). Information on such heterogeneities, which is valuable information to assess overall policy effects, are precisely the kind of findings our method aims to uncover.

2 | METHODS

2.1 | Transition matrices and mobility indices

A transition matrix depicts the probability a child will have adult earnings (Y^c) in a specific income bracket given that his/her parents' income (Y^p) was in a certain income bracket.[3] Transition matrices have several advantages over mobility measures that focus only on the average degree of transition (e.g., intergenerational elasticities or correlation coefficients). Transition matrices provide more information about mobility across the *entire* distribution, allow for asymmetric patterns across the distribution (e.g., more mobility at the top than the bottom), and allow subgroup comparisons across the entire distribution (Black & Devereux, 2011; Jäntti et al., 2006).[4]

More specifically, let there be m income brackets (defined as equal percentile groups) with boundaries $0 < \zeta_1 < \zeta_2 < \cdots < \zeta_{m-1} < \infty$ for the parental distribution and $0 < \xi_1 < \xi_2 < \cdots < \xi_{m-1} < \infty$ for the children's distribution. Also let π denote the unconditional probability of falling into a bracket (e.g., $\pi = 0.25$ for a quartile transition matrix—the empirical application of this paper). A transition matrix (P) is an $m \times m$ matrix with elements p_{ij} that represent the conditional probability that a child is in income bracket j given his/her parents were in income bracket i or

$$p_{ij} = \frac{\Pr(\zeta_{i-1} \leqslant Y^p < \zeta_i \text{ and } \xi_{j-1} \leqslant Y^c < \xi_j)}{\Pr(\zeta_{i-1} \leqslant Y^p < \zeta_i)},$$

where $\sum_{j=1}^{m} p_{ij} = 1$. Here we see that each p_{ij}, and thus the matrix P, is dependent on the distributions of children's incomes conditional on parental income group—specifically its relation to the ξ_j boundary cut-off points.

The top panel in Table 1 (“Empirical matrix") provides a quartile transition matrix derived for white males from the NLSY79 data (Section 3 provides a full discussion of the data used). Entries in Table 1 (top panel) indicate the probability a son will be in each quartile of the income distribution conditional on parents having been in a specific quartile of the income distribution. The top left entry (41.5) tells us that for parents in the bottom quartile of the income distribution their son has about a 42% chance of ending up in the bottom quartile of his generation's income distribution. Similarly, the top right entry (12.4) tells us that the same son has only about a 12% chance of ending up in the top quartile of the income distribution. Comparatively, a son whose parents were in the top quartile of the income distribution has a 13% and 42% chance of ending up in the bottom and top quartile, respectively, of their generation's income distribution.

The intergenerational mobility portrayed in a transition matrix can be summarized through mobility indices, $M(P)$, which map the transition matrix P into a scalar value (Formby et al. 2004). These mappings can be useful in ranking transition matrices according to various welfare criteria (e.g., Dardanoni, 1993; Kanbur & Stiglitz, 2015; Maasoumi, 1998). A number of mobility indices have been proposed, each reflecting a different way to measure mobility.[5] Therefore, researchers have a degree of discretion in how they summarize the transition matrix. While our decomposition approach extends to any of these proposed summary indices, we consider two indices shown by Dardanoni (1993) to be “coherent” in that they rank order transition matrices in accordance with a class of social welfare functions.[6]

The first index is a measure proposed by Bartholomew (1982) and can be interpreted as the expected number of boundaries a child might cross:

$$M_1 = \sum_i \sum_j \pi_i p_{ij} |i - j|. \tag{1}$$

The second index is related to the second largest eigenvalue (λ_2) of the transition matrix, which is considered a measure of the rate at which a Markov chain converges to its steady state (i.e., the larger the λ_2, the slower the convergence). Because the second largest eigenvalue is also related to the correlation of status between generations or the speed of escape from initial conditions, it is commonly proposed as a measure of social immobility (Maasoumi, 1998; Sommers & Conlisk, 1979; Theil, 1972). We focus on a similar measure based on the second largest eigenvalue ($\hat{\lambda}_2$) of a “symmetrized” version of the transition matrix P (Sommers & Conlisk, 1979):

$$M_2 = 1 - |\hat{\lambda}_2|, \tag{2}$$

[3]There is a distinction between “size” transition matrices and “quantile” transition matrices. The former defines boundaries of the matrix exogenously—for example, every \$10,000—while the latter defines them endogenously—for example, every 25th percentile. We focus solely on quantile transition matrices. Interested readers can refer to Formby, Smith, and Zheng (2004) and references therein for an in-depth discussion of size and quantile matrices.

[4]Transition matrices, however, are not without shortcomings. For example, a main criticism, voiced by Bhattacharya and Mazumder (2011), is the reliance on arbitrary cut-offs in the distribution.

[5]See Maasoumi (1998) or Checchi, Ichino, and Rustichini (1999) for relatively comprehensive overviews of summary mobility measures and Formby et al. (2004) for a discussion of their asymptotic properties.

[6]We thank an anonymous referee for referring us to the Dardanoni (1993) article.

TABLE 1 Transition matrices and implied effects for white males in NLSY79

Matrix	Parental quartile	Childs' quartile			
		1st	2nd	3rd	4th
Empirical matrix	1st	41.47 (2.35)	27.15 (2.17)	19.03 (1.98)	12.36 (1.73)
	2nd	28.45 (2.04)	24.34 (1.96)	26.54 (1.87)	20.67 (1.95)
	3rd	17.07 (1.95)	28.25 (2.05)	29.30 (2.11)	25.37 (2.15)
	4th	13.02 (1.85)	20.26 (2.03)	25.13 (1.89)	41.60 (2.17)
Independent matrix	1st	25.00	25.00	25.00	25.00
	2nd	25.00	25.00	25.00	25.00
	3rd	25.00	25.00	25.00	25.00
	4th	25.00	25.00	25.00	25.00
Counterfactual matrix	1st	32.70 (2.48)	25.68 (2.68)	20.14 (3.11)	21.48 (3.06)
	2nd	27.20 (2.17)	24.76 (2.42)	28.05 (2.92)	19.99 (2.43)
	3rd	21.10 (2.013)	26.79 (2.44)	28.86 (2.13)	23.25 (2.25)
	4th	19.00 (1.77)	22.78 (2.30)	22.95 (2.19)	35.27 (2.24)
Implied composition effect	1st	8.76*** (2.64)	1.48 (2.74)	−1.11 (2.64)	−9.13*** (2.54)
	2nd	1.25 (2.21)	−0.42 (2.39)	−1.51 (2.93)	0.68 (2.25)
	3rd	−4.03* (2.09)	1.46 (2.29)	0.44 (2.34)	2.12 (1.84)
	4th	−5.99*** (1.76)	−2.52 (2.08)	2.18 (2.07)	6.33*** (1.88)
Implied structure effect	1st	7.70*** (2.48)	0.68 (2.68)	−4.86 (3.11)	−3.52 (3.06)
	2nd	2.20 (2.17)	−0.24 (2.42)	3.05 (2.92)	−5.00** (2.43)
	3rd	−3.90* (2.01)	1.79 (2.44)	3.86* (2.13)	−1.75 (2.25)
	4th	−5.99*** (1.78)	−2.22 (2.30)	−2.05 (2.19)	10.27*** (2.24)

Note. Standard errors, based on 200 bootstraps, are in parentheses. Incomes are adjusted by state-level cost-of-living indices. All results are multiplied by 100 for readability. For effects, asterisks denote statistical significance: *** 1% level; ** 5% level; * 10% level.

TABLE 2 Mobility indices

Index	*M* (NLSY 1979)	*M* (independent matrix)	*M* (counterfactual)
M_1	0.98 (0.027)	1.25	1.11 (0.029)
M_2	0.69 (0.027)	1.00	0.85 (0.029)

Note. These are summary index values defined in Section 2.1 for the empirical matrix based on white males from the 1979 NLSY (panel 1 in Table 1), index values for the independent matrix (panel 2 in Table 1), and for the counterfactual matrix (panel 3 in Table 1). Standard errors in parentheses.

where the symmetrized matrix $\hat{P}$ is defined as

$$\hat{P} = (1/2)[P + \Pi P'\Pi^{-1}]$$

and Π is the diagonal matrix with the steady-state vector along the diagonal (e.g., $\pi = 0.25$ across the diagonals in our 4×4 matrix setting). This measure (M_2) can be thought of as a short-run correlation measure, whereas a measure based on the actual matrix is more of a long-run correlation measure. Thus, for our intergenerational application, a measure based on the symmetrized version may be more appropriate (Sommers & Conlisk, 1979). Moreover, Dardanoni (1993) shows that the symmetrized index (M_2) is coherent in relation to a class of social welfare functions, whereas the unsymmetrized index is not. Table 2, column 1 and 2, compares the M_1 and M_2 values for the NLSY79 mobility matrix for white males to the independent mobility matrix.

Most of the existing literature on intergenerational mobility use transition matrices or mobility indices as the final output—that is, a simple way to summarize empirical estimates of economic mobility (e.g., Checchi et al., 1999; Corak & Heisz, 1999; Dearden, Machin, & Reed, 1997; Jäntti et al., 2006; Peters, 1992). What we wish to do is take this a step further and understand the potential driving forces behind the transition matrix and summary indices; specifically, what is driving the immobility observed in our empirical matrix.

To evaluate such forces, we define an *independent* matrix to which we compare our actual (empirically estimated) matrix. Thus our comparison matrix is one where there is no link between parental income and children's income (a 4×4 matrix with all entries of 0.25 in our quartile matrix setting). This matrix can roughly be thought of as a "perfect mobility" matrix (Dardononi, 1993; Jäntti et al., 2006) and we seek to understand why we diverge from this hypothetical world. Comparisons of this nature parallel the common question in the intergenerational mobility literature of "what explains a parent–child income correlation value of Z," which essentially asks why the correlation value is different from zero. Therefore, our use of the independent matrix as our point of comparison is analogous to the comparison of the correlation value to zero.[7] For example, consider the probability that a son ends up in the top quartile given his parents were also in the top quartile (bottom right entry: p_{44}). The difference between the empirical matrix and the independent transition matrix is 17% (42 − 25); that is, given parents from the top quartile, males in our sample are 17% more likely to end up in the top quartile than if mobility were reflected by the independent mobility matrix. What explains this difference? Alternatively, we can compare summary indices of these matrices. For example, what explains the −0.27 difference between the M_1 index value of the NLSY79 mobility matrix for white males and the M_1 value of the independent matrix (0.98 − 1.25)?

We seek to explain how much of these differences are due to differences in returns to characteristics between children from different households (referred to as the "structure" effect), how much is due to differences in observed characteristics between these children (referred to as the "composition" effect) and how much of the composition effect is due to specific characteristics. The indices presented above and the underlying transition matrix are functions of conditional distributions of incomes (i.e., distributions of children's incomes conditional on parental income). In order to address our questions of interest, we must conduct counterfactual experiments on the conditional distributions and then inquire what the counterfactual conditions imply for the matrix and indices. In this vein, we connect the literature on economic mobility with the literature on decomposition methods.

2.2 | Aggregate decomposition

In line with our application (quartile transition matrices), we begin with four non-overlapping groups $g \in \{1, 2, 3, 4\}$: groups of children from households in each quartile of the parental income distribution.[8] For children in each group $g \in \{1, 2, 3, 4\}$, we observe an outcome Y^g (child's adult log income) and a d-dimensional vector of observables X^g with distributions F_Y^g and F_X^g, respectively, along with a conditional cumulative distribution function (CDF) $F_{Y|X}^g$, which is implicitly defined by a wage structure $Y^g = w_g(X, \epsilon)$.

Let $\nu(F_Y^1, F_Y^2, F_Y^3, F_Y^4)$ be a function of our empirical transition matrix based on the observed distributions F_Y^1, F_Y^2, F_Y^3, and F_Y^4; for example, a summary index (M_i) or a specific entry in the transition matrix (p_{ij}). Analogously, define the same function based on our independent matrix as $\nu(F_Y^B, F_Y^B, F_Y^B, F_Y^B)$; the independent matrix is defined as the case where all groups have the same income distribution as some baseline group: F_Y^B (a topic we pick up in section 2.4). Now, define the overall "mobility gap" as follows:

$$\Delta_O^\nu = \nu(F_Y^1, F_Y^2, F_Y^3, F_Y^4) - \nu(F_Y^B, F_Y^B, F_Y^B, F_Y^B).$$

This is the gap between what we actually observe and a world in which children's incomes are independent of their parents'. To identify how much of this mobility gap is due to structure or composition effects, we rely on the use of counterfactual distributions. Specifically, for groups $g \neq g'$, define the following counterfactual distribution:

$$F_Y^{g|g'}(y) = \int F_{Y|X}^g(y, x)\mathrm{d}F_X^{g'}(x). \tag{3}$$

This is a counterfactual distribution of incomes based on individuals with characteristics like those from group g' and a wage structure (i.e., returns to those characteristics) like those from group g. The composition effect, or the portion of the overall mobility gap explained by differences in observed characteristics, is defined as

$$\Delta_X^\nu = \nu(F_Y^1, F_Y^2, F_Y^3, F_Y^4) - \nu(F_Y^{1|B}, F_Y^{2|B}, F_Y^{3|B}, F_Y^{4|B}). \tag{4}$$

[7]The proposed method can also be extended to compare two empirical transition matrices, for example that of the USA and the UK or even of a single country across two points in time. Further, we could compare our empirical matrix to a "perfect *im*mobility" matrix. Thus, rather than attempt to understand a mobility "gap," one could investigate what underlies the mobility "observed." Either of these alternatives requires a simple choice of the appropriate comparison transition matrix.

[8]The choice to partition into four groups reflects our dataset. With larger sample sizes, one might consider quintile or decile transition matrices.

The structure effect, or the portion explained by differences in returns to characteristics, is defined as[9]

$$\Delta_S^{\nu} = \nu(F_Y^{1|B}, F_Y^{2|B}, F_Y^{3|B}, F_Y^{4|B}) - \nu(F_Y^B, F_Y^B, F_Y^B, F_Y^B). \tag{5}$$

Together, these two components represent an "aggregate" decomposition of the overall mobility gap: $\Delta_O^{\nu} = \Delta_X^{\nu} + \Delta_S^{\nu}$.

Decomposing the overall mobility gap into separate structure and composition effects will provide interesting insight into sources of immobility. However, we are also interested in identifying which characteristics (or covariates) drive the composition effect; that is, what portion of the composition effect can be assigned to specific variables of interest? To investigate this question, we employ a "detailed" decomposition of the composition effect in Equation 4.[10] Several decompositions have been proposed (see Chernozhukov, Fernandez-Val, & Melly, 2013; DiNardo, Fortin, & Lemieux, 1996; Machado & Mata, 2005) but these are, in general, "path dependent" (i.e., they depend on the order of the covariate inclusion in the decomposition).[11] We apply a non-path-dependent decomposition based on copula theory introduced by Rothe (2015). In what follows, we briefly outline our specific use of his procedure and intuition; however, interested readers should refer to Rothe (2015) for a complete description of the procedure in a more traditional decomposition setting.

2.3 | Detailed decomposition

Rothe (2015) shows that the detailed decomposition cannot be additively decomposed into the marginal components of variables alone; the decomposition must contain interaction terms and a dependence structure (the intuition to each is explained in detail below). To identify these components and to see how our decomposition proceeds, we refer to the concept of the copula, which was first introduced by Sklar (1959).[12] Sklar's theorem states that any d-dimensional distribution F_X^g can be decomposed into two parts: the d marginal distributions $F_{X_i}^g$ for the random variables $(X_1, X_2, \ldots, X_d)$ and the copula function C^g, which is a multivariate CDF that captures the dependence structure of the distribution:

$$F_X^g(x) \equiv C^g(F_{X_1}^g(x_1), \ldots, F_{X_d}^g(x_d)) \text{ for } g \in \{1, 2, 3, 4\}.$$

In other words, the dependence structure of the covariates in F_X^g—captured by the copula function—can be separated from the individual marginal distributions: the $F_{X_i}^g$'s. Also note that the copula function takes as its arguments $F_{X_i}^g(x_i)$'s, which are uniform random variables, rather than the x's themselves.[13]

Now consider a d-dimensional product set $\{1, 2, 3, 4, B\}^d$, where an element of the product set (denoted with bold font) represents a set of d covariate marginal distributions. Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ represent a generic element of this set. For example, if $\boldsymbol{\omega} = (B, B, 1, 1, \ldots, 1)$, this would denote the set of covariate marginal distributions where the distributions for covariates one and two are equal to that of group B, and the remaining covariate distributions are equal to that of group 1. Using this notation, we extend the counterfactual setup from Equation 4. Let the counterfactual distribution of Y, where the wage structure is of group g, the copula function is of group g', and the marginal distribution of the lth covariate is equal to the group denoted in $\boldsymbol{\omega}_l$, be

$$F_Y^{g|g',\boldsymbol{\omega}}(y) = \int F_{Y|X}^g(y, x) \mathrm{d}F_X^{g',\boldsymbol{\omega}}(x), \tag{6}$$

[9]The sequence of the decomposition (i.e., whether to introduce the composition or structure effect first) is a choice faced by all decomposition analyzes. We choose this sequence because it prices characteristics in the composition effect at prices individuals actually face. Further, since we only perform a detailed decomposition on the composition effect, we believe this sequence is more appropriate for our method. This sequence choice differs from the decomposition literature focused on discrimination which typically prices characteristics according to the base group—the 'non-discriminatory' returns. However, as will be discussed further in the results section, we do not interpret the structural effect in our setting as a "discrimination" effect.

[10]Performing a detailed decomposition on the structure effect has conceptual problems regarding the choice of omitted/baseline group for covariates of interest even for simple linear models (see Fortin, Lemieux, & Firpo, 2011, for a discussion), and therefore we do not pursue such a decomposition.

[11]This issue of path dependence extends to other decomposition settings; in particular, there is a large body of literature that seeks to decompose inequality measures into portions explained by between-group effects (i.e., inequality across groups) and within-group effects (i.e., inequality within groups). Shorrocks (2013) proposed a solution to this issue, which he termed the "Shapley decomposition." This approach amounts to averaging components of the decomposition over all possible decomposition paths. A similar solution has also been used in the decomposition of the structure effect in Oaxaca–Blinder decomposition settings (see Fortin et al., 2011), for a discussion). While this type of approach is feasible in the current setting (specifically in the decomposition of the total marginal effect discussed below), we believe the approach taken here more closely aligns with assessing possible policy changes - for example, we can learn the limits of the effect of improving education outcomes in lower income houses.

[12]See Rothe (2015) for a specific discussion or Nelsen (2006) for a general handling of copula theory.

[13]In simplest terms, a copula is a multivariate distribution that takes standard uniform random variables as its arguments. Converting the observed covariates into uniform random variables by mapping them through their marginal distributions allows us to decompose a multivariate distribution into individual marginals and the dependence structure. The main benefit of this process is that it allows the multivariate distribution to be modeled via two easier-to-handle operations (individual marginals and a dependence structure). For our setting, it allows us to model the two components separately, simulate counterfactual distributions that combine these aspects as we see fit, and thus simulate various policy outcomes more flexibly.

with

$$F_X^{g',\boldsymbol{\omega}}(x) \equiv C^{g'}(F_{X_1}^{\omega_1}(x_1), \dots, F_{X_d}^{\omega_d}(x_d)).$$

Now the composition effect can be decomposed into a *total marginal effect* (Δ_M^v), which is due to differences in the marginal distributions between groups, and a *dependence effect* (Δ_D^v), which is due to differences between groups in their copulas:

$$\Delta_X^v = \Delta_M^v + \Delta_D^v.$$

Specifically, if we define $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}$ and $\mathbf{B}$ to be specific elements of Ω such that $\mathbf{1} = (1, 1, \dots, 1)$, $\mathbf{2} = (2, 2, \dots, 2)$, …, and $\mathbf{B} = (B, B, \dots, B)$—that is, all marginal distributions are from the same group—then the total marginal effect and the dependence effect are defined as[14]

$$\Delta_M^v = v(F_Y^1, F_Y^2, F_Y^3, F_Y^4) - v(F_Y^{1|1,\mathbf{B}}, F_Y^{2|2,\mathbf{B}}, F_Y^{3|3,\mathbf{B}}, F_Y^{4|4,\mathbf{B}});$$
$$\Delta_D^v = v(F_Y^{1|1,\mathbf{B}}, F_Y^{2|2,\mathbf{B}}, F_Y^{3|3,\mathbf{B}}, F_Y^{4|4,\mathbf{B}}) - v(F_Y^{1|B,\mathbf{B}}, F_Y^{2|B,\mathbf{B}}, F_Y^{3|B,\mathbf{B}}, F_Y^{4|B,\mathbf{B}}).$$

Next, we want to further decompose the total marginal effect (Δ_M^v) into portions attributable to specific covariates (and their interactions).[15] With d covariates and four groups, there are several potential counterfactual distributions for each Y^g that can be derived from Equation 6, not all of which are of direct interest for our decomposition. Recall that the aggregate decomposition, and thus the composition effect, was derived with group B as the baseline group to which each Y^g was compared. Therefore, to decompose the total marginal effect, we need to consider all potential counterfactuals for groups 1, 2, 3, and 4 relative to group B. That is, we need to evaluate each counterfactual where all four groups take group B's marginal distributions for some set of covariates but maintain their own marginal distributions for the remaining covariates. This requires additional notation. Consider a d-dimensional product set $\{0, 1\}^d$ where an element of the product set (denoted with a bold font and a tilde) represents a mapping for the set of the four elements $\boldsymbol{\omega}^i$ with $i = \{1, 2, 3, 4\}$. Let $\tilde{\boldsymbol{\omega}}$ be a generic element of this set and define this mapping such that $\boldsymbol{\omega}_l^i = B$ if $\tilde{\boldsymbol{\omega}}_l = 1$ and $\boldsymbol{\omega}_l^i = i$ if $\tilde{\boldsymbol{\omega}}_l = 0$. For example, if $\tilde{\boldsymbol{\omega}} = (1, 1, 0, 0, \dots, 0)$, then $\boldsymbol{\omega}^1 = (B, B, 1, 1, \dots, 1)$, $\boldsymbol{\omega}^2 = (B, B, 2, 2, \dots, 2)$, etc. In addition, let $\tilde{\mathbf{e}}^l$ be the lth unit vector such that all entries equal 0 except the lth, which equals 1. Then, given a feature v of interest, we can denote

$$\beta^v(\tilde{\boldsymbol{\omega}}) = v(F_Y^1, F_Y^2, F_Y^3, F_Y^4) - v\left(F_Y^{1|1,\boldsymbol{\omega}^1}, F_Y^{2|2,\boldsymbol{\omega}^2}, F_Y^{3|3,\boldsymbol{\omega}^3}, F_Y^{4|4,\boldsymbol{\omega}^4}\right).$$

This should be interpreted as the effect of a counterfactual alteration where, for elements of $\tilde{\boldsymbol{\omega}}$ with $\tilde{\boldsymbol{\omega}}_l = 1$, the mapped marginal distributions of groups 1, 2, 3, and 4 are altered to the marginal distributions of group B while holding all else constant (including the copula). For example, if $\tilde{\boldsymbol{\omega}} = (1, 1, 0, 0, \dots, 0)$, then $\beta^v(\tilde{\boldsymbol{\omega}})$ represents the effect on our transition matrix of changing the marginal distributions of the first two variables for all groups to those of group B while holding all else constant.

Now, letting $|\tilde{\boldsymbol{\omega}}| = \sum_{l=1}^d \tilde{\boldsymbol{\omega}}_l$, define

$$\Delta_M^v(\tilde{\boldsymbol{\omega}}) = \beta^v(\tilde{\boldsymbol{\omega}}) + \sum_{1 \leqslant |\tilde{\boldsymbol{\omega}}'| \leqslant |\tilde{\boldsymbol{\omega}}|-1} (-1)^{|\tilde{\boldsymbol{\omega}}|-|\tilde{\boldsymbol{\omega}}'|} \beta^v(\tilde{\boldsymbol{\omega}}').$$

Thus $\Delta_M^v(\tilde{\mathbf{e}}^l) = \beta^v(\tilde{\mathbf{e}}^l)$. Also, note that if $|\tilde{\boldsymbol{\omega}}| = 2$ then $\Delta_M^v(\tilde{\boldsymbol{\omega}})$ is equal to the effect $\beta^v(\tilde{\boldsymbol{\omega}})$ with the individual "parts" of the effect from $\beta^v(\tilde{\boldsymbol{\omega}}')$ where $|\tilde{\boldsymbol{\omega}}'| = 1$ removed. Thus the "full" counterfactual of changing, for instance, the distribution of the first two covariates to that of group B for all groups 1, 2, 3, and 4 is split into two "direct effects" [$\beta^v(\tilde{\mathbf{e}}^l) = \Delta_M^v(\tilde{\mathbf{e}}^l)$, where $|\tilde{\mathbf{e}}^l| = 1$ and $l = 1, 2$] and one "interaction effect" [$\Delta_M^v(\tilde{\boldsymbol{\omega}})$, where $|\tilde{\boldsymbol{\omega}}| = 2$, $\tilde{\boldsymbol{\omega}} = (1, 1, 0, \dots, 0)$].

Therefore, the detailed decomposition (of the composition effect) contains three parts: (1) "direct contributions" from each covariate due to differences between groups in the marginal distribution of these variables [$\Delta_M^v(\tilde{\mathbf{e}}^l)$]; (2) several "interaction effects," which are due to interactions between the marginals ($\Delta_M^v(\tilde{\boldsymbol{\omega}})$ where $|\tilde{\boldsymbol{\omega}}| > 1$); and (3) a "dependence effect" due to different dependence structures among the covariates between the groups (Δ_D^v). Thus we have

$$\Delta_X^v = \sum_{l=1}^d \Delta_M^v(\tilde{\mathbf{e}}^l) + \sum_{|\tilde{\boldsymbol{\omega}}|>1} \Delta_M^v(\tilde{\boldsymbol{\omega}}) + \Delta_D^v, \tag{7}$$

where the interaction effects are summed over all possible interactions.

[14]Note that the term $v(F_Y^{1|B,\mathbf{B}}, F_Y^{2|B,\mathbf{B}}, F_Y^{3|B,\mathbf{B}}, F_Y^{4|B,\mathbf{B}})$ is identical to the counterfactual term in the aggregate decomposition, that is, $v(F_Y^{1|B}, F_Y^{2|B}, F_Y^{3|B}, F_Y^{4|B})$.

[15]Rather than taking this approach, one could choose to deviate at this point and take an approach more in line with that of Shorrocks (2013) and simply average over all possible paths of equating the marginal to obtain only "direct" marginal effects in the spirit of the Shapely value (see footnote 11).

The direct effects are perhaps the simplest to understand: They are the portion of the overall gap that would be closed if we equalized the distribution of some specific variable across children from different households, for example, eliminated education differences along parental incomes. The interaction terms come about because we are dealing with nonlinear functions: The effect of changing two sets of characteristics will invariably have a different effect from the simple sum of the two direct effects; the interaction effects account for this discrepancy. The intuition behind the dependence structure is a bit more nuanced. Not only do children from different households have different distributions of individual characteristics, but the characteristics potentially covary together differently. Imagine a simple scenario with two groups (1 and 2) and two characteristics (A and B) both distributed uniform $[-1, 1]$ for both groups. Furthermore, assume both groups have the same structural model defining an outcome of interest: $Y = A + B$. Both covariate distributions are identical, as is the structure effect. But what if for group 1 the relationship between A and B (the dependence structure) were $A = B$, while for group 2 it were $A = -B$? Despite being identical in all other regards, the two groups have drastically different dependence structures and yield vastly different outcomes for the distribution of Y. The dependence effect captures this aspect of the mobility gap.

2.4 | Choice of counterfactual

The counterfactuals considered above were defined by a baseline group that must be explicitly chosen. That is, we must choose an appropriate F_X^B to generate the various counterfactual distributions. The choice of F_X^B has no effect on the independent matrix values; for our application, the independent matrix is a 4×4 matrix with all entries 0.25 regardless of the baseline distribution chosen. However, the choice of F_X^B will affect the counterfactual transition matrices that determine our decomposition components. The baseline group defines the distribution of covariates that will be assigned to all children (regardless of parental quartile) in the counterfactual distributions. Therefore, the resulting counterfactual transition matrices will represent a world in which all children have the covariate distribution of the baseline group for the covariates of interest in that particular counterfactual but maintain their own distributional aspects for the remaining covariates, as well as their own returns to these characteristics. In the traditional decomposition setting, the counterfactual is simply the distribution of characteristics of one of the groups in question (e.g., distribution of men's characteristics in a comparison of men's and women's incomes), and the interpretation of results varies based on the selected baseline group. We follow a similar practice here. With multiple choices for the baseline group, we base our choice on what we believe to be relevant policy possibilities, particularly a policy of increased investment. The aim of such policies may be to increase investment in children from lower-income households to align their characteristics with children from the highest-income households. The baseline distribution required to analyze such a policy would therefore be the distribution of children from the top quartile households. This counterfactual is essentially asking: "What if all children had characteristics like those at the top?" We follow this approach in our application and therefore set $F_X^B = F_X^4$. While we believe this is a valid counterfactual with potential policy implications and relevance, one could easily explore alternative baseline distributions. For example, one could use the unconditional distribution of all children as the baseline; this counterfactual would ask: "What if all children had the average characteristics of the population?" and be more more reflective of a redistribution policy.

2.5 | Estimation

The decomposition effects are estimated by simulating the relevant counterfactual distributions. Thus, as a first step, we must estimate each component used in the construction of these counterfactuals. These include the marginal distributions of the covariates for each group, the conditional CDF of $Y|X$ for each group, and the copula function for each group. Once these components are estimated, any counterfactual outcome can be numerically approximated using the estimated components for population components (Equation 6).

Marginal distributions are simply estimated with the empirical CDF. Most statistical software packages have built-in functions that can easily perform this; for example, the "ecdf" function in R estimates the CDF, and the "quantile" function estimates the 'inverse' CDF. This amounts to estimating $\Pr(x \leqslant c)$ with the empirical counterpart; that is, $\frac{1}{n}\sum_{i=1}^{n} I(x_i \leqslant c)$, where $I(\cdot)$ is the indicator function.

The conditional CDFs are estimated using the distributional regression approach suggested by Foresi and Peracchi (1995); this is based on estimating multiple standard probit models where we vary the cut-off across a grid along the outcome space.[16] Specifically, we repeatedly estimate $\Pr(y_i \leqslant \tilde{y}|x) = \Phi(x\beta_{\tilde{y}})$ for $\tilde{y} \in \mathcal{Y}$, where X is a vector of covariates (e.g., education, experience,

[16] Koenker and Bassett (1978) propose an alternative approach based on quantile regression. We refer readers interested in the relationship between the alternative approaches to Koenker, Leorato, and Peracchi (2013).

cognitive, and non-cognitive measures) and Y is log income. Thus for any vector $X = x$ we have an approximation to the conditional CDF represented by a grid of estimated probabilities (i.e., $Pr(y_i \leqslant \tilde{y}|x)$ for select $\tilde{y}$). The approximation of the conditional CDF is completed by interpolating these estimated grid points to create an approximately continuous conditional CDF.

Lastly, the copula is estimated with the Gaussian copula model:

$$C_{\Sigma}(u) = \Phi_{\Sigma}^{d}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)). \tag{8}$$

This model does not impose the joint distribution of X to be Gaussian; rather, it assumes the dependence structure of the transformed variables $u_i = F_i(x_i)$ follows the Gaussian model.[17] For estimation, we rely on the maximum pseudo-likelihood approach (Genest, Ghoudi, & Rivest, 1995).[18] Specifically, this step estimates the covariance matrix of the multivariate Gaussian distribution underlying our copula. The resulting estimate is the one that maximizes the likelihood of the "pseudo-data"—data that have been converted to "rankings" through the empirical CDF of each marginal.[19]

Our steps to simulate a counterfactual transition matrix are as follows:

1. Estimate the copulas, conditional CDFs and marginal distributions as described above.
2. Simulate N pseudo-observations from a copula of choice, which yields a multivariate distribution of covariate "rankings." For our application, we draw $N = 50,000$ observations from a Gaussian multivariate distribution (given our choice of copula and estimate of the covariance structure) and then pass these draws through a standard Gaussian CDF to yield pseudo-observations or "rankings."
3. Pass the pseudo-observations through the "inverse" CDF of a set of marginals of choice, providing a distribution of covariates (X's).
4. Pass the set of X's from step 3 through a conditional CDF of choice to get N conditional CDFs.
5. For each $\tilde{y}$, take the average $\Pr(y_i \leqslant \tilde{y})$ across the N estimated conditional CDFs to get an unconditional CDF.
6. Interpolate the estimated grid probabilities from step 5 to create an approximately continuous distribution.
7. Repeat steps 2–6 for each group underlying the counterfactual transition matrix (e.g., for each group of children from different parental income quartiles).
8. Cast the set of simulated distributions into a counterfactual transition matrix.

Once the counterfactual transition matrix is constructed, we can use it to assign portions of the mobility gap to specific effects (i.e., structure, composition, dependence, direct marginal, etc.).

3 | DATA

The primary source of data for our analysis is the 1979 National Longitudinal Survey of Youth (NLSY79), including the restricted-use geocode data.[20] The NLSY79 is a panel survey of youths aged 14–22 in 1979. It includes a cross-sectional representative survey ($n = 6,111$), an oversample of minorities and poor whites ($n = 5,295$), and a sample of military respondents ($n = 1,280$).[21] We use only the cross-sectional representative survey.

We limit the sample to white males who reported living with a parent for at least two of the first three years of the survey and with reported parental income for those years.[22] A key variable of interest is parental status based on both parents' (average) household income (identified as full household income minus any income reported by the respondent).[23] The outcome of interest

[17] While the Gaussian copula is fairly flexible and a common choice, it is not the only option. For a robustness check, we also estimate our decomposition using Student's t copula, which assumes the dependence structure of the transformed variables follows a multivariate t-distribution and allows for greater tail dependence. We find no substantial differences from our main findings.

[18] Specifically, we use the R package "copula" and the methods within (Hofert, Kojadinovic, Maechler, & Yan, 2015; Hofert & Maechler, 2011; Kojadinovic & Yan, 2010a; Yan, 2007). Through simulation, Kojadinovic and Yan (2010b) show that the pseudo-likelihood approach performs favorability over several other approaches.

[19] For example, assume a two-dimensional case with a given actual data point of covariates of $X = (x_1, x_2) = (35, 50)$. This would be converted to pseudo-data of perhaps $\tilde{X} = (\tilde{x}_1, \tilde{x}_2) = (0.25, 0.8)$ if 35 was the 25th percentile of the marginal distribution of x_1 and 50 was the 80th percentile of the marginal distribution of x_2. The covariance matrix estimate is maximized over this pseudo-data; in particular, these pseudo-observations are passed through an inverse standard Gaussian CDF yielding data on a multivariate Gaussian distribution on which the covariance structure is maximized (i.e., Equation 8).

[20] The use of the NLSY79 geocode data is subject to a special agreement with the Bureau of Labor Statistics (BLS). The views expressed here do not necessarily reflect the views of the BLS.

[21] The oversample of military and poor whites was discontinued in 1984 and 1990, respectively

[22] We exclude individuals who lived with a spouse or child during these years.

[23] Measurement error in parental income is a common concern in the mobility literature. Although some research indicates that transition matrices are less susceptible to such errors relative to other mobility measures (Nybom & Stuhler, 2016, the concern is still valid, particularly in the corners of the transition matrix (Nybom & Stuhler, 2016; O'Neill, Sweetman, & Van De Gaer, 2007).

TABLE 3 Summary statistics: NLSY79 white males

Variable	Parental income quartile Q1	Q2	Q3	Q4	All
Parental income	12,342	25,190	35,539	58,746	32,974
	(4,741)	(2,961)	(3,674)	(15,748)	(19,050)
Youth income	19,443	23,672	27,014	35,370	26,381
	(15,504)	(16,724)	(18,418)	(24,677)	(20,016)
Youth income rank	0.37	0.47	0.53	0.63	0.50
	(0.27)	(0.28)	(0.26)	(0.28)	(0.29)
Experience	13.4	13.9	14.0	13.9	13.8
	(4.0)	(3.5)	(3.2)	(3.3)	(3.5)
Education	12.7	13.1	13.9	14.9	13.6
	(2.5)	(2.3)	(2.4)	(2.6)	(2.6)
AFQT	0.10	0.43	0.59	0.80	0.48
	(1.02)	(0.87)	(0.86)	(0.80)	(0.93)
Rotter	8.8	8.5	8.4	8.2	8.5
	(2.4)	(2.2)	(2.3)	(2.5)	(2.4)
Esteem	21.8	22.1	22.7	23.2	22.5
	(3.9)	(3.8)	(4.0)	(4.1)	(4.0)
Perlin	21.9	22.5	22.7	23.1	22.5
	(2.9)	(3.1)	(3.0)	(3.0)	(3.0)
Birth year	61.2	61.4	61.3	61.2	61.3
	(2.03)	(2.01)	(2.09)	(2.04)	(2.05)

Note. Standard deviations in parentheses. Incomes are constant 1982–84 dollars and adjusted for state-level cost of living indices. AFQT score is a standardized measurement.

is the individual's economic status based on their average reported wage and salary income between 1996 and 1998. All incomes are deflated to 1982–1984 dollars using the CPI. Further, since regional differences in cost-of-living may lead to imperfect and potentially biased measures of real income (Fuchs, 2004), we adjust parental and child incomes according to the state-level cost of living indices discussed by Berry, Fording, and Hanson (2000).[24]

The sample is further limited to individuals not enrolled in school over the period of interest and with available Armed Forces Qualifying Test (AFQT) scores. The final sample includes 1,321 individuals born between 1957 and 1964 with a mean age of 34.7 during our outcome years of interest (1996–1998).[25] Table 3 provides summary statistics for the entire sample and by parental income quartile.

The variables we include in our decomposition are based on an extended Mincer equation. The traditional Mincer equation includes education, experience, and experience squared (Mincer, 1974). We extend this basic model to include other variables that have been related to income determination. In particular, we include a measure of cognitive ability (*AFQT*) and three measures of non-cognitive ability (*Esteem*, *Rotter*, and *Perlin*).

The NLSY79 does not provide a direct measure of experience. Therefore, we construct a measure of "full-time equivalent" (FTE) years of experience using the weekly array of hours worked.[26] One FTE year of experience is assumed to equal 52 weeks times 40 hours (hours worked are top coded to 40). A few older individuals in our sample completed their education

[24]This step requires the use of the restricted NLSY79 geocode data, which includes respondents' state of residence. We link respondents' state of residence to year/state-level costs of living. Specifically, we use the index provided on Richard Fording's website (https://rcfording.wordpress.com/datasets/) and thank him for providing these publicly.

[25]The literature on intergenerational mobility has identified the possibility of life-cycle bias depending on age at which incomes are measured (Böhlmark & Lindquist, 2006; Haider & Solon, 2006; Nybom & Stuhler, 2016). However, this literature seems to indicate that such biases, at least for left-hand-side measurement error, are minimized or eliminated when youths reach their mid-30s, and this informs our choice of survey years in a trade-off with higher attrition later in the survey. Furthermore, Nybom and Stuhler (2016) find that transition matrices are much more immune to these biases than alternative measures of mobility. However, for a robustness check, we reestimate our analysis by rolling forward our sample window. We remeasure incomes between 1998 and 2000, which gives us an average age of over 37 years. We find minimal differences and no changes in our qualitative results. Regarding right-hand-side bias due to the age of parents' incomes, we have less control over this due to the nature of the survey. However, our average parental age is 46 years. In a "both-hand-side" measurement error simulation by Nybom and Stuhler (2016), they fix parental age at 45 and find this adds little bias to "bottom to top" transitions in a transition matrix. While not definitive, we take these to suggest that life-cycle biases are unlikely to be the main drivers of our main results.

[26]Our measure of experience is very similar to, but slightly different from, the measure used by Regan and Oaxaca (2009).

prior to the beginning of the survey and therefore were already working during the first round of interviews in 1979. Without information on previous work experience for these individuals, we construct the following "pre-survey" estimate of FTE years ($FTE_{<79}$) based on age, years of schooling, and FTE years of experience earned in the initial survey year: $FTE_{<79} = (Age_{79} - \text{Years of schooling}_{79} - 6) \cdot FTE_{79}$. We then add the pre-survey FTE years to the (observed) survey FTE years.

Educational attainment is measured as years of schooling. The measure of cognitive ability used in our analysis is Armed Forces Qualifying Test scores (AFQT). Since different individuals took the test at different ages, the measure used is from an equi-percentile mapping used across age groups to create age-consistent scores (Altonji, Bharadwaj, & Lange, 2012). The use of AFQT scores as a measure of ability warrants a brief discussion. Some argue AFQT scores are proxies for IQ scores while others draw serious doubts to this interpretation Ashenfelter and Rouse (2000). Others question what it is exactly that IQ scores measure, noting large changes in IQ scores over time (Flynn, 2004). Therefore, we simply interpret AFQT scores as some combination of innate ability and accumulated human capital as a youth that is valued in the labor market. However, for ease of expression, we will refer to AFQT scores as our measure of "cognitive ability."

We use three measures for non-cognitive ability. First, we use information from the Rosenberg Self-Esteem Scale Rosenberg (1965). This scale contains 10 statements on self-approval and disapproval; we use a summary measure of the individual's responses to these 10 statements (*Esteem*). Second, we use a summary measure from the Rotter-Locus of Control Scale (*Rotter*), which measures the "extent to which individuals believe they have control over their lives through self-motivation or self-determination (internal control) as opposed to the extent that the environment (that is, chance, fate, luck) controls their lives (external control)" (Bureau of Labor Statistics, 2015). *Rotter* and *Esteem* were measured in the first two rounds of the survey (1979 and 1980, respectively). Therefore, the third measure we include is the Perlin Mastery Scale, measured in 1992 when respondents were in their early 30s. The Perlin Mastery Scale measures the extent to which individuals "perceive themselves in control of forces that significantly impact their lives" (Bureau of Labor Statistics, 2015).

A brief discussion regarding identification is warranted. That is, what can we actually learn from our empirical application? The first step of our approach is an aggregate decomposition that attempts to separate the overall mobility gap into composition and structure effects. In order to correctly identify these effects, we must assume that, conditional on the observable characteristics, any unobserved characteristics affecting wages are identically distributed for all groups (i.e., across parental quartiles).[27] If this assumption does not hold, our structure effect may incorrectly pick up differences in unobserved productivity characteristics. Given our robust set of cognitive and non-cognitive measurements, we believe this is a tenable assumption in our application. Thus we feel our aggregate decomposition identifies meaningful structure and composition effects. The second step of our approach is a detailed decomposition to break down the composition effect into effects from individual characteristics (as well as interaction and dependence effects). To attach causal interpretations to our detailed decomposition, an additional assumption must hold: any unobservable characteristics that affect wages are independent of our observed characteristics.[28] The endogenous nature of education in explaining incomes is well documented and thus we believe this is likely an untenable assumption in our setting. Therefore, like much of the mobility and decomposition literature, we interpret the detailed decomposition results as descriptive in nature and as a first-order approximation to underlying drivers of income persistence.

4 | RESULTS

We begin by investigating the decomposition results for our two summary indices M_1 and M_2 (Table 4). We then take a closer look at what is driving those results. We do this by first discussing the aggregate decomposition and counterfactual transition matrix as a whole (Table 1). Then, we present detailed decomposition results for the transition matrix entries for those coming from households in the top quartile of the earnings distribution (p_{4j} for $j = 1, 2, 3, 4$) (Table 5) and transition matrix entries for those coming from households in the bottom quartile of the earnings distribution (p_{1j} for $j = 1, 2, 3, 4$) (Table 6).[29]

Within each decomposition (Tables 4 –6), our presentation of results follows our theoretical discussion of the decomposition method. First, we report the overall difference and its aggregate decomposition into the structure and composition effects. Second, we report the initial step in the detailed decomposition of the composition effect (i.e, the dependence and total marginal

[27]This assumption is more generally referred to as "conditional independence" or "ignorability." See Fortin, Lemieux, and Firpo (2011) for a detailed discussion of identifying assumptions in decomposition methods.
[28]This assumption is more generally referred to as "independence."
[29]The overall actual–benchmark difference reported in Tables 4 –6 will differ slightly from the differences from Tables 1 and 2 because the decomposition results are based on simulated data rather than the empirical data (i.e., we simulate our actual empirical transition matrix to construct our decomposition rather than use the empirical matrix itself; see section 2.5). However, the estimated values are very similar, suggesting that our parametric specification of the conditional CDF and copula function fit the data reasonably well.

TABLE 4 Index decompositions

	M1		M2	
Total difference: Δ_o^v	−27.23***	(2.65)	−30.57***	(2.73)
Structure effect: Δ_s^v	−14.15***	(2.86)	−14.87***	(2.95)
Composition effect: Δ_x^v	−13.08***	(2.47)	−15.70***	(2.66)
Dependence effect: Δ_d^v	0.83*	(0.48)	0.96	(0.93)
Marginal effect: Δ_m^v	−13.92***	(2.45)	−16.66***	(2.60)
"Direct" contribution to marginal effect: $\Delta_m^v(e^j)$				
AFQT	−3.82***	(1.46)	−4.68***	(1.46)
Education	−5.96**	(2.36)	−6.19**	(2.47)
Experience	−1.22*	(0.68)	−1.26*	(0.69)
Esteem	−0.25	(0.56)	−0.42	(0.58)
Rotter	−0.11	(0.50)	−0.19	(0.52)
Perlin	−0.52	(0.74)	−0.60	(0.75)
"Two-way" interactions: $\Delta_m^v(k)$ with $\|k\| = 2$				
AFQT:Education	−1.37	(1.13)	−2.23*	(1.35)
AFQT:Experience	−0.17	(0.51)	0.07	(0.48)
AFQT:Esteem	−0.04	(0.35)	0.12	(0.29)
AFQT:Rotter	0.20	(0.27)	0.10	(0.26)
AFQT:Perlin	−0.60	(0.68)	−0.25	(0.52)
Esteem:Education	−0.32	(0.43)	−0.27	(0.50)
Esteem:Experience	−0.01	(0.17)	−0.03	(0.16)
Esteem:Rotter	0.06	(0.10)	0.07	(0.10)
Esteem:Perlin	0.06	(0.23)	0.02	(0.19)
Rotter:Education	0.12	(0.30)	0.12	(0.35)
Rotter:Experience	−0.02	(0.13)	−0.06	(0.14)
Rotter:Perlin	0.11	(0.18)	0.15	(0.18)
Perlin:Education	−0.79	(0.72)	−1.08	(0.80)
Perlin:Experience	0.08	(0.41)	0.02	(0.31)
Education:Experience	−0.50	(0.63)	−0.78	(0.64)

Notes. Standard errors, based on 200 bootstraps, are in parentheses. Asterisks denote statistical significance: ***1% level; **5% level; *10% level. All results are multiplied by 100 for readability.

effects). Then, we report the covariate direct contributions to the marginal effect followed by two-way interaction contributions to the marginal effect; higher-order interaction terms are not reported for economy.

In many traditional decomposition settings, the structure effect is interpreted as a discrimination effect. We believe that would be an inappropriate label for this setting; instead, a more befitting interpretation would be some form of (loosely termed) "privilege"; for example, parental connections or parental knowledge/awareness of job market opportunities. In terms of the latter example, if wealthier parents pass on greater knowledge/awareness of educational and job market opportunities to their sons, such benefits would appear to the econometrician as higher returns for productive characteristics. Similarly, since job search is costly, sons of wealthier parents may have the financial flexibility to spend more time searching for a better job match, and the structure effect may partially reflect differences in match quality.

4.1 | Index decompositions

Comparing the aggregate decomposition results for the two indices (M_1 and M_2), the results appear fairly consistent; the composition and structure effect each account for about half of the observed mobility gap. Therefore, differences in characteristics and differences in returns to those characteristics have equal weight in explaining the total observed gap in the mobility indices. Moreover, they both identify similar drivers within the detailed decompositions: the composition effect appears to be driven by the marginal effect, and AFQT and education appear to be the main drivers of the marginal effect. AFQT accounts for around 30% of the marginal effect (15% of the total gap), whereas education accounts for 40–45% of the marginal effect (20% of the total gap). Experience is also a minor contributor to the marginal effect, explaining about 10% of the marginal effect for both

TABLE 5 Decomposition of top (parental) quartile

	1st Q		2nd Q		3rd Q		4th Q	
Total difference: Δ_o^v	−11.98***	(1.72)	−4.74**	(2.01)	0.13	(2.23)	16.60***	(2.17)
Structure effect: Δ_s^v	−5.99***	(1.86)	−2.22	(2.09)	−2.05	(2.07)	10.27***	(2.18)
Composition effect: Δ_x^v	−5.99***	(1.66)	−2.52	(1.93)	2.18	(1.85)	6.33***	(1.70)
Dependence effect: Δ_d^v	−1.14***	(0.43)	1.33***	(0.45)	1.59***	(0.570)	−1.79***	(0.58)
Marginal effect: Δ_m^v	−4.85***	(1.51)	−3.85**	(1.89)	0.58	(1.87)	8.12***	(1.80)
"Direct" contribution to marginal effect: $\Delta_m^v(e^j)$								
AFQT	−1.28**	(0.54)	−0.76	(1.07)	0.44	(1.28)	1.59*	(0.95)
Education	−2.05**	(0.84)	−1.78	(1.66)	−0.39	(1.81)	4.21***	(1.55)
Experience	−0.21	(0.35)	−0.03	(0.38)	0.19	(0.38)	0.06	(0.52)
Esteem	−0.09	(0.19)	0.01	(0.32)	−0.22	(0.37)	0.29	(0.36)
Rotter	0.09	(0.14)	−0.03	(0.26)	−0.67	(0.41)	0.61*	(0.34)
Perlin	−0.62*	(0.36)	−0.27	(0.57)	0.23	(0.53)	0.65	(0.45)
"Two-way" interactions: $\Delta_m^v(k)$ *with* $\lvert k\rvert = 2$								
AFQT:Education	−0.76	(0.84)	−0.90	(1.44)	1.57	(1.64)	0.10	(1.13)
AFQT:Experience	0.08	(0.29)	0.02	(0.51)	−0.07	(0.56)	−0.04	(0.32)
AFQT:Esteem	−0.03	(0.172)	0.05	(0.348)	0.27	(0.35)	−0.29	(0.23)
AFQT:Rotter	0.00	(0.11)	−0.09	(0.25)	0.62	(0.39)	−0.54*	(0.28)
AFQT:Perlin	0.11	(0.37)	0.32	(0.79)	0.12	(0.74)	−0.55*	(0.32)
Esteem:Education	−0.18	(0.19)	−0.22	(0.36)	0.71	(0.49)	−0.31	(0.36)
Esteem:Experience	−0.00	(0.09)	0.01	(0.17)	−0.00	(0.19)	0.00	(0.11)
Esteem:Rotter	0.01	(0.05)	−0.02	(0.10)	0.15	(0.15)	−0.13	(0.15)
Esteem:Perlin	−0.02	(0.13)	0.03	(0.23)	0.14	(0.25)	−0.15	(0.13)
Rotter:Education	−0.00	(0.13)	−0.02	(0.27)	0.10	(0.41)	−0.08	(0.32)
Rotter:Experience	0.02	(0.06)	−0.05	(0.13)	0.02	(0.18)	0.01	(0.14)
Rotter:Perlin	0.03	(0.10)	−0.10	(0.19)	0.33	(0.23)	−0.25*	(0.15)
Perlin:Education	−0.37	(0.42)	0.16	(0.80)	−0.11	(0.84)	0.31	(0.52)
Perlin:Experience	0.03	(0.21)	−0.06	(0.38)	0.05	(0.34)	−0.02	(0.14)
Education:Experience	−0.14	(0.32)	−0.21	(0.57)	0.01	(0.63)	0.35	(0.41)

Note. Standard errors, based on 200 bootstraps, are in parentheses. Asterisks denote statistical significance: ***1% level; **5% level; *10% level. All results are multiplied by 100 for readability.

measures (or about 5% of the total observed gap). The AFQT–education interaction stands out for both measures but is only significant for M_2. The interaction term implies complementary effects between the two measures. Based on the estimated direct and interaction effects, about 80% of the composition effect (or about 40% of the total gap) would be closed if both AFQT and education were equalized across households. Overall, while the two summary indices measure mobility differently, they portray very similar pictures regarding the main drivers of measured immobility in the USA.

Summary measures such as M_1 and M_2 provide a useful overall assessment of underlying mobility drivers, but do not reveal potential heterogeneities in where and how certain factors play their driving role. In particular, the decompositions of M_1 and M_2 indicate that AFQT and education are the main drivers of the mobility gap. However, these results do not identify how these effects are distributed across the income distribution. A closer look at the entire empirical and counterfactual transition matrices, along with decompositions of specific transition matrix entries, can help reveal potential heterogeneities.

4.2 | Aggregate counterfactual transition matrix

The third panel in Table 1 presents the estimated counterfactual transition matrix. This is the transition matrix we would observe if there were no differences in the distribution of *characteristics* between sons from different households (i.e., if composition effects were removed). Remaining differences between the counterfactual matrix and the independent matrix represent the role played by the structure effect (i.e., differences in *returns* to characteristics). These differences are presented in the bottom two panels of Table 1.

TABLE 6 Decomposition of bottom (parental) quartile

	1st Q		2nd Q		3rd Q		4th Q	
Total difference: Δ_o^v	16.47***	(2.29)	2.15	(2.12)	−5.97***	(1.98)	−12.64***	(1.60)
Structure effect: Δ_s^v	7.70***	(2.75)	0.68	(2.82)	−4.86*	(2.93)	−3.52	(3.07)
Composition effect: Δ_x^v	8.76***	(2.61)	1.48	(3.12)	−1.11	(2.70)	−9.13***	(2.50)
Dependence effect: Δ_d^v	0.46	(0.55)	−0.74	(0.58)	−1.45*	(0.74)	1.73**	(0.74)
Marginal effect: Δ_m^v	8.30***	(2.34)	2.22	(2.97)	0.33	(2.69)	−10.85***	(2.50)
"Direct" contribution to marginal effect: $\Delta_m^v(e^j)$								
AFQT	3.78**	(1.93)	−0.04	(1.95)	−1.51	(1.64)	−2.22	(1.36)
Education	2.45	(2.30)	−0.20	(3.09)	1.82	(2.70)	−4.07*	(2.25)
Experience	1.26	(0.88)	−0.22	(0.71)	−0.33	(0.59)	−0.71*	(0.42)
Esteem	0.08	(0.67)	0.76	(0.73)	−0.76	(0.67)	−0.08	(0.52)
Rotter	−0.17	(0.54)	−0.13	(0.62)	0.81	(0.66)	−0.51	(0.48)
Perlin	0.088	(0.93)	−0.25	(1.04)	−0.27	(0.96)	0.44	(0.61)
"Two−way" interactions: $\Delta_m^v(k)$ *with* $\lvert k\rvert = 2$								
AFQT:Education	0.80	(1.63)	1.91	(2.15)	−1.19	(1.7)	−1.51*	(0.85)
AFQT:Experience	0.08	(0.62)	0.24	(0.76)	−0.23	(0.61)	−0.10	(0.26)
AFQT:Esteem	0.50	(0.42)	−0.67	(0.66)	0.10	(0.49)	0.07	(0.22)
AFQT:Rotter	−0.12	(0.35)	0.56	(0.53)	−0.69	(0.49)	0.25	(0.29)
AFQT:Perlin	0.87	(0.90)	0.11	(1.12)	−0.83	(0.90)	−0.15	(0.32)
Esteem:Education	0.46	(0.55)	−0.59	(0.79)	0.14	(0.62)	−0.00	(0.40)
Esteem:Experience	0.06	(0.22)	−0.04	(0.34)	−0.02	(0.21)	0.00	(0.08)
Esteem:Rotter	−0.03	(0.11)	0.05	(0.16)	−0.09	(0.15)	0.06	(0.08)
Esteem:Perlin	0.16	(0.29)	−0.20	(0.43)	−0.00	(0.31)	0.04	(0.10)
Rotter:Education	−0.19	(0.41)	0.54	(0.65)	−0.20	(0.62)	−0.15	(0.47)
Rotter:Experience	−0.06	(0.15)	0.15	(0.24)	−0.06	(0.19)	−0.02	(0.11)
Rotter:Perlin	−0.16	(0.23)	0.36	(0.34)	−0.28	(0.30)	0.08	(0.14)
Perlin:Education	0.44	(0.99)	0.68	(1.29)	0.11	(0.97)	−1.23**	(0.59)
Perlin:Experience	−0.01	(0.48)	−0.17	(0.63)	0.15	(0.46)	0.03	(0.12)
Education:Experience	0.21	(0.65)	−0.08	(0.94)	0.53	(0.66)	−0.66	(0.51)

Note. Standard errors, based on 200 bootstraps, are in parentheses. Asterisks denote statistical significance: ***1% level; **5% level; *10% level. All results are multiplied by 100 for readability.

Comparing the bottom three panels in Table 1, we find that removing differences in characteristics shrinks the empirical–benchmark gap for most (but not all) entries and that the roles of the structure and composition effects vary substantially across entries. For example, consider sons from households in the lowest-income quartile (top row in each panel). The composition effect explains a majority of the total mobility gap for those ending up in either the lowest quartile or the highest quartile of their own generation's income distribution (first and last columns); in particular, differences in characteristics explain over 50% of the over-representation in the lowest quartile and over 70% of the under-representation in the highest quartile. Similarly, for sons from households in the highest income quartile (bottom row in each panel), the composition effect also appears to play the largest role in the tails of the income distribution. However, for sons from middle-income households (quartiles 2 or 3), the total mobility gap is relatively small and driven primarily by the structure effect in most cases.[30] Overall, "equalizing" characteristics appear to have the largest effects in the tails of the income distribution. Therefore, to better understand the role of the composition effect within the tails of the distribution (i.e., role of specific covariates), we look at detailed decompositions for sons from households in the top quartile of the parental distribution and sons from households in the bottom quartile of the parental distribution.

[30]One exception is what explains the relative absence of sons from the third quartile households (third row) from the bottom quartile of earnings (first column); this appears to have a relatively strong composition component.

4.3 | Matrix results: Top quartile of the parental distribution

We first focus on the decomposition of sons from households in the top quartile of the parental earnings distribution (p_{4j} for $j = 1, 2, 3, 4$) reported in Table 5. In this case, the total difference represents the difference between the entries in the last row of the "Empirical Matrix" in Table 1 and in the last row of the independent matrix (0.25 for all cells; second panel in Table 1). The total difference is significant for all but the third quartile; the structure and composition effects are statistically significant for the first and fourth quartiles but not the second and third quartiles. The composition effect explains around 50% of the mobility gap for the bottom quartile but only about 38% for the top quartile. These results indicate that differences in characteristics explain less than half of the large presence of sons from the top households ending up in the top quartile of their earnings distribution, and explains half of their relative omission from the bottom quartile.

In the detailed decomposition, the dependence effect is statistically significant for all quartiles; this implies that differing interrelations among productive characteristics between sons from different households are themselves important drivers of income persistence.[31] The total marginal effect is statistically significant for the first, second, and fourth quartiles. Looking at the direct contributions to the total marginal effect, AFQT has a statistically significant effect at the first and fourth quartiles, accounting for about 10% of the overall mobility gap in both. Similarly, education plays a statistically significant and large role in the first and fourth quartiles, accounting for about 16% of the overall gap in the first quartile and 25% of the overall mobility gap in the fourth quartile. Statistical significance of the non-cognitive measures are somewhat sporadic, with the *Perlin* measure playing a role in the first quartile and the *Rotter* measure in the fourth quartile. Similarly, only a few of the two-way interactions appear to be statistically significant in driving mobility and only at the fourth quartile.

Analyzing these results we see some important heterogeneity overlooked by the summary indices. While the index results point to a 50:50 split for the composition–structure effects, the results here show that the main driver varies depending on the entry of interest. Most notably, the composition effect only accounts for 38% of the overabundance of these sons in the upper quartile of incomes. This leaves a relatively large role for the structure (privilege) effect; this implies that more than just equating characteristics is needed to close the gap at the upper income levels. For example, if the structure effect partly reflects differences in parental knowledge/awareness of job market opportunities, then educating families on career path opportunities may be a promising mechanism to help close this gap. More heterogeneity emerges in the detailed decomposition, particularly the effect of education. These results imply that if educational differences were removed, such that everyone had education levels equal to those coming from the top quartile households, then the estimated mobility gaps would shrink significantly, although not equally across the transition matrix. That is, the differences in educational outcomes across households explains quite a bit of the overabundance of wealthier sons in the upper quartile but explains considerably less about their relative absence from the bottom quartile.

4.4 | Matrix results: Bottom quartile of the parental distribution

Next, we consider sons from households in the bottom quartile of the parental earnings distribution (p_{1j} for $j = 1, 2, 3, 4$) reported in Table 6. In this case, the total difference represents the difference between the entries in the first row of the "Empirical Matrix" in Table 1 and in the first row of the independent matrix (0.25 for all cells; second panel in Table 1). The composition effect is estimated at 53% for the bottom quartile and 72% for the top quartile.[32] Therefore, differences in characteristics explain 53% of the overabundance of sons from bottom-quartile homes that end up in the bottom quartile of their generation's earnings distribution, and it explains 72% of their relative absence in the upper quartile.

In the detailed composition, the marginal effect appears to drive the composition effect in both the first and fourth quartiles. The dependence effect again plays a role, but is only statistically significant in the upper half of the distribution.[33] Looking at the direct contributions to the marginal effect, only three effects are statistically significant. The effect of AFQT at the first

[31] The dependence effect illustrates that sons are pushed out of the second and third quartiles (the center) and into the first and fourth (the tails). This is driven primarily by the education-experience correlation copula parameter. This parameter is highly negative for sons from the wealthiest homes and steadily shrinks in magnitude across lower-income households to essentially zero for sons from the poorest households (these results are not reported but are available by request). From the summary statistics in Table 3, we see this is partly due to the large amount of "free" time for sons from poorer households (i.e., the summation of experience and education is lower). The dependence effect is what would result if we s et all sons' copula parameters to those from the wealthiest homes. This would have the effect of imposing the (larger) negative correlation between education and experiences for all sons, and therefore shrink the tails of the income distribution for the sons from the bottom three quartiles of households (i.e., those with high education would then have lower experience and those with low education would then get more experience). Thus, if the bottom three groups of sons are, relatively, pulled out of the tails of the distribution, then sons from the top would be pushed in.

[32] Aggregate decomposition components are not statistically significant for the second and third quartiles although the point estimates put about 80% of the gap at the third quartile on the structure effect.

[33] These results mirror those from the wealthiest homes. See footnote 31 for an explanation of these results.

quartile is significant and quite large, accounting for 23% of the overall mobility gap for that quartile. The other statistically significant effects are in the fourth quartile, where education explains about 33% of the overall gap and experience plays a small explanatory role (6%). Further, the AFQT–education and education–Perlin two-way interaction terms play a fairly large and significant role in the fourth quartile.

The decomposition of specific portions of the transition matrix again reveals heterogeneity in effects that are masked by looking at the decomposition of the summary indices. While the composition effect explains nearly three-quarters of the relative absence of these sons from the top earning quartile, it only explains half of their overabundance in the bottom. This implies that quite a different strategy is needed to "unstick" these sons from the lowest-income levels—again, perhaps equalizing information on potential career opportunities. In the detailed decomposition, we again see important heterogeneity across the tails of the matrix. While education is a strong driver in explaining why sons from the poorest households do not end up in the upper quartile of earners, it does a substantially weaker job at explaining why they seem to be stuck in the bottom quartile (33% of the gap vs. a statistically insignificant 15%). Thus equalizing the educational outcomes across households would appear to help the highest-achieving sons from the poorest homes; it would not do nearly as much in helping the lowest-achieving sons from these households. On the other hand, equalizing AFQT across households (which could be interpreted as reducing the quality gap in primary and secondary schooling) would have a substantial effect in helping the lowest-achieving sons from these homes. Notably, this latter finding is overlooked in the summary indices which simply points to AFQT as a lesser driver than education; in reality, their importance plays out in different parts of the distribution.

5 | CONCLUSION

Social scientists have long been interested in intergenerational mobility. Over the past few decades, great gains have been made in estimating mobility, both in mean effects (e.g., intergenerational elasticity of income) and through alternative measurements that provide a more complete picture of mobility (e.g., transition matrices). As a result, interest has started to turn from estimating more accurate point estimates to developing a better understanding of the forces that drive income persistence. While much work has been done to understand the driving mechanisms of mean effects, the lack of an appropriate framework has limited similar work on transition matrices and related indices. In this paper, we use recent advances in the decomposition literature to fill this gap.

We introduced a method to decompose transition matrices and related indices. Given a benchmark matrix of interest, the method decomposes the difference between the empirical matrix and benchmark matrix into portions attributable to differing characteristics between children from different economically advantaged households (a composition effect) and differing returns to these characteristics (a structure effect). Our approach also includes a detailed decomposition of the composition effect based on copula theory. While the method we presented is not a decomposition in the traditional sense, it draws directly on the decomposition literature; the procedure entails simulating several simultaneous counterfactuals, which are then recast into counterfactual transition matrices and from which we can identify decomposition effects.

To examine drivers of economic persistence in the USA, we apply our method to the intergenerational mobility of white males surveyed by the 1979 NLSY. We base our decomposition on an extended Mincer equation that includes education, experience, cognitive, and non-cognitive measures. We apply our method to two summary indices and to specific transition matrix entries. We find that the relative importance of the structure versus composition effects varies substantially across the measures of interest; these findings highlight the importance of being able to understand the driving forces of mobility beyond mean effects. Further, we found that the summary measures mask important heterogeneity in the roles of specific characteristics. In particular, both summary indices identified education and AFQT as the main drivers of the composition effect; about 40% of the total gap observed in the mobility indices would be closed if both AFQT and education were equalized across households. However, detailed decompositions of specific matrix entries revealed significant heterogeneity in the effects across the income distribution. For example, while education is a strong driver in explaining why sons from the poorest households do not end up in the upper quartile of earners, it does a substantially weaker job at explaining why they seem to be stuck in the bottom quartile. On the other hand, AFQT has a larger effect at pulling them out of the lowest quartile than it does in helping them break into the top quartile.

ACKNOWLEDGMENTS

The views expressed in this paper are those of the authors alone and do not necessarily reflect those of the Office of the Comptroller of the Currency or the US Department of the Treasury. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS. The authors would like to thank participants at seminars at Indiana–Purdue University Fort Wayne, Iowa State University, The Office of the Comptroller of the Currency—Risk Analysis Division, The University of Oklahoma, and participants at the Missouri Valley Economic Association's 2015 annual conference. This work has also benefited from the comments of three anonymous referees. All errors are our own.

REFERENCES

Altonji, J., Bharadwaj, P., & Lange, F. (2012). Changes in the characteristics of American youth: Implications for adult outcomes. *Journal of Labor Economics*, *4*, 783–828.

Ashenfelter, O., & Rouse, C. (2000). Schooling, intelligence, and income in America, *Meritocracy and economic inequality*, (chapter 5). Princeton, NJ: Princeton University Press.

Bartholomew, D. (1982). *Stochastic Models for Social Processes* (3rd ed.). London, UK: Wiley.

Berry, W., Fording, R., & Hanson R. (2000). An annual cost of living index for the American states, 1960–1995. *Journal of Politics*, *62*(2), 550–567.

Bhattacharya, D., & Mazumder, B. (2011). A nonparametric analysis of black–white differences in intergenerational income mobility in the United States. *Quantitative Economics*, *43*(1), 139–172.

Björklund, A., Lindahl, M., & Plug, E. (2006). The origins of intergenerational associations: Lessons from Swedish adoption data. *Quarterly Journal of Economics*, *121*(3), 999–1028.

Black, S., & Devereux, P. (2011). Recent developments in intergenerational mobility, *Handbook of Labor Economics* (pp. 1487–1541), Vol. *4B*. Amsterdam, Netherlands: Elsevier.

Blanden, J., Gregg, P., & Macmillan, L. (2007). Accounting for intergenerational income persistence: Noncognitive skills, ability and education. *Economic Journal*, *117*, C43–C60.

Blau, P., & Duncan, O. (1967). *The American Occupational Structure*. New York, NY: Wiley.

Böhlmark, A., & Lindquist, M. (2006). Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden. *Journal of Labor Economics*, *24*(4), 879–896.

Bowles, S., & Gintis, H. (2002). The inheritance of inequality. *Journal of Economic Perspectives*, *16*(3), 3–30.

Bureau of Labor Statistics (BLS) (2015). NLSY79 Appendix 21: Attitudinal scales. *Bureau of Labor Statistics: National Longitudinal Survey of Youth 1979*. Retrieved June 29, 2015 from https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-appendix-21-attitudinal-scales#references

Cardak, B., Johnston, D., & Martin, V. (2013). Intergenerational income mobility: A new decomposition of investment and endowment effects. *Labour Economics*, *24*, 39–47.

Checchi, D., Ichino, A., & Rustichini, A. (1999). More equal but less mobile? Education financing and intergenerational mobility in Italy and in the US. *Journal of Public Economics*, *74*, 351–393.

Chernozhukov, V., Fernandez-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*, 2205–2268.

Corak, M., & Heisz, A. (1999). The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data. *Journal of Human Resources*, *34*, 504–533.

Dardanoni, V. (1993). Measuring social mobility. *Journal of Economic Theory*, *61*(2), 372–394.

Dearden, L., Machin, S., & Reed, H. (1997). Intergenerational mobility in Britain. *Economic Journal*, *107*(440), 47–66.

DiNardo, J., Fortin, N., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica*, *64*, 1001–1044.

Erikson, R., & Goldthorpe, J. (2002). Intergenerational inequality: A sociological perspective. *Journal of Economic Perspectives*, *16*(3), 31–44.

Flynn, J. (2004). IQ trends over time: Intelligence, race, and meritocracy, *Meritocracy and Economic Inequality*, (chapter 3). Princeton, NJ: Princeton University Press.

Foresi, S., & Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, *90*, 451–466.

Formby, J., Smith, J., & Zheng, B. (2004). Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics*, *120*, 181–205.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics, *Handbook of Labor Economics* (pp. 1–102), Vol. *4(A)*. Amsterdam, Netherlands: Elsevier.

Fuchs, V. (2004). Reflections on the socio-economic correlates of health. *Journal of Health Economics*, *4*(23), 653–661.

Genest, C., Ghoudi, K., & Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*, 543–552.

Haider, S., & Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, *96*(4), 1308–1320.

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2015). Copula: Multivariate dependence with copulas. R package version 0.999-13.

Hofert, M., & Maechler, M. (2011). Nested Archimedean copulas meet R: The nacopula package. *Journal of Statistical Software*, *39*(9), 1–20.

Jäntti, M., Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Österbacka, E., ..., & Eriksson, T. (2006). American exceptionalism in a new light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States (IZA Discussion Paper No.1938): Bonn, Germany: IZA—Institute of Labor Economics.

Kanbur, R., & Stiglitz, J. (2015). Dynastic inequality, mobility and equality of opportunity. (Center for Economic Policy Research, Discussion Paper Series No. 10542). Retrieved July 15, 2015 from http://voxeu.org/sites/default/files/file/DP10542.pdf

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koenker, R., Leorato, S., & Peracchi, F. (2013). Distributional vs.quantile regression, (Einaudi Institute for Economics and Finance, Working Paper 29/13). Retrieved form https://ideas.repec.org/p/eie/wpaper/1329.html

Kojadinovic, I., & Yan, J. (2010a). Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, *34*(9), 1–20.

Kojadinovic, I., & Yan, J. (2010b). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, *47*, 52–63.

Lefgren, L., Lindquist, M., & Sims, D. (2012). Rich dad, smart dad: Decomposing the intergenerational transmission of income. *Journal of Political Economy*, *120*, 268–303.

Liu, H., & Zeng, J. (2009). Genetic ability and intergenerational income mobility. *Journal of Population Economics*, *22*, 75–95.

Maasoumi, E. (1998). On mobility, *Handbook of Applied Economic Statistics*, (chapter 5, pp. 119–173). New York, NY: Marcel Dekker.

Machado, J., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*, 445–465.

Mayer, S., & Lopoo, L. (2008). Government spending and intergenerational mobility. *Journal of Public Economics*, *92*, 139–158.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the United States using social security income data. *Review of Economics and Statistics*, *87*(2), 235–255.

Mincer, J. (1974). *Schooling, Experience and Earnings*. New York, NY: National Bureau of Economic Research.

Nelsen, R. (2006). *An Introduction to Copulas*. New York, NY: Springer.

Nybom, M., & Stuhler, J. (2016). Biases in standard measures of intergenerational income dependence. *The Journal of Human Resources* (forth coming).

O'Neill, D., Sweetman, O., & Van De Gaer, D. (2007). The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility. *Journal of Economic Inequality*, *5*(2), 129–178.

Peters, E. (1992). Patterns of intergenerational mobility in income and earnings. *Review of Economics and Statistics*, *74*, 456–466.

Regan, T., & Oaxaca, R. (2009). Work experience as a source of specification error in earnings models: Implications for gender wage decompositions. *Journal of Population Economics*, *22*, 463–499.

Richey, J., & Rosburg, A. (2017). Changing roles of ability and education in U.S.intergenerational mobility. *Economic Inquiry*, *55*(1), 187–201.

Rosenberg, M. (1965). Society and the adolescent self-image, Princeton, NJ.

Rothe, C. (2015). Decomposing the composition effect. *Journal of Business and Economic Statistics*, *33*(3), 323–337.

Shea, J. (2000). Does parents' money matter? *Journal of Public Economics*, *111*(3), 358–368.

Shorrocks, A. (2013). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *Journal of Economic Inequality*, *11*, 99–126.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de L'Institut de Statistique de L'Universite de Paris*, *8*, 229–231.

Solon, G. (1999). Intergenerational mobility in the labor aarket, *Handbook of labor economics*, chapter 29, Vol. *3A*. Amsterdam, Netherlands: Elsevier.

Sommers, P., & Conlisk, J. (1979). Eigenvalue immobility measures for Markov chains. *Journal of Mathematical Sociology*, *6*, 253–276.

Theil, H. (1972). *Statistical Decomposition Analysis*. Amsterdam, Netherlands: North-Holland.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, *21*(4), 1–21.

How to cite this article: Richey J, Rosburg A. Decomposing economic mobility transition matrices. *J Appl Econ*. 2017;1-18. https://doi.org/10.1002/jae.2578