Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Decomposing joint distributions via reweighting functions: an application to intergenerational economic mobility

Jeremiah Richey & Alicia Rosburg

To cite this article: Jeremiah Richey & Alicia Rosburg (2019): Decomposing joint distributions via reweighting functions: an application to intergenerational economic mobility, Econometric Reviews

To link to this article: https://doi.org/10.1080/07474938.2019.1697088

Published online: 11 Dec 2019.

ECONOMETRIC REVIEWS
https://doi.org/10.1080/07474938.2019.1697088

Taylor & Francis
Taylor & Francis Group

Decomposing joint distributions via reweighting functions: an application to intergenerational economic mobility

Jeremiah Richey[a] and Alicia Rosburg[b]

[a]Credit Risk Analysis Division, Office of the Comptroller of the Currency, Washington, DC, USA; [b]Department of Economics, University of Northern Iowa, Cedar Falls, Iowa, USA

ABSTRACT
We introduce a method that extends the traditional Oaxaca-Blinder decomposition to both the full distribution of an outcome of interest and to settings where group membership varies along a continuum. We achieve this by working directly with the joint distribution of outcome and group membership and comparing it to an independent joint distribution. Like all decompositions, we assume the difference is partially due to differences in characteristics between groups (a composition effect) and partially due to differences in returns to characteristics between groups (a structure effect). We use reweighting functions to estimate a counterfactual joint distribution representing the hypothetical if characteristics did not vary according to group while returns to characteristics did. The counterfactual allows us to decompose differences between the empirical and independent distributions into composition and structure effects. We demonstrate the method by decomposing multiple measures of immobility for white men in the U.S.

KEYWORDS
Decomposition; counterfactual joint distribution; intergenerational mobility

JEL CLASSIFICATION
C14; C20; J31; J62

1. Introduction

A common problem faced by economists is one of understanding the differences in an outcome between groups, such as wages between men and women. A method often used to investigate such differences is a variation of the Oaxaca-Blinder (OB) decomposition (Blinder, 1973; Oaxaca, 1973).[1] The decomposition begins with the recognition that the two groups likely have different distributions of characteristics valued in the labor market, and they may also receive different returns to these characteristics. The method is then based on a simple hypothetical scenario. Suppose a counterfactual world in which women were to retain the returns they actually received for their characteristics but had identical distributions of characteristics as men. We would then have three sets of wages—(a) women's actual wages, (b) men's actual wages, and (c) women's counterfactual wages. By comparing (a) to (c), we can identify what portion of the total gender wage gap is due to woman having different distributions of characteristics than men, and by comparing (b) to (c), we can identify the portion due to women receiving different returns to characteristics than men. The former is commonly termed a 'composition' or explained effect. The latter portion is the 'structure' or unexplained effect and is commonly interpreted in the wage gap literature as some form of discrimination.

The traditional OB decomposition focuses on mean differences between two groups; the counterfactual is the average earnings of women if they had the average characteristics of men.

CONTACT Jeremiah Richey Jeremiah.richey@occ.treas.gov Credit Risk Analysis Division, Office of the Comptroller of the Currency, Washington, D.C., USA.

[1]See Black and Devereux (2011) for a fairly extensive review of the decomposition literature in economics.

However, recent literature has extended this focus in two main directions. One body of research has expanded the focus to the entire distribution of earnings (Chernozhukov et al., 2013; DiNardo et al., 1996; Firpo et al., 2007; Machado and Mata, 2005; Rothe, 2015). While there are several approaches to achieve this, they all aim to characterize the full counterfactual distribution of women's earnings if they had the characteristics of men. This is then compared to the actual distribution of women's and men's earnings to derive composition and structure effects comprising the gender gap. Methods in this strand of literature allow one to investigate things such as differences in the 10th or 90th quantiles of the earnings distributions between men and women.

A second body of research retains the focus on mean differences but seeks to expand the application to problems where the group distinction varies along a continuum, such as socioeconomic status (Nopo, 2008; Ulrick, 2012). These methods still focus on the expectation of the outcome of interest conditional on group membership as in the traditional OB decomposition, but they apply flexible estimation strategies to handle continuous group indexing. Methods in this second strand of literature allow one to investigate mean differences in outcomes between people from different percentiles of the continuum such as the average college grade gap between students from the 10th and 75th income percentile (Ulrick et al., 2018). The counterfactual in this extension is the average grade students from households in the 10th income percentile would obtain if they had the characteristics of students from households in the 75th income percentile.

This paper seeks to extend the traditional OB decomposition in both of these directions simultaneously—to the full distribution of the outcome of interest and to settings where group distinction varies along a continuum. In other words, we seek to retain the broad view of the first extension by keeping interest on the full distribution of the outcome of interest, but also seek the flexibility to allow the group indexing to be continuous as in the second extension. We overcome technical hurdles to doing this by approaching the problem from a different angle.

Although the technical approaches may differ, all decompositions are essentially interested in some comparison between conditional distributions—how outcomes differ between groups. This is true whether the focus of interest is differences in averages or quantiles of some outcome and regardless of whether the 'groups' are discrete or vary along a continuum. In the traditional OB approach, as with the extensions cited above, the mechanics of the decomposition also operate on conditional distributions or functions thereof; the literature that follows the methodology of DiNardo et al. (1996) work directly with separate distributions conditional on group membership, while the literature that follows the approach of Nopo (2008) and Ulrick (2012) work with flexible forms of the conditional expectation function (i.e., expected outcome conditional on group membership). Instead, the mechanics of our decomposition work directly with the joint distribution of the outcome and group membership. A joint distribution framework more easily lends itself to applications where groups vary along a continuum. The conceptual framework for our alternative approach is that asking what explains differences in outcomes between groups is equivalent to asking why the joint distribution of outcomes and group membership is not independent. Our decomposition is then concerned with estimating a counterfactual joint distribution that addresses 'what would the joint distribution look like if there were no relation between group assignment and characteristics but there remained a relation between group assignment and returns to characteristics?' This counterfactual is then compared with the observed joint distribution of outcomes and groups and an independent joint distribution of outcomes and groups.

Extracting useful information directly from joint distributions, however, can be conceptually challenging as well as data intensive. Thus, as in nearly all empirical studies, we analyze joint distributions by characterizing various conditional distribution functions. Therefore, after estimating the counterfactual joint distribution, we investigate how this counterfactual distribution differs from the empirical or independent joint distribution by comparing various bivariate regressions of outcomes on group assignment (e.g., mean, quantile, nonparametric). These comparisons allow us to decompose the empirical-independent difference for these measures into composition and structure effects. In other words, while our decomposition conceptually relies on joint distributions (empirical, independent, and counterfactual), we rely on estimates of conditional distribution functions in our empirical application in order to extract easily usable and understandable relationships.[2]

Methodologically, the approach is an extension of the reweighting method of DiNardo et al. (1996) to joint distributions. Their method asks 'what would womens' wage distribution look like if they had the characteristics of men?', and the counterfactual distribution is estimated through a weighting scheme based on the probability of being a man given your characteristics. In a parallel setting, our approach would ask 'what would the joint distribution of gender and wages look like if characteristics were independent of gender?' Similar to DiNardo et al. (1996), we estimate this counterfactual joint distribution through a weighting scheme, but instead, based on the full distribution of gender conditional on characteristics. This approach allows us to go beyond settings with binary groups, such as the gender gap, to tackle problems where groups of interest vary along a continuum without relying on arbitrary segmentation of the continuum, as can be found in some applications (Cameron and Heckman, 2001; Frenette, 2007; Richey and Rosburg, 2018).

We illustrate the ease and power of our method with an application to the intergenerational economic mobility of U.S. white males surveyed in the 1979 National Longitudinal Survey of Youth (NLSY). The objective of the application is to better understand the differences in incomes of children that come from homes with different parental income levels. We base our wage structure on an extended Mincer equation that includes education, experience, and cognitive and non-cognitive measures and consider multiple immobility measures including the standard intergenerational elasticity of income (IGE), quantile regression counterparts to the IGE, and nonlinear versions of both. While the interpretation of the estimated composition effect is fairly straightforward—children from wealthier homes likely have higher levels of characteristics valued in the labor market (e.g., education) which lead to higher wages—the interpretation of the estimated structure effect is less straightforward. In traditional decomposition settings, such as gender and racial wage studies, the structure effect is commonly interpreted as a measure of discrimination. Richey and Rosburg (2018), however, argue that a more fitting interpretation in the mobility context is a household advantage linked to higher parental incomes, such as professional connections of the parents, parental knowledge/awareness of job market and education opportunities, or perhaps greater financial flexibility to facilitate job search. All of these would result in higher returns to similar productive characteristics much in the way discrimination in gender/race wage gap studies shows up as differing returns to such characteristics.

Results from our application reveal substantial, and previously unobserved, nonlinearities in the empirical and counterfactual relationships. Linear mean and quantile IGE regressions tend to mask these nonlinearities and therefore produce somewhat misleading results. By applying our decomposition method to nonlinear models, we reveal heterogenous effects across conditional quantiles of offsprings' incomes as well as between low and high parental income levels. For example, parental income appears to act as a safety net with a larger effect on lower quantiles than on higher ones, an effect previously noted by Eide and Showalter (1999). But, this effect is confined to lower income households and is driven by the composition component. Policies aimed at the composition effect, or equalizing characteristics such as education, would therefore likely benefit the worst performing sons from the poorest households. Conversely, as parental income increases it has a larger effect on the top quantiles than on lower ones causing a widening of the right tail of the conditional distribution. But, this effect is confined to higher income

[2]More specifically, our decomposition methodology is wholly concerned with joint distributions as is reflected in the model notation. Our application, however, is concerned with understanding differences in joint distributions and relies on conditional distribution functions (e.g., quantile and mean bivariate regressions). So while the mechanics of estimating our counterfactual operate on joint distributions directly, pattern extraction from the joint distributions operate on conditional distribution functions within our application.

households and is primarily driven by the structure component. These results suggest a strong role of 'advantage' in explaining the high incomes of the best performers from the highest income households. While it is important to note that identification of these effects in our empirical application rely on a key identification assumption (ignorability) and that drivers of immobility and wage gaps may evolve over time (Albrecht et al., 2015; Maasoumi and Wang, 2019; Richey and Rosburg, 2017), the descriptive nature of our results still illustrate the importance of looking beyond simple mean or linear relationships, and thus, highlight the flexibility and usefulness of the proposed method.

2. Method

To ease exposition, we discuss the details of our method in light of our application. Let offspring's income be denoted y_c and parental income be denoted y_p.[3] We investigate the empirical joint distribution—$f(y_c, y_p)$—and seek to understand why the two variables are not independent, i.e., why do we not observe $f(y_c, y_p) \equiv f(y_c)f(y_p)$. To better understand this difference, we focus our empirical application on specific estimates of conditional distribution functions, such as the slope parameter from linear regression of y_c on y_p (i.e., the IGE), and denote such distributional measures as $\nu(\cdot)$. Thus, we investigate various Measures of Immobility (MIs)[4]:

$$\Delta^{\nu}_{MI} = \nu(f(y_c, y_p)) - \nu(f(y_c)f(y_p)). \tag{1}$$

Now, let x represent individual attributes valued in the labor market; we seek to identify what aspect of the MI is caused by children from wealthier homes having different distributions of characteristics—a composition effect—and what aspect is caused by children from wealthier homes receiving different returns to their characteristics in the labor market—a structure effect. Note that any joint distribution of y_c and y_p can be written as: $f(y_c, y_p) = \int f(y_c|y_p, x)f(x|y_p)dxf(y_p)$. Thus, the composition effect comes into play via $f(x|y_p)$—how the distribution of individual attributes vary with parental income. The structure effect comes into play via $f(y_c|y_p, x)$—how the distribution of y_c conditional on x varies with y_p, or 'how returns to x vary with y_p'. For brevity, we refer to these pieces as the 'composition component' and the 'structure component' of the joint distribution; these are the components we will manipulate to form our counterfactual and achieve our decomposition.

If there were only two groups, say children from households below and above median household income, this problem could be framed in the traditional OB setting. In that case, separate wage structures would be made explicit for children from above and below the median, i.e., $f_g(y|x) = m_g(x)$ for $g = \{1, 2\}$, rather than be implicitly defined within $f(y_c|x, y_p)$. Similarly, in such binary settings, separate distributions of characteristics would be made explicit for children from above and below the median, i.e., $f_g(x)$ for $g = \{1, 2\}$, rather than be implicitly defined within $f(y_p, x)$. However, to allow for the continuous nature of parental income, we do not revert to such explicit conditioning but rather work directly within the framework based on $f(y_c|x, y_p)$ and $f(y_p, x)$.

To aid our discussion, we introduce notation for joint distributions where subscripts on a distribution denote whether the structure and composition component (with the former listed first) come from the actual ('a') or independent ('i') distribution (i.e., $f_{struct=\{a,i\}|comp=\{a,i\}}$). Therefore, we can express the actual joint distribution of (y_c, y_p) as $f_{a|a}$:

[3]Our actual focus, like much of the mobility literature, is log earnings; we refer to 'income' for short.

[4]The reader should note that for all distributional features of interest we will investigate - various slope parameters of mean and quantile regressions - the corresponding parameter for the independent joint distribution is zero, so for our purposes $\nu(f(y_c)f(y_p)) = 0$.

$$f_{a|a}(y_c, y_p) = \int f(y_c|x, y_p) f(x|y_p) dx f(y_p). \tag{2}$$

The independent joint distribution can be expressed as $f_{i|i}$ and results from a scenario in which the distribution of x and the conditional distribution of y_c are independent of parental income:

$$f_{i|i}(y_c, y_p) = \int f(y_c|x) f(x) dx f(y_p). \tag{3}$$

Now, consider the counterfactual joint distribution $f_{a|i}$:

$$f_{a|i}(y_c, y_p) = \int f(y_c|x, y_p) f(x) dx f(y_p) \tag{4}$$

which represents the counterfactual joint distribution that answers the question 'what if the distribution of productive characteristics were independent of parental income levels, but returns to those characteristics remained dependent on parental income levels?'

With these three distributions, we can identify the composition and structure effect of the MI of interest. The composition effect answers the question: what is the difference between the MI derived from the empirical joint distribution and one derived from our hypothetical distribution where all children had the same distribution of characteristics? It is denoted: $\Delta_X^\nu = \nu(f_{a|a}) - \nu(f_{a|i})$. Conceptually, this tells us how much of the MI is due to the varying distribution of characteristics related to parental income levels. The structure effect answers the question: what is the difference between the MI derived from our hypothetical distribution where all children had the same distribution of characteristics and one derived from the independent joint distribution? It is denoted: $\Delta_S^\nu = \nu(f_{a|i}) - \nu(f_{i|i})$. This tells us how much of the MI is due to returns to characteristics varying with parental income levels.

Together, the composition and structure effects represent a decomposition of the MI:

$$\Delta_{MI}^\nu = \Delta_X^\nu + \Delta_S^\nu. \tag{5}$$

Importantly, note that:

$$f_{a|i}(y_c, y_p) = \int f(y_c|x, y_p) f(x) dx f(y_p) \tag{6}$$

$$= \int f(y_c|x, y_p) \frac{f(x)}{f(x|y_p)} f(x|y_p) dx f(y_p) \tag{7}$$

$$= \int f(y_c|x, y_p) \psi(y_p, x) f(x|y_p) dx f(y_p) \tag{8}$$

where the 'reweighting' functions can be rewritten utilizing Bayes' rule as $\psi(y_p, x) = \frac{f(x)}{f(x|y_p)} = \frac{f(y_p)}{f(y_p|x)}$. Therefore, the counterfactual joint density can be rewritten as:

$$f_{a|i}(y_c, y_p) = \int f(y_c|x, y_p) \frac{f(y_p)}{f(y_p|x)} f(x|y_p) dx f(y_p). \tag{9}$$

From the above derivation, one can see that our method extends the approach of DiNardo et al. (1996) to joint distributions and that their approach is naturally embedded within ours as a conditional version (conditioning on a value of the membership index). Specifically, consider a version of our counterfactual that is conditioned on a specific level of $y_p = y_p^\dagger$ using the formulation in Eq. (7):

$$f_{a|i}(y_c|y_p = y_p^\dagger) = \int f(y_c|x, y_p = y_p^\dagger) \frac{f(x)}{f(x|y_p = y_p^\dagger)} f(x|y_p = y_p^\dagger) dx. \tag{10}$$

This parallels the counterfactual in DiNardo et al. (1996). Applying their approach to the gender gap where the counterfactual is assumed to be the wage distribution of women if they had the characteristics of men yields:

$$f_{w|m}(y) = \int f(y|x, sex = w)\frac{f(x|sex = m)}{f(x|sex = w)}f(x|sex = w)dx. \tag{11}$$

The comparison of Eqs. (10) and (11) highlights the relationship between the two approaches with a key difference. The counterfactual in Eq. (11) assumes one group (women) has the characteristics of the other group (men). Instead, our counterfactual (Eq. (10)) assumes every person, as well as the specific group defined by $y_p^\dagger$, has the unconditional distribution of characteristics ($f(x)$). By focusing on the full joint distribution, our counterfactual simultaneously equalizes all groups' distribution of characteristics and naturally extends the DiNardo et al. (1996) reweighting approach. In practice, this allows us to apply regression methods to the full joint distribution and obtain more general results. Further, such conditioning as in DiNardo et al. (1996) is only practical (or even feasible) with a limited number of groups. In settings with a continuum of groups, one would need to set arbitrary groupings and work with multiple conditional distributions (e.g., Richey and Rosburg, 2018). Working directly with the joint distribution is more natural and allows flexible analysis without the burden of multiple conditional distributions. Moreover, as we will see in our results, a discrete approach necessarily limits what one can learn about the counterfactual and is contingent upon the chosen segmentation.

The weights, $\psi(y_p, x)$, can be estimated by using several well-documented methods. The distribution $f(y_p)$ can be estimated via kernel density estimation while $f(y_p|x)$ can be estimated with the nonparametric conditional density estimator of Hall et al. (2004).[5,6] Once weights are estimated, they can be used in the estimation of any function of the joint distribution to obtain a counterfactual version. This is done by re-estimating weighted versions of regressions of interest using the weights $\psi(y_p, x)$. Standard errors for counterfactual regression parameters can be calculated by bootstrapping the two-step process: (1) estimate weights and (2) estimate weighted regressions.

Thus, our approach can be implemented through the following multi-step process[7]:

1. Estimate parameters of interest from regressions of y_c on y_p for the empirical and independent joint distribution—$\nu(f_{a|a})$ and $\nu(f_{i|i})$.[8]
2. Estimate $f_i(y_p)$ for each individual $i = \{1, 2, 3 \ldots N\}$ via a kernel density estimator.
3. Estimate $f_i(y_p|x)$ for each individual $i = \{1, 2, 3 \ldots N\}$ via the nonparametric method of Hall et al. (2004).
4. Construct $\psi_i(y_p, x) = \frac{f_i(y_p)}{f_i(y_p|x)}$ for each individual $i = \{1, 2, 3 \ldots N\}$.

[5]We use the KernSmooth package in R for $f(y_p)$ and the 'np' package for $f(y_p|x)$.

[6]Like all nonparametric estimators, dimensionality can be a concern for the Hall *et al.* (2004) method. However, there is some work, particularly Izbicki and Lee (2016), which shows the method performs rather well in simulation exercises based on sample sizes very similar to our application (n = 1,000 compared to our n = 1,357), though the computational time is shown rather large; and our application is based on x of dimension six, whereas their simulations go as high as twenty. Thus, while this limitation should be noted, there is some evidence regarding stability of the estimator indicating it should not be an overwhelming concern.

[7]We note here that this reweighting approach is not the only way one may go about estimating such counterfactuals, much as there are several approaches to estimating counterfactuals in the traditional discrete group decomposition literature (Chernozhukov *et al.*, 2013; DiNardo *et al.*, 1996; Firpo *et al.*, 2007; Machado and Mata, 2005; Rothe, 2015). In a previous version of this paper, under a different title, we proposed an alternative, more 'brute force' approach to the counterfactual (Richey and Rosburg, 2016). That approach paralleled Machado and Mata (2005) and took an 'estimate-and-simulate' approach that was very computationally intensive and required estimation of complex interactive conditional CDFs (i.e., $F(y_c|x, y_p)$). The approach provided here builds off the foundation of that paper, but provides what we believe is a simpler estimation procedure that avoids estimation of complex interactive conditional CDFs.

[8]Note again, for all of our applications the parameter of interest for the independent distribution is zero.

5. Estimate counterfactual parameters via weighted versions of regressions of interest—$\nu(f_{a|i})$.
6. Construct composition effects as $\nu(f_{a|a}) - \nu(f_{a|i})$.
7. Construct structure effects as $\nu(f_{a|i}) - \nu(f_{i|i})$. In our application, this is equivalent to our counterfactual estimates ($\nu(f_{a|i})$) since, as noted in footnote 4, regression coefficients for the independent case are zero.
8. Bootstrap steps 2–7 to construct standard errors for counterfactual/decomposition parameters of interest.

There are two implicit choices we have made in constructing our counterfactual: the 'order' of the decomposition and our choice for $f(x)$. Here, we briefly discuss these choices and possible alternatives.

All decompositions must choose which counterfactual to consider (i.e., the order of the decomposition). Above, we considered the counterfactual if the distribution of characteristics was independent of parental income but returns to characteristics were allowed to vary with parental income as empirically observed. An alternative is the counterfactual if the returns to characteristics were independent of parental income but characteristics varied with parental income as empirically observed. If one did wish to consider the alternate counterfactual—$f_{i|a}(y_c, y_p) = \int f(y_c|x)\tilde{\psi}f(x)dxf(y_p)$, where $\tilde{\psi} = \frac{f(y_p|x)}{f(y_p)}$—one would first need to obtain the 'independent data' to reweight. However, obtaining this independent data is not trivial. While this could be done by randomizing parental income over the other variables and then continuing as above, this approach would likely add substantial noise to the estimator without a substantially large data set. Thus, we do not pursue this counterfactual here. Which counterfactual is more appropriate will depend on the question of interest. For example, the counterfactual we consider here is most appropriate for questions related to the effects of equalizing characteristics, such as how equalizing education and test scores might shrink various MI.

Our chosen counterfactual also requires a choice regarding the distribution of characteristics assumed to prevail in the independent distribution. A computationally appealing choice, and the one assumed above, is the observed unconditional distribution of characteristics—$f(x)$. This aligns with Nopo's (2008) definition of the composition effect—the part due to differences in characteristics between individuals and the 'average'. This, however, is in contrast with most of the traditional decomposition literature. Most of the literature selects a 'base group' and the counterfactual assumes all individuals had the distribution of the base group, e.g., $f^{\dagger}(x)$; this is also the strategy chosen by Richey and Rosburg (2018), within a transition matrix setting, whereby their counterfactual is if children from all family quartiles had the distribution of characteristics as those coming from the wealthiest quartile of homes. From an interpretation standpoint, this alternative does have some advantages; the counterfactual would then consider outcomes if all children were 'lifted up' to the characteristics of those from the wealthiest homes. Conversely, our choice equalizes all children attributes to the average, which necessarily lifts up some while pushing others down. However, there are tradeoffs to the base group approach. If one wished to consider a base group $f^{\dagger}(x)$ for the counterfactual, there would be an additional term in the constructed weights: $\frac{f^{\dagger}(x)}{f(x)}$. The additional term adds noise to the estimation of the weights. Therefore, while feasible, we do not pursue a base group approach due to this concern.

Our approach aims to separate out composition and structure effects. To correctly identify these effects, we must assume that, conditional on the observable characteristics, any unobserved characteristics affecting wages are identically distributed across parental income levels. For example, if wealthier families had more 'motivated' or 'unmotivated' children (unobserved), the decomposition will still be identified as along as the distribution of motivation is identical across parental income when conditioned on observables. This assumption is generally referred to as 'conditional independence' or 'ignorability'.[9] If this assumption does not hold, our structure effect may incorrectly pick up differences in unobserved productivity characteristics. In absence of this assumption, our results would be, like many in the decomposition literature, merely descriptive in nature.

3. Application: intergenerational economic mobility

It is well established that many labor market outcomes, such as earned income, differ systematically according to parental income. A simple regression of log offsprings' incomes on log parental incomes provides the IGE, a common measure of the intergenerational link within the economic mobility literature; IGE estimates for the U.S. range from 0.30 to 0.60.[10] What is less well known is what drives this persistence of income across generations. Interest in this latter question has initiated several avenues of research. For example, researchers have attempted to identify the causal effect of parental incomes or investments on children incomes (Cardak et al., 2013; Shea, 2000), the split between nature and nurture in the connection (Björklund et al., 2006), and specific environmental drivers of the link (Chetty and Hendren, 2018); others have attempted to account for intergenerational persistence through mitigating factors (Blanden et al., 2007; Richey and Rosburg, 2017).

This paper differs from much of the mobility literature because, rather than identifying roles played by specific variables, we seek to understand how much of the intergenerational persistence is due to children from wealthier homes having different levels of characteristics and how much is due to these children receiving different returns for these characteristics. We recognize, however, that many of the effects identified in the aforementioned strands of mobility research are related to both varying levels of characteristics and varying returns to characteristics. For example, a neighborhood effect, as in Chetty and Hendren (2018), could lead to greater educational achievement but also better networking that provides greater returns for those educational gains. Thus, we believe investigating intergenerational mobility through a decomposition lens provides a complimentary perspective to previous findings. Our paper is most closely related to Richey and Rosburg (2018) who also take a decomposition approach to understanding economic mobility. Their research takes an existing discrete decomposition method (Rothe, 2015) and applies it to discrete groupings of individuals within a transition matrix setting. As such, their analysis relies on arbitrary segmentations of the distribution. The approach proposed here removes the segmentation restrictions imposed by a transition matrix setting. To illustrate how our proposed method provides a more nuanced picture of intergenerational mobility, we will provide a comparison of our model results to results based on a discretized approach.

3.1. Measures of immobility

The intergenerational mobility literature provides various measures that represent different ways to summarize or capture some aspect of the joint parent-offspring distribution of incomes. Our method can decompose any of these measures by simply re-estimating weighted versions of the measure of interest with our calculated weights—$\psi(y_p, x)$. Here, we focus on the traditional linear IGE, quantile regression counterparts to the IGE, and nonlinear versions of both. Other popular

[9]See Fortin *et al.* (2011) for a detailed discussion of identifying assumptions in decomposition methods. Also note ignorability is a less restrictive assumption than *independence*, which would require the unobservables to be independent of the covariates. Only ignorability is needed for identification of the structure-composition decomposition.

[10]See Mazumder (2005) or Black and Devereux (2011) for an overview of this literature. The large range in the estimated IGE arises from a variety of data issues including life-cycle and measurement error biases (Böhlmark and Lindquist, 2006; Haider and Solon, 2006; Nybom and Stuhler, 2017).

methods for measuring (im)mobility, not covered here, include transition matrices and directional mobility measures (Bhattacharya and Mazumder, 2011).

Much of the economic mobility literature estimates intergenerational mobility by modeling expected log earnings of a child $[ln(y_c)]$ as a linear function of log parental earnings $[ln(y_p)]$:

$$E\left[ln(y_c)|y_p\right] = \alpha + \beta ln(y_p). \tag{12}$$

The value β is the IGE and $(1-\beta)$ is a measure of intergenerational economic mobility. This simple relationship is the workhorse of much of the existing literature on economic mobility. By comparing the observed (actual) IGE with the counterfactual IGE, we can ascertain what portion of the observed IGE is structural or compositional in nature.

While the standard IGE approach tells us how the conditional mean of offspring income varies with parental income, it is often informative to look beyond the mean effect. A natural extension to the basic IGE is through quantile regressions. The goal of a quantile regression, first introduced by Koenker and Bassatt (1978), is to identify the effect of an explanatory variable on different quantiles of the conditional distribution of the dependent variable. For example, previous results for the U.S. have shown parental income has a larger effect at the lower end of children conditional income distribution, which may reflect parental income acting as a safety net (Eide and Showalter, 1999). Our method allows us to investigate not only how the relationship varies across the distribution but also what mechanism (compositional or structural) drives the relationship across the conditional distribution. We model the conditional quantile of offspring log income as:

$$Q_{ln(y_c)|ln(y_p)}(\tau) = \alpha_\tau + \beta_\tau ln(y_p) \tag{13}$$

where τ is the quantile of interest. The coefficient vector β_τ will, in general, differ for each quantile. Our decomposition method will identify the portion of each β_τ that is structural or compositional in nature.

Both of the above models are specified with linear functional forms. While such linear approximations to the conditional earnings function are quite common in the economic mobility literature, a key theme in the work by Becker and Tomes (1979, 1986) is that, in the face of credit constraints, nonlinearities may exist in the relationship between generations.[11] Furthermore, if nonlinearities exist, simple linear IGE estimates can cloud mobility comparisons (Bratsberg et al., 2007). Thus, more flexible estimation procedures through higher order polynomials, kinked/segmented regressions, or nonparametric approaches may be more appropriate for understanding the intergenerational link. Our method can handle any such approach by estimating an empirical and counterfactual version of each model with appropriate weights. In our application, we explore high order polynomial and kinked regression versions of the mean and quantile regressions to allow nonlinearities in both the empirical and counterfactual relationships. Finally, to compare our results with those derived from a discrete approach, we parallel our regression results by estimating slope parameters between the discrete distribution's statistics of interest (see Section 3.4).

3.2. Data

The data for our application is the 1979 National Longitudinal Survey of Youth (NLSY79). The NLSY79 is a panel survey of youths aged 14–22 in 1979. It includes a cross-sectional representative survey ($n = 6{,}111$), an over sample of minorities and poor whites ($n = 5{,}295$), and a sample of military respondents ($n = 1{,}280$).[12] We use only the cross-sectional representative survey.

[11]See Corak and Heisz (1999) and Grawe (2004) for work along these lines.

[12]The over sample of military and poor whites were discontinued in 1984 and 1990, respectively.

Table 1. Summary statistics–NLSY79 white males.

Variable	Mean	St. Dev.	Variable	Mean	St. Dev.
Parental income	33,542	18,237	Age	33.6	2.14
Offspring income	16,584	11,143	Rotter	8.42	2.33
Experience	12.90	3.34	Esteem	22.51	3.96
Education	13.72	2.58	Perlin	22.65	3.02
AFQT	0.52	0.91			

Notes: Incomes are constant 1982–1984 dollars. AFQT score is standardized.

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 since parents' (average) income is a key variable of interest, to those with reported parental income for those years; parental income is measured as the sum of income from both parents averaged over observation years.[13] Offspring income is average reported wage and salary income from 1994, 1996, and 1998 and limited to those not enrolled in school. All incomes are deflated to 1982–1984 dollars using the consumer price index. The data exhibit long left tails for both income distributions, and we trim the data based on a rule of thumb of two times the interquartile range below the first quartile for both parental and offspring incomes. The final sample includes 1,357 individuals with a mean age of 33.6.[14] Table 1 provides summary statistics.

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 noncognitive 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.[15] One FTE year of experience is assumed to equal 52 weeks times 40 hours (hours per week worked are top coded to 40). A few older individuals in our sample completed their education prior to the beginning of the survey and 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: $FTE_{<79} = (\text{Age}_{79} - \text{ Years of Schooling}_{79} - 6) \cdot FTE_{79}$. 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 ability used in our analysis is Armed Forces Qualifying Test (AFQT) scores. 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 et al., 2012). We use three measures for noncognitive ability. First, we use information from the Rosenberg Self-Esteem Scale (1965). The Rosenberg Self-Esteem 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)' (BLS, 2015). *Rotter* and *Esteem* were measured in the first two

[13]We exclude individuals who lived with a spouse or child during these years. Measurement error in parental income is a common concern in the mobility literature. However, recent research indicates that some mobility measures are less susceptible to such errors relative to others (Nybom and Stuhler, 2017).

[14]The literature on intergenerational mobility has identified the possibility of life-cycle biases in estimates depending on age at which children are surveyed (Böhlmark and Lindquist, 2006; Haider and Solon, 2006; Nybom and Stuhler, 2017). However, this literature seems to indicate such biases are minimized or eliminated when youths reach their mid-30s.

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

Table 2. Decomposition of linear and segmented IGEs for white males in the NLSY79.

Estimated Regression	Relationship/ Effect	Linear results IGE		Segmented regressions results: Low-side IGE		High-side IGE		Break Point	
	Empirical/total	0.346	(0.029)	0.208	(0.047)	0.596	(0.060)	10.37	(0.34)
Mean	Counterfactual/structure	0.244	(0.029)	0.100	(0.061)	0.501	(0.069)	10.36	(0.30)
	Composition	0.102	(0.020)	0.108	(0.036)	0.096	(0.037)	–	–
	Empirical/total	0.396	(0.083)	0.396	(0.083)	0.396	(0.083)	–	–
10q	Counterfactual/structure	0.335	(0.086)	0.098	(0.152)	0.532	(0.110)	10.03	(0.57)
	Composition	0.062	(0.067)	0.299	(0.145)	−0.135	(0.082)	–	–
	Empirical/total	0.316	(0.028)	0.259	(0.041)	0.662	(0.146)	10.71	(0.57)
50q	Counterfactual/structure	0.232	(0.033)	0.180	(0.049)	0.503	(0.154)	10.65	(0.60)
	Composition	0.084	(0.027)	0.078	(0.028)	0.159	(0.137)	–	–
	Empirical/total	0.423	(0.060)	0.013	(0.134)	0.828	(0.077)	10.35	(0.18)
90q	Counterfactual/structure	0.177	(0.063)	−0.213	(0.184)	0.850	(0.130)	10.34	(0.26)
	Composition	0.245	(0.055)	0.226	(0.111)	−0.021	(0.108)	–	–

Notes: Standard errors are in parentheses. Standard errors are based on 500 bootstraps, which incorporate both stages of estimation. For segmented regression results, the standard errors are calculated with a fixed break point within the bootstrap procedure based on the empirically estimated break point.

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 late 20 s or early 30 s. The Perlin Mastery Scale measures the extent to which individuals ‘perceive themselves in control of forces that significantly impact their lives’ (BLS, 2015).

3.3. Results

Before discussing our empirical results, let us overview the structure of how we present our results. Table 2 reports our linear and segmented IGE results and the relationship between estimated effects. The four horizontal panels present results for the mean, 10th, 50th, and 90th quantile regression results. Each of these panels contains three rows: ‘Empirical/Total’, ‘Counterfactual/Structural’, and ‘Composition’. The ‘Empirical/Total’ row presents the regression results for the empirically observed relationship, so the first entry of 0.346 is the linear IGE from our actual data set. It is important to note that since this relationship for the independent joint distribution is zero, these values also represent the MI of interest. The ‘Counterfactual/Structure’ row represents the regression results for our counterfactual joint distribution, which is achieved with weighted counterparts to the results in the first row. Recall that the structure effect is the difference between the counterfactual and the independent relationship ($\Delta_S^\nu = \nu(f_{a|i}) - \nu(f_{i|i})$). Again, since an independent relationship would lead to regression parameters equivalent to zero, the counterfactual and structure effect are equivalent. The ‘Composition’ row presents our estimated composition effect or the difference between the empirical relationship and the counterfactual ($\Delta_X^\nu = \nu(f_{a|a}) - \nu(f_{a|i})$). The first column of results are estimates for slope parameters for linear specifications. Columns two through four present results for segmented regression specifications. The ‘Low-Side’ column presents slope parameters estimated below the break point in parental income. The ‘High-Side’ column presents slope parameters estimated above the break point. Finally, the ‘Break Point’ column presents estimated kinks in the linear relationship. Figure 1 provides a visual depiction of our nonlinear analysis. The predicted empirical and counterfactual relationships for the conditional mean and select quantiles are based on high order polynomial regressions (fifth order in parental log income). The high-ordered models are demanding on our data and confidence bands (not shown) tend to be large and overlap; nonetheless, they allow for a visual depiction of the nonlinear relationships. While these polynomial regressions are more flexible than our kinked regressions, the slope estimates from the segmented regressions are useful for comparative purposes.

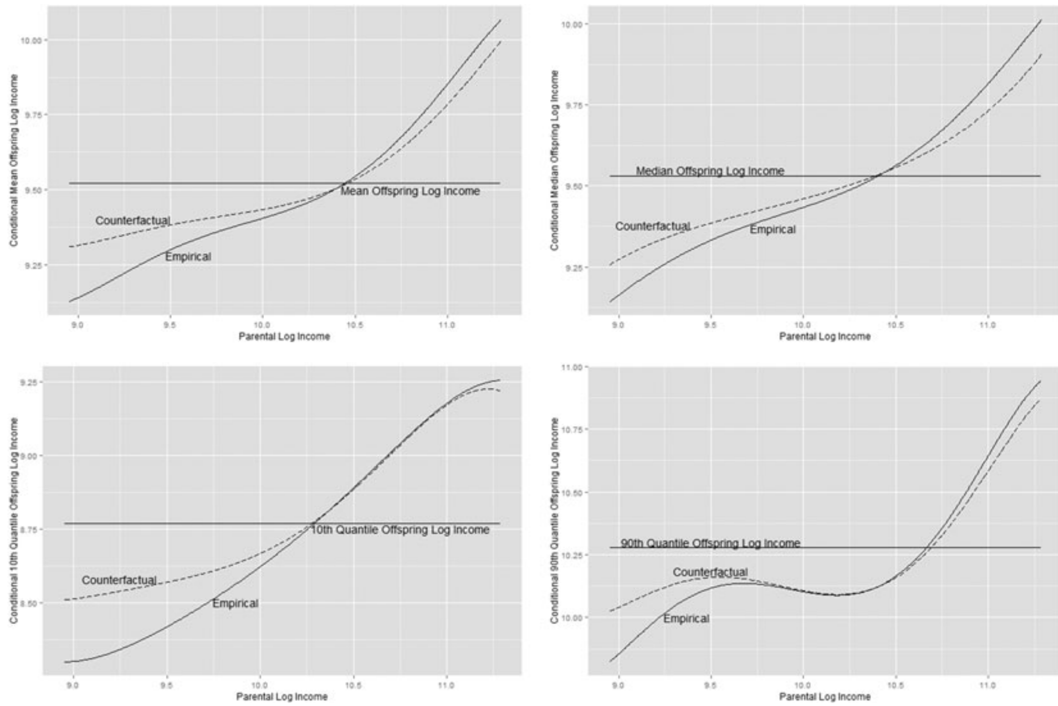

Figure 1. Top left: mean relationship; top right: median relationship; bottom left: 10th quantile relationship; bottom right: 90th quantile relationship.

First, let us discuss the linear specification results provided in the first column of Table 2. The empirical standard IGE is estimated at 0.35 and implies that as parental income increases by 1%, offsprings' expected income increases by 0.35%; this is on the lower end of the IGE estimates reported in existing literature.[16] The decomposition shows that the structure effect accounts for about 70% of the IGE while the composition effect accounts for the other 30%. Therefore, in explaining (im)mobility for white males in the U.S., differences in returns to characteristics across households has about double the explanatory power as differences in offspring characteristics across households. Conditional on the ignorability assumption holding, our results suggest there is a considerable unobserved labor market advantage related to parental income beyond the role played by higher levels of productive characteristics.

Linear quantile IGE counterparts reveal a U-shaped effect with larger effects in the tails (i.e., 10th and 90th quantiles). As parental income increases, offsprings' conditional wage distribution exhibits compression at the lower end (decreasing 10–50 gap) and an extending upper tail (increasing 50–90 gap). This increased effect at the lower end is consistent with Eide and Showalter (1999) and suggests that parental income has a safety-net effect. Decomposition results suggest the structure effect is the main driver at the 10th quantile at 85% of the total, which then decreases to 42% at the 90th quantile as the composition effect plays a larger role. Therefore, the compression at the lower end of the distribution as parental income increases appears to be driven by the structure effect (the structure effect increases from 0.23 to 0.33 from the median to the 10th quantile) while the extending tail appears to be driven by the composition effect (the composition effect increases from 0.08 to 0.25 from the median to 90th quantile).

[16]It is not unexpected that our IGE estimate is on the lower end of the range reported in existing literature (0.30 - 0.60). IGE estimates tend to be lower with shorter income averages (see Mazumder (2005) or Black and Devereux (2011) for discussions of measurement error as it relates to this issue), and we use a two-year average. Ideally, we would use a longer observation time frame but doing so has serious effects on our sample size and hinders our ability to carry out the decomposition.

At first glance, these results seem to imply some important takeaways regarding how policies aimed at the structure or composition effect might affect mobility; that is, policies aimed at the composition effect will have a greater effect on the outcomes of the right tail of the conditional distribution while policies aimed at the structure effect will have a greater effect on the left tail of the conditional distribution. However, as noted previously, linear specifications may not be a good approximation to the empirical or counterfactual intergenerational relationships. Therefore, we turn to our nonlinear results.

The polynomial conditional mean results, as well as segmented mean regressions, reveal nonlinearities in the intergenerational relationship with a larger IGE at higher levels of parental income. The empirical IGE is 0.20 at low parental income levels ('Low-Side IGE') and 0.60 at high parental income levels ('High-Side IGE'). Moreover, decompositions illustrate that removal of the composition effect has differing effects for the average son from lower income households than higher income households. The segmented regressions suggest an even 50–50 structure-composition split at the lower end of parental income but an 85–15 split at higher income levels. Therefore, the 70–30 structure-composition split suggested by the linear results masks some of the underlying heterogeneity.

A comparison of the linear and nonlinear quantile regression results reveals similar differences. While the empirical 10th quantile conditional relationship is linear in parental income (thus, the identical linear and low/high side IGE and no estimated break point), the 50th quantile has a nonlinear relationship similar to that of the mean. The 90th quantile is extremely nonlinear with no discernable relationship at low parental income levels and a very high estimated IGE at high parental income levels (0.83). These results imply a compression of the left tail of the distribution as parental income increases only among sons from lower income households (0.40 IGE at the 10th quantile vs. 0.26 IGE at the 50th quantile for the 'Low-Side IGE'); for sons from higher income households ('High-Side IGE'), the 10–50 spread increases as parental income increases (0.40 IGE at the 10th quantile vs. 0.66 IGE at the 50th quantile). In a similar pattern, we see a shrinking 50–90 gap among sons from lower income households but a widening gap for sons from higher income households. This implies a uniform compression of the conditional distribution as parental income increases at low levels and a uniform spreading at higher parental income levels—a more complex relationship than the linear results are able to portray. Recall, the linear results suggested a simple compression at the left tail and extension at the right tail.

Decomposition results suggest similar heterogeneity in the drivers of these relationships. At low parental income levels, the composition effect dominates at the 10th conditional quantile—75% of the empirical relationship based on segmented regressions—and is the driver of the left tail compression as parental income increases at low levels. Therefore, the 'safety net' effect at low parental income levels appears to be driven by the composition effect; linear results suggested the structure effect was the driver. If we focus on the spreading 50–90 gap at high parental income levels, it is driven by the structure effect, which increases from 0.50 to 0.85 from the 50th to 90th quantile. The linear results suggested the widening 50–90 gap was due to the composition effect. Looking closer, this appears to be an anomaly of the very flat empirical relationship at low levels changing to a negative relationship predicted by the counterfactual (though with large standard errors and not statistically different from zero). The statistical insignificance of the structure effect in the low-side for the 90th quantile makes it hard to assign the key driver to the compression of the 50–90 gap at lower parental income levels. Similarly, the statistically insignificant composition effects in the high-side for the 10th and 50th quantiles make it hard to assign roles for the widening 10–50 gap at higher parental income levels.

Together, these results indicate a very complex relationship with differing trends at low and high parental income levels. Differences are found not only in the empirical relationship but also in the decompositions; the role of the structural and compositional components in different segments of the conditional distribution of offspring income varies with the parental income range (i.e., 'Low-Side' vs 'High-Side'). For example, the composition effect is the driver of the safety net effect at low parental income levels. Thus, policies aimed at the composition effect would likely benefit the worst performing sons from the poorest households. Alternatively, the spreading 50–90 gap at high parental income levels is driven by the structure effect. This indicates a strong role of 'advantage' in explaining the high incomes of the best performers from the highest income households.

While not directly comparable, these two findings are generally in line with the findings of Richey and Rosburg (2018). They find that the composition effect accounts for about 60% of the increased likelihood of children from the poorest quartile of households ending up in the lowest income quartile as compared to children from the second poorest household quartile. They also find the composition effect only accounts for about 25% of the increased likelihood children from the wealthiest quartile end up in the highest income quartile as compared to children from the second wealthiest homes. Furthermore, again noting differences in methods and application, Ulrick et al. (2018) also find that differences in college grades is better explained by differences in characteristics at the lower end of the socio-economic continuum. These similarities suggest our findings may be indicative of a general relationship between parental incomes and several important economic outcomes.

While all of these takeaways hinge on the ignorability assumption, we believe that such nuanced findings (even if only descriptive in nature) provide interesting contributions to the mobility literature that at a minimum point towards areas of potential further study. Moreover, the ability of the procedure to reveal such heterogeneous effects highlights the flexibility of our proposed decomposition method. Our results are also contingent on the observed cohort—the 1979 NLSY. Drivers of mobility, similar to drivers of wage gaps, are likely to evolve over time (Albrecht et al., 2015; Maasoumi and Wang, 2019; Richey and Rosburg, 2017). The NLSY has a second survey (the 1997 NLSY) that includes a later generation of children, but this cohort does not include the noncognitive ability measures and the most recent available data is the 2015 wave. Using parallel assumptions for data selection would provide respondents at an average age of 30.7 compared to 33.6 in our 1979 NLSY data set. As discussed in the data section, life-cycle bias is a major concern in the mobility literature. Not only is the lower average age in the 1997 cohort a concern, the cohort also has an increase in education attainment which (on average) would have delayed their working wage profiles compared to the 1979 cohort. Given these concerns, along with our interest in the tails of the distribution and the increased effect at higher parental incomes where the bias may be more prevalent, we do not believe the currently available 1997 NLSY cohort would provide a suitable comparative analysis, and therefore, we do not pursue it here.

3.4. Comparison with a discrete approach

The method proposed and applied here is a natural extension and improvement over the approach provided in Richey and Rosburg (2018). That article, like others in the decomposition literature that address continuum groups, relies on discretizing the group membership variable. In the case of Richey and Rosburg (2018), the discretization of choice is quartiles, which naturally fit the article's application of interest—transition matrices. However, a discrete approach is more restrictive than the approach proposed here and has several limitations if applied outside of the transition matrix setting. In particular, it imposes that everyone in a given (discretized) group has the same structure effect. Results will therefore depend on (and vary with) the chosen discretization. Further, as will be illustrated below, results from a discrete approach are simply a discrete number of conditional distributions such that 'quantile regressions' would simply yield three 'slopes' between the four distribution's quantiles. Moreover, to form such slopes, levels for

Table 3. Decomposition of slopes in mean/quantiles in children incomes between parental income quartile groups based on discrete approach for white males in the NLSY79.

		Mean	10th Quantile	50th Quantile	90th Quantile
First to second	Empirical	0.168 (0.069)	0.294 (0.178)	0.224 (0.076)	–0.111 (0.116)
Parental	Counterfactual/structure	0.020 (0.068)	0.105 (0.173)	0.141 (0.081)	–0.242 (0.148)
Quartile	Composition	0.148 (0.051)	0.189 (0.137)	0.082 (0.064)	0.131 (0.122)
Second to third	Empirical	0.402 (0.144)	0.677 (0.385)	0.273 (0.176)	0.488 (0.217)
Parental	Counterfactual/structure	0.162 (0.128)	0.191 (0.366)	0.088 (0.157)	0.242 (0.310)
Quartile	Composition	0.239 (0.079)	0.486 (0.239)	0.185 (0.101)	0.247 (0.218)
Third to Fourth	Empirical	0.575 (0.097)	0.428 (0.228)	0.542 (0.115)	0.952 (0.258)
Parental	Counterfactual/structure	0.395 (0.094)	0.373 (0.273)	0.248 (0.126)	0.672 (0.203)
Quartile	Composition	0.180 (0.055)	0.055 (0.172)	0.293 (0.089)	0.280 (0.182)

Notes: Standard errors are in parentheses. Standard errors are based on 500 bootstraps, which incorporate both stages of estimation.

parental income must be chosen which leads to results that are conditioned on another (researcher-selected) choice.

To illustrate the advantages of our proposed approach, we estimate parallel results to those in Table 2 using a discrete approach. We first discretize the data into quartiles according to parental incomes. Then, we conduct four decompositions using the approach of DiNardo et al. (1996) (see Eq. (10)) where there are four groups with $y_p^\dagger$ defined as being in each parental income quartile. This yields four counterfactual distributions from which we can calculate means and select quantiles as well as changes between the groups. To estimate 'slope' parameters to compare with the results in Table 2 and Fig. 1, we use the mean of the parental incomes in each group to estimate simple slopes between the successive groups.[17] For example, the top left entry in Table 3 reports the empirical 'slope' between the means of incomes from children from the first and second parental quartile. This slope is estimated by dividing the difference in the means between these children incomes by the difference in the mean of their parental incomes. Similarly, the second column of the first row divides the difference in the 10th quantile of these children incomes by the difference in the average of their parental incomes. These yield quasi-slope parameters across parental income levels to compare to our main regression parameter estimates as well as our nonparametric graphs. In what follows, we refer to the results and findings based on our proposed method as the 'main results'.

The 'slope' results for the discrete approach are given in Table 3. The mean results (first column) are generally consistent with our main results: the empirical relationship increases as parental income increases, and the composition effect dominates on the lower end while the structure effect becomes increasingly important at the top end. The results for the decomposition at the 10th quantile (second column) also generally align with our main results, although the magnitudes of the empirical relationship present a changing relationship (0.29 to 0.67 to 0.43) not observed in our main results; this difference is likely due to the arbitrary segmentation and the fairly large standard errors. The 50th quantile results (third column) differ in magnitude to our main results. The structure effect is $<50\%$ of the decomposition at the top end of parental income in the discretized results (0.25 of 0.54) but explains almost all of the observed immobility in our main results. The 90th quantile results also, in general, keep the general sense of our main results with the overall relationship increasing with parental income and the structure effect becoming increasingly important; however, these results leave about 25% of the explanatory power to the composition effect at the top, something not observed in our main results.

While the discrete results align with some of the major takeaways from our main results, they also highlight some of the shortcomings of a discrete approach compared to our proposed approach. In particular, with the discrete approach, we cannot be confident in the magnitudes of

[17] While we have the actual parental income for each individual, the decomposition treats everyone in each quartile as simply a group member and thus it would be incorrect to reinsert this information to attempt actual regressions.

the effects as they are dependent on the segmentation. This creates some results that seem to contradict our main results (changing magnitudes of the 10th quantile effect, the smaller role for the structure effect at the 50th quantile at the top end, and the non-negligible role of the composition effect at the 90th quantile at the top end). Also, due to the segmentation of the data, there is increased uncertainty in the 'slope' estimates, which creates challenges for inference. Further, while some of the major takeaways of the results from both methods generally align, it is more natural to work with one joint distribution and analyze direct regression results rather than comparing quantiles and means between arbitrary discrete groups.

4. Conclusion

Recent advances in decomposition methods have expanded the applicability beyond their traditional settings. Many methods have been introduced to evaluate differences across the full distribution of the outcome of interest, and to a lesser degree, some have been introduced to expand beyond discrete group settings. This paper proposes a simple method that extends the traditional decomposition method in both directions simultaneously. Our method achieves this by reframing the problem: rather than ask why groups differ in their outcomes, we ask why the joint distribution of group membership and outcome are not independent. Thus, the focus is on decomposing the difference between the empirical and an independent joint distribution of outcome and group index. Reweighting functions are used to generate a counterfactual joint distribution which represents outcomes if characteristics were independent of group membership but returns to characteristics still depended on group membership; this counterfactual allows us to identify what part of the relationship between the outcome and the group membership is due to characteristics varying with group membership (a composition effect) and what part is due to returns to characteristics varying with group membership (a structure effect). This method can easily be applied to any distributional feature of interest of joint distributions.

We applied the proposed method to the economic mobility of a cohort of white males surveyed in the 1979 NLSY and investigated multiple versions of the IGE. We find a standard IGE of 0.35 with a 30–70 composition-structure split in explanatory power. However, this simple mean relationship hides large nonlinearities across parental income levels as well as large heterogeneity in drivers across the son's conditional quantiles. Quantile and nonlinear results suggest a more complex relationship that can easily be overlooked or misrepresented by focusing on mean or linear specifications. Overall, our application highlights the important role that varying returns to characteristics across parental income levels plays in explaining intergenerational mobility, something that, to our knowledge, has not been explored in previous mobility literature.

While we focus our application on economic mobility, the method proposed here can be applied to any joint distribution where the relationship is believed to be due to differing levels of intermediate variables and differing returns to them across group membership. Furthermore, while we focus on mean and quantile regressions as well as nonlinear versions of each, the method can be applied to any function of joint distributions such as transition matrices, directional migration measures, or other more complex measures.

Acknowledgments

The views expressed in this paper are our own and do not reflect the views of the Office of the Comptroller of the Currency, the US Department of the Treasury, or any federal agency. The views in the paper do not establish supervisory policy, requirements, or expectations. We would like to acknowledge useful feedback and comments from seminar participants at the Office of the Comptroller of the Currency Economics Department, the editor, Esfandiar Maasoumi, as well as an associate editor and three anonymous referees. All errors are our own.

References

Albrecht, J., van Vuuren, A., Vroman, S. (2015). The black-white wage gap among young women in 1990 vs. 2011: the role of selection and educational attainment. *Labour Economics* 33:66–71. doi:10.1016/j.labeco.2015.02.006

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. doi:10.1086/666536

Becker, G., Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy* 87(6):1153–1189. doi:10.1086/260831

Becker, G., Tomes, N. (1986). Human Capital and the rise and fall of families. *Journal of Labor Economics* 4(3, Part 2):S1–S39. doi:10.1086/298118

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. doi:10.3982/QE69

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. doi:10.1162/qjec.121.3.999

Black, S., Devereux, P. (2011). Recent developments in intergenerational mobility. In: Card, D., Ashenfelter, O., eds. *Handbook of Labor Economics*, Amsterdam: Elsevier, Volume 4(B), pp. 1487–1541.

Blanden, J., Gregg, P., Macmillan, L. (2007). Accounting for intergenerational income persistence: noncognitive skills, ability and education. *The Economic Journal* 117(519):C43–C60. doi:10.1111/j.1468-0297.2007.02034.x

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *The Journal of Human Resources* 8(4):436–455. doi:10.2307/144855

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. doi:10.1086/506489

Bratsberg, B., Røed, K., Raaum, O., Naylor, R., JäNtti, M., Eriksson, T., ÖSterbacka, E. (2007). Nonlinearities in intergenerational earnings mobility: consequences for cross-country comparisons. *The Economic Journal* 117(519):C72–C92. doi:10.1111/j.1468-0297.2007.02036.x

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

Cameron, S., Heckman, J. (2001). The dynamics for educational attainment for black, Hispanic, and white males. *Journal of Political Economy* 109(3):455–499. doi:10.1086/321014

Cardak, B., Johnston, D., Martin, V. (2013). Intergenerational income mobility: a new decomposition of investment and endowment effects. *Labour Economics* 24:39–47. doi:10.1016/j.labeco.2013.05.007

Chernozhukov, V., Fernandez-Val, I., Melly, B. (2013). Inference on counterfactual distributions. *Econometrica* 81: 2205–2268. doi:10.3982/ECTA10582

Chetty, R., Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. *The Quarterly Journal of Economics* 133(3):1107–1162. doi:10.1093/qje/qjy007

Corak, M., Heisz, A. (1999). The intergenerational earnings and income mobility of Canadian men: evidence from longitudinal income tax data. *The Journal of Human Resources* 34(3):504–533. doi:10.2307/146378

DiNardo, J., Fortin, N., Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. *Econometrica* 64(5):1001–1044. doi:10.2307/2171954

Eide, E., Showalter, M. (1999). Factors affecting the transmission of earnings across generations: a quantile regression approach. *The Journal of Human Resources* 34(2):253–267. doi:10.2307/146345

Firpo, S., Fortin, N., Lemieux, T. (2007). Decomposing wage distributions using recentered influence function regressions. Mimeo: University of British Columbia. http://www.economics.uci.edu/files/docs/micro/f07/lemieux.pdf

Fortin, N., Lemieux, T., Firpo, S. (2011). Decomposition methods in economics. In: Ashenfelter, O., Layard, R., Card, D., eds. *Handbook of Labor Economics*, New Holland: National Bureau of Economic Research, Inc, Volume 4(A), pp. 1–102.

Frenette, M. (2007). Why are youth from lower-income families less likely to attend university? evidence from academic abilities, parental influences, and financial constraints. Analytical Studies Branch Research Paper Series, No. 295. Statistics Canada, Catalogue No. 11F0019MIE - No. 295.

Grawe, N. (2004). Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints. *The Journal of Human Resources* 39(3):813–827. doi:10.2307/3558998

Haider, S., Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review* 96(4):1308–1320. doi:10.1257/aer.96.4.1308

Hall, P., Racine, J., Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99(468):1015–1026. doi:10.1198/016214504000000548

Izbicki, R., Lee, A. (2016). Nonparametric conditional density estimation in a high-dimensional regression setting. *Journal of Computational and Graphical Statistics* 25(4):1297–1316. doi:10.1080/10618600.2015.1094393

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica* 46(1):33–50. doi:10.2307/1913643

Maasoumi, E., Wang, L. (2019). The gender gap between earnings distributions. *Journal of Political Economy* 127(5):2438–2504. doi:10.1086/701788

Machado, J., Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* 20(4):445–465. doi:10.1002/jae.788

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. doi:10.1162/0034653053970249

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

Nopo, H. (2008). An extension of the Blinder-Oaxaca decomposition to a continuum of comparison groups. *Economics Letter* 100:292–296. doi:10.1016/j.econlet.2008.02.011

Nybom, M., Stuhler, J. (2017). Biases in standard measures of intergenerational income dependence. *Journal of Human Resources* 52(3):800–825. doi:10.3368/jhr.52.3.0715-7290R

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review* 14(3): 693–709. doi:10.2307/2525981

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(2):463–499. doi:10.1007/s00148-007-0180-5

Richey, J., Rosburg, A. (2016). Understanding intergenerational economic mobility by decomposing joint distributions. Working paper. Available at: https://ideas.repec.org/p/pra/mprapa/72665.html. Last accessed 26 November 2019.

Richey, J., Rosburg, A. (2017). Changing roles of ability and education in U.S. intergenerational mobility. *Economic Inquiry* 55(1):187–201. doi:10.1111/ecin.12362

Richey, J., Rosburg, A. (2018). Decomposing economic mobility transition matrices. *Journal of Applied Econometrics* 33(1):91–108. doi:10.1002/jae.2578

Rothe, C. (2015). Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes. *Journal of Business & Economic Statistics* 33(3):323–337. doi:10.1080/07350015.2014.948959

Shea, J. (2000). Does parents' money matter? *Journal of Public Economics* 111(3):358–368. doi:10.1016/S0047-2727(99)00087-0

Ulrick, S. (2012). The Oaxaca decomposition generalized to a continuous group variable. *Economics Letters* 115(1): 35–37. doi:10.1016/j.econlet.2011.11.037

Ulrick, S., Mongeon, K., Giannetto, M. (2018). A continuous income-based grade gap decomposition on university course grades. *Applied Economics* 50(31):3388–3404. doi:10.1080/00036846.2017.1420893