Relative Distributional Methods

by

Mark S. Handcock, Martina Morris and Annette D. Bernhardt*

March 14, 1997

Submitted to *Sociological Methodology.*

ABSTRACT

Relative distribution methods are a non–parametric statistical framework for analyzing data in a fully distributional context. The methods combine the graphical tools of exploratory data analysis with a framework for statistical decomposition and inference. The relative distribution is similar to a density ratio, and is based on the direct comparison of one distribution to another. It is technically defined as the random variable obtained by transforming a variable from a comparison group by the cumulative distribution function (CDF) of that variable for a reference group. This transformation produces a set of observations, the relative data, that represent the rank of the original comparison value in terms of the reference group's CDF. The relative data preserve the information needed to compare the two original distributions. The density and CDF of the relative data can therefore be used to fully represent and analyze distributional differences. Analysis can move beyond comparisons of means and variances to fully tap the information inherent in distributions. The analytic framework is general and flexible, as the relative density is decomposable into location, shape and covariate effects.

Keywords: Grade transformation; Graphical methods; Inequality; Lorenz curve.

---

* Mark S. Handcock is Associate Professor in the Department of Statistics, and Martina Morris is Associate Professor in the Departments of Sociology and Statistics, both at The Pennsylvania State University, University Park, PA 16802. Both are members of the Population Research Institute. Annette D. Bernhardt is Senior Research Associate, Institute on Education and the Economy, Teachers College, Columbia University, New York, NY 10027.

1. Introduction and motivation

In social science research, theories regarding differences among groups or changes over time often imply properties of distributions that are not well captured by the usual summary measures of location and variation. Consider some of the current controversies regarding growing inequality in earnings, racial differences in test scores, and the socio–economic correlates of birth weights and survival times. The patterns of inequality that animate the debates in these fields are complex – they comprise the usual mean–shifts and changes in variance, but also more subtle comparisons of changes in the upper and lower tails of the distributions. While survey and census data on such attributes contain a wealth of distributional information, traditional methods of data analysis leave much of this information untapped. And while exploratory data analysis techniques provide important complementary tools for traditional methods, methods that combine the exploratory power of EDA with a framework for statistical inference and estimation remain rare.

In this paper we present a statistical approach to full distributional comparison based on the relative distribution. The PDF and CDF associated with the relative distribution capture information equivalent to the density ratio between two distributions. In contrast to the Lorenz curve, which compares a single distribution to exact equality, the relative distribution directly compares two distributions to each other. In this sense, the relative distribution plays the same role in inter–distributional comparison that the Lorenz curve plays in intra–distributional comparison: it preserves all of the information necessary to compare two distributions. The relative distribution also provides a general integrated framework for analysis: a graphical component that simplifies exploratory data analysis and display, a statistically valid basis for the development of hypothesis–driven summary measures, and the potential for decomposition that enables one to examine complex hypotheses regarding the origins of distributional changes within and between groups. We demonstrate the use of the relative distribution for each of these analytic tasks in this paper. The integration of the different components of analysis, the capability of representing full distributional information, and the flexibility for representing the factors that are of substantive interest, make the relative distribution approach well suited to emerging research questions in many fields.

The paper is divided into five sections. In the remainder of this section we describe the data that will be used in the examples. In the next section, the relative distribution is formally defined. In section three, we present a decomposition of the relative distribution into median and shape shifts. In section four we develop summary measures to quantify and statistically test two natural hypotheses about distributional change. The first measure captures the contribution of components such as the median and shape shifts to the overall difference between the distributions, and the second captures the extent to which the shape difference takes the form of relative polarization, or rising inequality. In section five, we present methods for covariate adjustment and analysis. The decomposition technique and summary statistics from the previous sections are applied here to separate the impact of compositional changes in population characteristics from other changes. In the case of discrete covariates, decomposition techniques can also be used to compare distributional changes within and between covariate levels. Much of the work presented here draws on and formalizes work developed in Morris (1994), Bernhardt (1995), and Morris (1996).

1.1 Data

The examples that will be used throughout this paper are motivated by recent research on growing wage inequality in the United States (for reviews see Danziger and Gottschalk 1993; Danziger and Gottschalk 1996; Karoly 1993; Levy and Murname 1992). One of the important questions to emerge in this research concerns the issue of lifetime wage mobility: to what extent does the observed cross–sectional growth in wage inequality translate into growing segmentation in the distribution of life–time wages? If workers' lifetime wage mobility is high, then the cross–sectional trends are less worrisome (for a review of literature in this area, see Atkinson, et al. 1992). For example, growing dispersion in wages could simply reflect a widening wage gap between younger and older workers. In this case, lifetime wage mobility would prevent a permanent segregation of high and low wage workers. Another possibility is that there is simply more volatility in wages now, potentially due to more frequent job changes, but that the long–term trajectories of workers' wages are no more polarized than before (Gottschalk and Moffit 1994; Haider 1997; Stevens 1996). On the other hand, there is some reason to expect that the cross–sectional trends do reflect growing labor market segmentation. Restructuring strategies at the firm level are changing the structure of internal labor markets and may be permanently changing the distribution of economic opportunities (Cappelli 1994; Harrison 1994). When firms replace on–the–job training and promotion with external hiring, or substitute temporary workers at the bottom of the job hierarchy, the traditional routes to career mobility are disrupted, especially for low–skill workers. If these are the forces driving the cross–sectional wage polarization, individual wage trajectories may also become more polarized, with some workers increasingly stuck in a series of low wage marginalized jobs, while others experience "winner take all" wage gains. This scenario suggests permanent changes in the distribution of mobility and the emergence of a more rigidly segmented labor market.

The longitudinal panels of the National Longitudinal Survey (NLS) data provide an opportunity to investigate these questions by comparing the wage growth profiles across cohorts. The data we use here come from two cohorts of the NLS, one initiated in 1966, the other in 1979. Both cohorts are 14–22 years old at the start of the survey and are followed for the first decade and a half of work experience (through 1981 for the original cohort - 16 years, and 1993 for the recent - 15 years). For simplicity here we restrict the examples to white males.

We focus on the life–cycle growth of "permanent earnings,"[1] highlighting the patterns of wage mobility. Permanent earnings for each respondent are represented as:

$$y_{it} = b_{0i} + b_{1i}\text{age}_t + b_{2i}\text{age}_t^2 \tag{1.1}$$

where y_{it} is the log of real wages for respondent i at time t. Each of the coefficients b_{0i}, b_{1i} and b_{2i} represent a combination of the fixed and random effects for the life–cycle growth in wages:

$$b_{ji} = \beta_j + \tau_{ji}, \qquad \tau_{ji} \sim N(0, \sigma_j^2) \quad j = 0, 1, 2 \tag{1.2}$$

where the β_j are the "fixed effects", and the τ_{ji} are independent random draws from Gaussian distributions. The fixed effects quadratic in age, $\beta_0 + \beta_1\text{age}_t + \beta_2\text{age}_t^2$, captures the mean growth in wages over the life–cycle, and the τ_{ji} capture the heterogeneity in individual profiles. We use this mixed–effects specification to fit a wage profile for each respondent over the 15 year observation period. An idealized distribution of these profiles is shown in Figure 1.

[1] Wages can be thought of as having a permanent and a transitory component. Permanent earnings are generally defined as the age–earnings profile predicted by a mixed effects model, where the effects of age on earnings are modeled with a fixed effect to capture the population average and a random effect to capture population profile heterogeneity. The residual from the predicted profiles is defined as the transitory variance in wages. For examples, cf. Gottschalk (1994) and Haider (1997).

Figure 1 about here

Wage growth for each respondent is then defined as

$$w_i = (y_{it} \mid \text{age}_t = 34) - (y_{it} \mid \text{age}_t = 16) \qquad i = 1, \ldots, n \tag{1.3}$$

All of the examples in this paper use the distributions of wage growth for the original and recent cohorts. Figure 2 shows the straight–forward density overlays for these two distributions. Several aspects of the relative wage growth in the two cohorts are apparent from this graph: the recent cohort experienced smaller average wage gains, these gains were more variable, and while the frequency of high wage gains was comparable to the original cohort, the frequency of low wage gains was much greater.

Figure 2 about here

The key theoretical questions here – how has the distribution of wage–growth changed, is there evidence of growing segmentation, and what is the role played by changes in covariates such as age and education – are questions that are well answered by relative distribution techniques.

2. The relative distribution

2.1 Definition

Let Y_0 be a random variable representing an outcome attribute for a population (e.g., permanent wage gains). We will call the population that generated Y_0 the *reference* population. Denote the cumulative distribution function (CDF) of Y_0 by $F_0(y)$ and the density by $f_0(y)$ (when the latter is defined). We do not place restrictions on the outcome space of the reference measurement, although in many applications it will only take on non–negative values

Suppose we observe another measurement of Y from a different population. We will call the population that generated Y the *comparison* population. It is assumed that Y has CDF $F(y)$ and density $f(y)$ (when the latter is defined). Typically Y is the measurement for a separate group or the same group during a later time period.

The objective is to study the differences between the distributions of the outcome attribute in the reference and comparison populations.

For continuous distributions, we define the *relative distribution* of Y to Y_0 to be the random variable:

$$R = F_0(Y) \tag{2.1}$$

So R is obtained from Y by transforming it by the function F_0. We will refer to the realization of R, r, as the *relative data*. If both Y and Y_0 are (absolutely) continuous distributions, R will be continuous with outcome space $[0, 1]$. We can express the CDF of R as

$$G(r) \;=\; F(F_0^{-1}(r)) \qquad 0 \le r \le 1.$$

where $F_0^{-1}(r)$ is the quantile function of F_0. The probability density function (PDF), when it exists, can be obtained as the derivative of $G(r)$:

$$g(r) = \frac{f(F_0^{-1}(r))}{f_0(F_0^{-1}(r))} \qquad 0 \le r \le 1$$

where r represents the proportion of values.

If the two distributions are identical then the PDF of the relative distribution is the uniform PDF. In general, the relative distribution is invariant to the scale of the distributions (up to a monotone transformation). For example, one obtains the same relative distribution from a comparison of log–wages as from the comparison of raw wages. We discuss the scale invariance in more detail in section 2.3 below.

The relative distribution is an intuitively appealing approach to the comparison problem because both the PDF and the CDF have clear, simple interpretations. The relative PDF $g(r)$ can be interpreted as a density ratio: the ratio of the fraction of respondents in the comparison group to the fraction in the reference group at a given level of the outcome attribute Y ($F_0^{-1}(r)$). This can be seen more easily by reexpressing the relative PDF explicitly in terms of the original measurement scale, y. Let the r^{th} quantile of R be denoted by the value y_r on the original measurement scale; so the y_r corresponding to r is $F_0^{-1}(r)$. The relative PDF is then:

$$g(y_r) = \frac{f(y_r)}{f_0(y_r)} \qquad y_r \ge 0$$

The relative CDF, $G(r)$, can be interpreted as the proportion of the comparison group whose attribute lies below the r^{th} quantile of the reference group. More technically: a proportion $G(r)$ of the Y are below the values of a proportion r of Y_0. Note that even though the relative CDF is explicitly scaled in terms of quantiles, the implicit unit of comparison is the value of the attribute on the original measurement scale ($y_r = F_0^{-1}(r) = F^{-1}(G(r))$ representing the cut–point). The relative data, r, also have a simple interpretation: they represent the rank that the original comparison value would have had in the reference CDF.

2.2 Estimating the Relative CDF and PDF

In this section we briefly consider the estimation of the relative CDF $G(r)$ and relative PDF $g(r)$. The methodology is developed in Handcock and Janssen (1996a), Cwik and Mielniczuk (1989; 1993), and also Li, Tiwari & Wells (1996). We will only give a brief outline here.

Let $Y_1, Y_2, \ldots, Y_m$ be independently and identically distributed from the distribution F. Similarly, let $Y_{01}, Y_{02}, \ldots, Y_{0n}$ be independently and identically distributed from the distribution F_0. We consider the situation where the $\{Y_j\}_{j=1}^m$ are independent of $\{Y_{0i}\}_{i=1}^n$ [2]. Let $F_m(y) = \frac{1}{m}\sum_{j=1}^m \mathcal{I}(Y_j \leq y)$ be the empirical distribution function of Y and $F_{n0}(y) = \frac{1}{n}\sum_{i=1}^n \mathcal{I}(Y_{0i} \leq y)$ be the empirical distribution function of Y_0. Here $\mathcal{I}(\cdot)$ is the indicator function. The natural estimator of the relative CDF $G(r) = F(F_0^{-1}(r))$ in (1.2) is:

$$G_{n,m}(r) = F_m(F_{n0}^{-1}(r)) \qquad 0 < r < 1 \tag{2.5}$$

This estimator can be viewed as a U–statistic with an estimated parameter ($\lambda = F_0^{-1}(r)$ is estimated by $\hat{\lambda} = F_{n0}^{-1}(r)$). Had we directly sampled the relative data we could estimate the PDF using well known one–sample non–parametric methods (see, for example, (Simonoff 1996)). Conceptually, we can considering a two–sample version of the usual one–sample framework by defining the *quasi relative data:* $\{R_1, R_2, \ldots, R_m\}$:

$$R_j = F_{n0}(Y_j) \qquad j = 1, \ldots, m$$

Note that the univariate sample $\{R_j\}_{j=1}^m$ are not independent as they depend on the $\{Y_{0i}\}_{i=1}^n$. However they will be close to uncorrelated (their pairwise correlation

[2] This is the usual situation when the samples are the result of repeated cross–sectional surveys. If the samples are a result of a longitudinal panel survey, they will be correlated. While the statistical development is similar, this case is not considered here.

is $O(n^{-1})$). The relative PDF can then be estimated by using a local–polynomial smoother (e.g., Fan and Gijbels 1996) to a binned version of the quasi relative data. Although the usual results for such estimators do not hold, the asymptotic joint distributions for the PDF and CDF can be derived. These results can be used to calculate confidence intervals and (simultaneous) confidence bands for $G(r)$ and $g(r)$ based on $G_{n,m}(r)$.

Software for the estimation of the relative CDF and PDF will be made available from the `statlib` archive at Carnegie–Mellon University; information on using `statlib` can be obtained by sending the message `send index` to the electronic mail address `statlib@lib.stat.cmu.edu`. In addition, information is available from the author's home pages (i.e., `http:\\www.stat.psu.edu`).

2.3 History and Previous Literature

2.3.1 Comparison to Lorenz Curves

One of the key application areas for relative distribution methods is the study of inequality. Lorenz curves (Lorenz 1905) (and the associated Gini index summary statistic) are the standard method used for inequality comparisons, so it is natural to ask what the relation is between Lorenz curves and the relative distribution.

Lorenz curves represent the outcome attribute in terms of cumulative share: the cumulative fraction of Y that is held by the $p\%$ of the population with lowest values of the attribute. It will coincide with a 45° line joining $(0,0)$ to $(1,1)$, when all individuals in the population receive the same value of of the attribute, and otherwise will curve below it. This curve characterizes the distribution, up to a multiplicative constant, in the sense that two distributions will have the same Lorenz curve if, and only if, they differ by a simple multiplicative factor.

One key difference between the relative distribution and Lorenz curve approaches is therefore the reference unit of measurement. Comparing two Lorenz curves, one can either express the difference in the fraction of Y held by a specific fraction p of two populations, or the difference in the two population fractions that hold a specific fraction of Y. The relative CDF and PDF, by contrast, represent the outcome attribute in terms of quantile or level: population fractions at different *levels* of Y. Using the relative distribution, one can compare the quantiles p_o and p_1 of the two populations at a specific value of the attribute, y.

The fundamental difference between the relative distribution and Lorenz curves is that the former permits direct distributional comparison, while the latter compares distributions indirectly via a third distributional standard. To compare two distributions using Lorenz curves, the Lorenz curve for each is first constructed by reexpressing the observed PDFs in terms of their deviation from exact equality (that is, no variance in the outcome attribute). Comparison between the two distributions is then based on how each Lorenz curve deviates from the distributional standard of exact equality. The relative distribution, by contrast, compares one CDF directly to the other.

The Lorenz curve for a distribution can be thought of as a comparison of the distribution of attribute share to attribute. The Lorenz curve attempts to capture the share of the total attribute for each level of the attribute. It is usually described in terms of income and we will do so here. The distribution of income for a population is the distribution of income for a person randomly chosen from the population of people. Alternatively, consider the the distribution of incomes (from which each of these dollars accrued) for a dollar randomly chosen from the population of dollars in the economy. Just as each person is associated with an income level, each dollar in the economy is associated with the income level that produced it. It is natural to call this the distribution of *level shares* of the total income. It represents the income share for a given level of income, i.e., the likelihood that a given dollar came from a person with that income. The level share distribution is also called the first or incomplete moment distribution. Note that the PDF of the level shares for the distribution Y is $yf(y)/E(Y)$, so that the shape of the level share distribution can be determined from that of the attribute distribution itself.

It is natural to compare the distribution of income shares to that of incomes. The distribution of income shares will always show more dispersion than that of incomes: higher incomes have a greater share of the dollars than of the people. The divergence between the distribution of income shares and incomes measures the strength of the concentration of the total income in the hands of those with high incomes. The two distributions will coincide only if all people earn the same income.

The Lorenz curve is the relative CDF of the distribution of income shares to that of incomes. Thus the Lorenz curve is an attempt to capture the quantile share information for a given distribution. Let R_L be the relative distribution of income shares to that of incomes (we will call it the *Lorenz relative distribution*). Let $G_L(r)$ be the corresponding relative CDF, i.e., the Lorenz curve of incomes. All summary measures based on the Lorenz curve can then be based on this relative distribution and this can sometimes provide simple interpretations of these measures. For example, the Gini index is defined to be twice the area between $G_L(r)$ and the 45^o line (which is the relative CDF corresponding to an egalitarian distribution in which all individuals have equal incomes). This is:

$$\text{Gini}(F) = 2E\Big(R_L\Big) - 1$$

Thus the Gini index is just a rescaled mean of the Lorenz relative distribution. The Lorenz relative density is

$$g_L(r) = F^{-1}(r)/E(Y)$$

and so will be non-decreasing starting from zero. The point where $g_L(r)$ crosses unity is the percentile (r_L) corresponding to the mean income. The Pietra index is defined by $\max_{0\leq r\leq 1}[r - G_L(r)] = r_L - G_L(r_L)$. This can be interpreted as the proportion of total income which would have have to be transferred from incomes above the mean to incomes below the mean to achieve an egalitarian distribution.

Interdistributional comparisons are usually made by comparing Lorenz curves. We have shown that these comparisons are implicitly comparisons of their Lorenz relative distributions. The Lorenz partial order is defined by $Y \leq_L Y_0$ (i.e., Y does not exhibit more inequality in the Lorenz sense than does Y_0), if the Lorenz curve for Y is greater than of equal to the Lorenz curve for Y_0 for each percentile. Thus the Lorenz partial order corresponds to the stochastic ordering of their corresponding Lorenz relative distributions.

A final key difference concerns the level of scale invariance. Lorenz curves are multiplicatively scale invariant, i.e., two distributions will have the same Lorenz curves if, and only if, they differ by a simple multiplicative constant. The same is true of the summary measures based on the Lorenz curves such as the Gini index, as well as other summary measures for inequality such as the Pietra, coefficient of variation and Kakwani indices. All are intrinsically tied to the original scale of measurement, up to a proportionate scale.

The relative distribution, by contrast, is invariant to *all monotonic* transformations of the original measurement scale. The comparison will thus be accurately and equivalently represented by comparisons of the raw attribute, the log–attribute, or any other monotonic transformation of the attribute, as long as there is a common, monotonic underlying utility function in the population. We shall call this the principle of *strong* scale invariance. Whenever this principle holds, the relative distribution plays the primary role in comparisons, in the sense that it contains all the information necessary for comparing distributions, making the minimal assumptions necessary for valid comparison. Holmgren (1995) shows that under appropriate technical conditions the relative distribution is the maximal invariant (loosely speaking, that any other quantity that contains the same information does not satisfy the principle of strong invariance – cf., Lehmann 1983). This does not mean that the relative distribution is inappropriate when the assumptions are not known to hold, only that comparisons may exist that can not be exclusively expressed in terms of the relative distribution, but require additional characteristics of the original distributions.

One approach is to look at measure of the difference in intradistributional inequality, such as the Gini indices. It has been shown by Holmgren (1995) that under appropriate technical conditions that any preference ordering between pairs of distributions can be expressed in terms of preference between their relative distributions. In this sense the relative distribution plays the same role for interdistributional comparisons as the Lorenz curve plays for intradistributional comparisons.

2.3.2 Other related literature

The relative CDF $G(r)$ is implicitly a theoretical p–p plot of F against F_0, an empirical version of which was considered by Wilk & Gnanadesikan (1968). It is the plot $\{(F(x), F_0(x)) : x \in \mathbb{R}\}$ which can be represented in the functional form $\{(r, G(r)) : 0 \leq r \leq 1\}$. (cf. also Chambers, et al. 1983). Holmgren (1995) gives a nice discussion of the merits of the relative CDF (p–p plots) compared to q–q plots. His theoretical justification and results carry over to relative distribution framework. One can also find a review of p–p plots and additional references in this book.

The relative CDF is also essentially equivalent to the receiver operating characteristic (ROC) used in the evaluation of the performance of medical tests for separating two populations (Begg 1991; Cambell 1994). In this context, Hsieh (1995) has used an empirical process approach to describe the properties of ROC curves. The precise relationship between the relative distribution and ROC curves is discussed in Li, Tiwari & Wells (1996).

The relative CDF and PDF has been explicitly studied in at three distinct threads. We will give only the briefest overview here. Parzen (1977; 1992) studies the relative CDF and PDF as part of "comparison change analysis". He refers to them as the "comparison distribution" and "comparison density," respectively. Although, the underling ideas have been floating around statistics for decades, Parzen appears to be the first to systematically study aspects of the relative PDF as a basis for interdistributional comparison. Parzen's students discuss the role of the relative PDF in distributional comparisons (Prihoda 1981), develop kernel density estimation for the relative PDF (Alexander 1989). and statistics based on the relative PDF for comparing distributions (Eubank, et al. 1987). In a separate literature, Cwik and Mielniczuk (1989; 1993) have investigated non–parametric density estimation for the relative PDF. They refer to refer to equation (1.1) as the *grade transformation*, as F_o can be thought of as a grading function. In their terminology the "relative density" is called a *grade density*. They develop a kernel estimate for the relative PDF and show uniform a.s. convergence, and develop a method for choosing an estimate which is appropriately smooth. Gijbels and Mielniczuk (1995) generalize these results (to the Radon–Nikodym derivative) and determine the rates of uniform a.s. convergence. In a third separate literature, Li, Tiwari & Wells (1996) develop the statistical properties of the relative CDF under the name of "inverse quantile density function." They gives references to related work, especially as related to the precise link between the relative CDF and ROC curves. Holmgren's work develops from this literature.

As noted above, closely related quantity to the relative PDF $g(r)$ is that of the *density ratio*:

$$h(x) = \frac{f(x)}{f_0(x)} \qquad\qquad -\infty < x < \infty$$

considered by Silverman (1978). It is a key element of discriminant analysis (Hand 1982) and likelihood–ratio methods. Note that $h(x) = g(F_0(x))$ and $g(r) = h(F_0^{-1}(r))$. Absava and Nadareishvili (1985) study non–parametric estimation of the density ratio.

These threads of research have been motivated by separate substantive research questions and and rarely reference each other. Similar, but more *ad hoc* methods have been developed in the social sciences, motivated by measurement issues for labor market inequality. It is quite likely that parallel developments of these concepts has been implicit in many fields.

2.4 Example

Figure 3 shows the relative CDF and PDF for the distribution of wage growth in the two NLS cohorts.

Figure 3 about here

The first panel represents the relative CDF of the wage growth, recent to original cohort. If the two distributions were identical, the curve would fall along the 45^o line. The differences in wage growth experienced by the two cohorts is easily described using this display using the horizontal and vertical gridlines. At the median of the original cohort wage growth, $r = 0.5$, $G(r) = 0.63$. This means that approximately 63% of the recent cohort experienced lower gains than the median respondent in the original cohort. We also see that 27% of the recent cohorts wage gains are in the bottom decile of the original cohort distribution, and the proportions for the two cohort converge for the upper deciles, becoming coincident above the 90^{th} percentile. The divergence between the two cohorts is greatest in the third decile; approximately 20% more of the recent cohort are below this wage–gain level than the original cohort.

The second panel shows the relative density of the wage growth, recent to original cohort. If the two distributions were identical, the relative density would be uniform. Compared to the relative CDF, the relative density provides a display that is easier to interpret visually: values above 1 represent more density in the comparative distribution, while values below 1 represent less, and the actual value is the fraction more (or less). The wage gains for the recent cohort are clearly shown here to have been concentrated in the bottom of the original cohort distribution. Nearly three times as many recent earners fell into the bottom decile of the wage growth experienced by the original cohort. The lower wage growth appears to have fallen hardest among those who would have experienced modest gains: in the middle of the original cohort's distribution (from the 30th to 80th percentile) there are now roughly a third less workers. At the top wage–growth levels, there is less discrepancy between the two cohorts, although the recent cohort is still somewhat less likely to achieve these gains. For example, the relative density at the 85^{th} percentile of original cohorts' wage growth ($F_0^{-1}(r) = 1.5$) is about equal to 0.80. This means that there are about 20% fewer recent cohort earners who attained this level of wage growth.

The relative density enhances comparison of the two PDFs in several ways.

First, it directly compares their relative frequency in terms of a ratio, which is easier to understand both visually and numerically. In contrast to the direct PDF overlay in Figure 2, which requires the viewer to construct the differences between the two curves at each point on the scale, the relative distribution codes this comparison directly. It provides a simple visual signal for information that is visible but not easy to process in the original PDF overlay. (Chambers, et al. 1983; Cleveland and Mcgill 1984). The graph of the relative density also provides a more intuitive and interpretable display than a comparable graph of the two Lorenz curves. In part this is again because the relative density graph codes the comparative information directly, so the viewer does not have to reconstruct it from the two individual curves. But in part it is also because the density ratio is simply easier for most people to interpret than a comparison of Lorenz curves.

The relative density graph remains close to the original data, allowing the researcher to isolate key characteristics of the individual distributions, as well as the comparison between them. The result is a much more accessible, intuitively meaningful and informative description of the data than that afforded by summary statistics such as the mean, interquartile range, or the Gini index. The real strength of this approach becomes apparent in the following sections, where a flexible set of decomposition techniques are developed that show the relative distribution is a tool for analysis as well as description.

3. Decomposing the relative distribution into location and shape components

To better capture the nature of differences between distributions, it is natural to ask how much of the difference could be explained in terms of a simple location shift. If the comparative distribution is a simple shifted version of the reference distribution, that is, $F(x) = F(x - c)$ or $F(x) = F(x \times c)$ for some constant c, then the comparison between the two distributions can be parsimoniously summarized by this shift. Differences that remain after a location adjustment represent differences in "shape" – a general concept that comprises spread, skew, and other distributional characteristics. In this section we develop a general approach to decomposing the overall relative distribution into two component relative distributions that represent differences in location and shape respectively.

Note that the concept of location is scale dependent. For example, a multiplicative scale change on the original scale represents an simple additive shift on the log–scale. While the relative distribution is scale–invariant, the decomposition developed below is not. The choice of scale is context dependent, and the analyst should choose the scale according to the nature of the data. In the discussion below, we work with an additive location shift, because the data units are differences in log wages. Had the units been wage ratios, it would have been more appropriate to use a multiplicative shift.

Let Y_A denote a random variable describing the comparison population location adjusted to have the same median as the reference population. For an additive shift, we define Y_A to be the random variable $Y - \rho$ where $\rho = \text{median}(Y) - \text{median}(Y_0)$[3]. We shall say that Y_A is Y *location adjusted* or *location matched* to Y_0. The CDF of Y_A can be written as $F_A(y) = F(y - \rho)$. Y_A defines a hypothetical comparison population which has the location (median) of the reference population, but the shape of the comparison population. For example, if we consider the wage growth distributions of the two NLS cohorts, the hypothetical distribution for the recent cohort's wage growth would have the same median as the original cohort and the same shape as the recent cohort.

Generalizing the notation of Section 1.2, let $R \equiv R_0^1 = F_0(Y)$ be the relative distribution of Y to Y_0. We can isolate the effect of the location difference on R_0^1 by comparing Y to its median–shifted version, Y_A. Both distributions have the same shape (that of the comparison distribution), they differ only in location. Let $R_A^1 = F_A(Y)$ be the relative distribution of Y to Y_A. We can interpret R_A^1 as the random variable describing the effect on Y of the location difference from Y_0. Note that R_A^1 will have a uniform distribution when the comparison and reference populations have the same location.

Similarly, we can isolate the effect of the shape difference by comparing Y_A to Y_0. These distributions have the same location, but retain the shapes of the comparison and reference distributions respectively. Let $R_0^A = F_0(Y_A)$ be the relative distribution of Y_A to Y_0. This random variable describes the effect on Y of the shape changes from Y_0, or equivalently, would define the relative distribution if the two populations had the same location. Note that R_0^A will have a uniform distribution

[3] Alternative methods for location matching can be used (e.g., multiplicative median shift, mean shifting). In that case the development given below is the same, with the alternative Y_A replacing the additive median shifted version used here.

when the *only* difference between the two distributions is a location shift.

These two effects form an exact decomposition of the relative distribution of Y to Y_0 in the sense that R_0^A is the relative distribution of R_0^1 to R_A^1. The decomposition can also be represented in terms of the density ratios from equation 2.4:

$$\frac{f(y_r)}{f_0(y_r)} = \frac{f(y_r)}{f_A(y_r)} \times \frac{f_A(y_r)}{f_0(y_r)} \tag{3.1}$$

or, in more heuristic terms:

$$\begin{array}{c}\text{overall relative}\\ \text{density}\end{array} = \begin{array}{c}\text{density ratio for}\\ \text{the location difference}\end{array} \times \begin{array}{c}\text{density ratio for}\\ \text{the shape difference}\end{array} \tag{3.2}$$

The density ratio for the shape effect is a proper relative density (i.e., it sums to 1). The density ratio for the location effect in general is not, because of the scale change imposed by using f_A rather than f_0 as the reference distribution for R_A^1. The density ratio preserves the cut–points, y_r, so that the location and shape effects are applied at the same value of y_r.

As the relative density can be displayed graphically, so can the terms of the decomposition, and these provide a very useful set of visual representations. Denote the densities of R_0^1, R_0^A and R_A^1 by g_0^1, g_0^A, and g_A^1, respectively. To adjust for the rescaling, denote the CDF of R_0^A by $F_0^A(r) = F_A(F_0^{-1}(r))$, $0 \leq r \leq 1$. Mathematically the relationship between the densities is:

$$g_0^1(r) = g_A^1(p) \times g_0^A(r) \qquad \text{where} \quad p = F_0^A(r), \quad 0 \leq r \leq 1 \tag{3.3}$$

Note that r is the percentile in the reference population for a given value of the attribute, y_r, and p is the percentile in the location–adjusted comparison population at that same value. By comparing plots of g_0^1, g_0^A, and g_A^1 side–by–side we can gauge the relative size and nature of the components.

We postpone an example until the end of the following section.

4. Summary measures

4.1 Conceptual issues

While graphical displays are a key part of the relative distribution framework, summary measures remain an important tool for the comparison of distributional change. A good summary statistic makes it possible to provide a simple and precise answer to a substantive question such as “has inequality in wage profiles grown significantly over the past 20 years?” or “has the upgrading in wage–gains been matched or exceed by the downgrading?” Several summary measures are currently available for comparing aspects of distributional shape, e.g., the Gini index, the Theil index and the coefficient of variation. The key challenge for such measures, however, is to summarize the right thing. As the “right thing” depends on the specific application, it would be useful to have a *framework* for developing summary measures, rather than a one–size–fits–all single measure. The relative distribution provides such a framework, and can be used as the basis for defining a wide and flexible range of summary measures. The generality of this framework for summary measure development is due to the fact that the relative distribution effectively captures *all* of the information that is necessary and sufficient for strongly scale–invariant comparison of distributions.

Two summary measures are developed here. The first provides an answer to the question, “How much does this component contribute to the difference between the two distributions,” and it plays a role similar to the partial R^2 in traditional linear modeling. The second provides an answer to a more specific shape–related question: “to what extent does the shape difference between the two distributions take the form of rising (or declining) inequality?” This statistic plays the same role for the shape component of the relative distribution as the Gini index plays for the Lorenz curve.

Summary measures based on the relative distribution are robust to both outliers and to deviations from assumptions. This robustness follows from two properties of the relative distribution: the rescaling of the comparison distribution to the reference distribution, and the absence of parametric assumptions. Outliers in either the reference or comparison distribution are not necessarily outliers in terms of the relative distribution. The rescaling maps the original units of both distributions to a rank measure (i.e., [0, 1]) moderating the influence of outliers. As a result, summary measures based on the relative distribution are less likely to be influenced by problem cases. The relative distribution, as well as the decomposition techniques, and natural summary measures in this framework are also fully non–parametric. They require minimal assumptions about the underlying distributions – either in terms of the individual distributions, or in terms of their relationship to one another. This actually distinguishes the relative distribution methods from other non–parametric approaches, most of which implicitly assume that the reference and comparison distributions have a well defined relationship to each other (e.g., are simply location shifted versions of each other) (Lehmann 1975).

4.2 Effect summary statistics

Summarizing the effect of location and shape changes on the overall relative distribution requires first a measure of the overall distributional difference. Perhaps the most commonly used measure of the divergence between two distributions is the *Kullback–Leiber divergence* defined by:

$$D(F; F_0) = \int_{-\infty}^{\infty} \log\left(\frac{f(x)}{f_0(x)}\right) dF(x)$$

$D(F; F_0)$ is also known as the information number, discrimination function, and "distance." While there are a number of other measures that could be used here (e.g., the χ^2 divergence and the L_1 distance), the Kullback–Leiber divergence has a simple interpretation in terms of the relative distribution, and it is decomposable into the location, shape and other components of interest here.

Kullback and Leiber (1951) motivated $D(F; F_0)$ in the following way. Suppose we know a single observation q is drawn from either F or F_0 and we wish to decide which. The log–Bayes factor for this problem is:

$$\log\left(\frac{f(x)}{f_0(x)}\right) = \log\left(\frac{P(F|x)}{P(F_0|x)}\right) - \log\left(\frac{P(F)}{P(F_0)}\right)$$

where $P(F|x), P(F_0|x)$ are the posterior probabilities of the hypotheses and $P(F)$, $P(F_0)$ are the corresponding prior probabilities. Now $D(F; F_0) = E_R[\log\left(\frac{f(x)}{f_0(x)}\right)]$ is just the expected information in a single observation from F for discriminating between the two hypotheses. Good (1950) referred to the log–Bayes factor as the "weight of evidence" for F over F_0. If the two distributions are similar, then a single observation will not provide a large amount of evidence for discriminating between them. The magnitude of $D(F; F_0)$ increases with the ease of discrimination. There are a number of alternative motivations for $D(F;F_0)$ that are context dependent; the appropriateness of it depends on the particular question asked, although it is often used as an omnibus measure of divergence (cf. also Akaike 1973; McCulloch 1989; Soofi 1994).

The Kullback–Leiber divergence and the relative distribution are closely linked (Mielniczuk 1992):

$$D(F;F_0) = \int_0^1 \log\Big(g(r)\Big) g(r)\ dr$$

Thus $D(F;F_0)$ can be rexpressed purely in terms of the relative density and with out separate reference to the underlying distributions. The expression on the right–hand side of the equation is just the (differential) *negative entropy* of the relative density (Shannon 1948). The entropy is also a widely used measure of the dispersion of a distribution (Theil and Laitinen 1980). In our context we can interpret $D(F;F_0)$ as the expected information for discriminating g from a uniform distribution based on a single observation from R.

The Kullback–Leiber divergence can be used to summarize the contributions of the location and shape effects to the overall difference between the two distributions. The most direct way is to compare the entropies of the three components (i.e., $D(F;F_0), D(F;F_A)$ and $D(F_A;F_0)$). Given the scale change in the location component, these entropies will not decompose directly. We can, however, use (3.3) directly to show

$$D(F;F_0) \quad = \quad D(F;F_A) \quad + \quad D_Y(F_A;F_0) \tag{4.1}$$

where

$$D_Y(F_A;F_0) = \int_0^1 \log\Big(g_0^A(r)\Big) g(r)\ dr$$

is a cross–entropy interpretable as the expected information for discriminating g_0^A from a uniform distribution based on a single observation from R. The relative sizes of these terms indicate the relative contributions of location and shape to the overall difference between the distributions.

Confidence intervals and p–values for these divergence measures can easily be determined by bootstrapping the comparative and reference distributions directly. We will give examples of this in the next sections.

4.3 Measuring Polarization

In this section we develop a measure of interdistributional inequality that isolates the shape component in the relative distribution (R_0^A) and measures the degree of polarization in this component. Polarization is of particular interest in the study of inequality, because it captures a discrepancy in outcomes that is hidden when only trends in averages are examined. While visual inspection of the original PDF overlay and the location–matched relative density is usually enough to determine whether polarization is occurring, quantifying and statistically comparing the observed patterns is desirable. The polarization index defined here and its decomposition provide a flexible and sensitive method for measuring the relative density in the center or tails of the distribution. It plays the same role as the difference in Gini coefficients, coefficients of variation, or variances of log–values in measuring interdistributional inequality (Grove and Hannum 1986), but without the ambiguity caused by crossing Lorenz curves.

Ideally, we would like a statistic that measures the deviations of the location–adjusted relative distribution from the uniform distribution (i.e., no divergence) and that emphasizes the deviations in both the upper and lower tails. Define the *median relative polarization index* (MRP) of Y relative to Y_0 as:

$$MRP(F;\ F_0) \quad = \quad 4\int_0^1 \mid r - \frac{1}{2} \mid g_0^A(r)\ dr \quad - \quad 1,$$

This measure weighs the mass in the upper and lower tails more heavily than the mass in the center. It is the mean absolute deviation about the median of the location–matched relative distribution, scaled to produce an index that varies between -1 and 1. Positive values represent more polarization, i.e., increases in the tails of the distribution, and negative values represent less polarization, i.e., convergence towards the center of the distribution. If the location–matched relative distribution is uniform the index will be zero. The MRP has a number of useful characteristics. For example, it has a natural symmetry property: the MRP of Y_0 to Y is equal to the negative of the MRP of Y to Y_0. It is also invariant to monotone transformations of the distributions: Let $h(\cdot)$ be a monotone function on the support of Y_0. Then the MRP of $h(Y)$ to $h(Y_0)$ is equal to the MRP of Y to Y_0.

The MRP is decomposable into the contributions made by components above and below the median of the relative distribution. We define the lower polarization index (LRP) by:

$$LRP(F;\ F_0) \quad = \quad 8\int_0^{\frac{1}{2}} \mid r - \frac{1}{2} \mid g_0^A(r)\ dr \quad - \quad 1$$

The upper polarization index (URP) is defined in a similar way:

$$URP(F;\ F_0) \quad = \quad 8\int_{\frac{1}{2}}^{1} \mid r - \frac{1}{2} \mid g_0^A(r)\ dr \quad - \quad 1$$

These indices decompose the MRP in the sense that:

$$MRP(F;\ F_0) \quad = \quad \frac{1}{2}LRP(F;\ F_0) \quad + \quad \frac{1}{2}URP(F;\ F_0)$$

The upper and lower indices have properties similar to the MRP. The scaling produces upper and lower indices that vary between -1 and 1, positive values representing more polarization, i.e., increases in the tail of the distribution, and negative values representing less polarization, i.e., convergence towards the center of the distribution. If the location–matched relative distribution is uniform below (above) the median then the lower (upper) index will be zero.

Inference can be obtained for the estimates of these indices. As above, we will investigate the properties of the natural estimator of $MRP(F;\ F_0)$:

$$\widehat{MRP(F;\ F_0)} \equiv MRP(F_m;\ F_{n0})$$

Define the *location–matched quasi relative data* $\{Q_1, Q_2, \ldots, Q_m\}$ by:

$$Q_j = F_{n0}(Y_j - \hat{\rho}) \qquad j = 1, \ldots, m$$

where $\hat{\rho} = \text{median}(F_m) - \text{median}(F_{n0})$ is the natural estimate of ρ. The estimate has the following appealing expression:

$$\widehat{MRP(F;\ F_0)} \quad = \quad \frac{4}{m}\sum_{j=1}^{m}\left|Q_j - \frac{1}{2}\right| \quad - \quad 1$$

This expression follows from the definitions of the empirical distribution function and some algebra. Note that the $\{Q_j\}_{j=1}^m$ are not independent as they depend on the $\{Y_{0i}\}_{i=1}^n$.

The natural estimators of LRP and URP are $\widehat{LRP(F; F_0)} \equiv LRP(F_m;\ F_{n0})$ and $\widehat{URP(F;\ F_0)} \equiv URP(F_m;\ F_{n0})$, respectively. They can be also be easily expressed in terms of the quasi–data:

$$\widehat{LRP(F;\ F_0)} = \frac{8}{m}\sum_{j=1}^{m}\left|Q_j - \frac{1}{2}\right|\mathcal{I}(Q_j \leq \frac{1}{2}) - 1$$

$$\widehat{URP(F;\ F_0)} = \frac{8}{m}\sum_{j=1}^{m}\left|Q_j - \frac{1}{2}\right|\mathcal{I}(Q_j > \frac{1}{2}) - 1$$

Often we would like to test if the median, upper or lower relative polarization indices in a given situation is statistically significantly different from zero. If the sample size is not small (i.e., $n, m > 30$), we can use the Gaussian approximation to the exact distributions of the estimate as the basis for a test. If the sample sizes are small the bootstrap can be used to determine the sampling distribution of the estimate and the corresponding critical values. We will not discuss these procedures in detail here. The reader is referred to Handcock & Janssen (1996b) where the statistical properties of these estimators for the polarization indices are developed, and confidence and significance bounds for $MRP(F;\ F_0), URP(F;\ F_0)$ and $LRP(F;\ F_0)$ are determined.

4.4 Example

Figure 4 presents the graphical display of the median and shape decomposition as applied to the relative distribution of wage gains. The first panel represents the overall relative density (and is the same as Figure 3b). The second panel represents the effect of the median shift in the wage gains between the two cohorts. These effects are quite large, and suggest that if distributional shape had remained constant, the lower median wage growth in the recent cohort would have led to more than 60% of their wage gains falling in the bottom half of the original cohort distribution, and only about half as many falling in the upper two deciles. Note, however, that the dramatic growth in the bottom decile is not exclusively due to the median downshift, as the bottom decile of panel (b) is about 1.8, well below the value of 2.7 observed in the actual data. The third panel is the location–matched relative distribution. It isolates the changes in the distributional shape of the relative wage gains between the two cohorts, and these effects are also striking. Even without the lower median, the greater dispersion of wage gains in the recent cohort would have led to relatively more low–growth earners, though this effect is confined to the bottom decile. The polarization hollows out the middle of the wage gain distribution, with a cumulative loss of nearly a third of recent earners in deciles 3 through 8. At the top of the distribution, however, the widening gap in wage gains works in the opposite direction from the location shift: operating by itself, it would have increased by over half the number of wage gains in the upper decile.

Figure 4 about here

The Kullback–Leiber divergences (entropy) are given on top of each figure. The overall dispersion between the wage growth of the two cohorts is 0.156 while the divergence due to the location difference is 0.066. Thus, overall, the location shift explains 42% of the distributional change observed. Based on the decomposition (4.1), the contribution of the shape difference in wage growth between the two cohorts is 0.090, or 58% of the total change. Confidence intervals and p–values for these divergence measures are summarized in Table 1.

Table 1 about here

The MRP index for the shape change displayed in panel (c) is 0.183 (95% CI 0.148 – 0.218), and is significant well below the $p = 0.05$ level. For comparison, two Gaussian distributions with the same MRP would have a standard deviation ratio of 1.34. The size and sign of the estimate confirm the impression left by the graphical display: there has been a significant growth in permanent wage inequality between the two cohorts. The full set of polarization indices is shown in Table 2.

Table 2 about here

The estimated lower and upper polarization indices for indicate that both the upper and lower parts of the distribution are significantly positively polarized. The lower index is slightly larger, indicating that there was greater polarization in the lower tail of the distribution than in the upper tail.

5. Covariate adjustment

In this section we present a method for adjusting the relative distribution for changes in the distribution of other covariates. This makes it possible to examine such questions as, "How would the wage–gain distributions have looked if the age distribution had remained constant?" or "How did median and shape changes combine to produce the changing returns to education?" For continuous covariates,the goal is to separate the relative distribution into a component that represents changes in the covariate profile, and a component that represents the residual changes. These residual changes are taken from the conditional distribution of the outcome attribute given the covariate level, and in one sense can be considered changes in the "returns" to that covariate. The method is similar to that used in the previous section: an adjusted population is created that matches the covariate distribution of the reference group, and using the conditional attribute distribution of the comparison group. Relative distributions of the adjusted population to the reference and comparison populations then isolate the composition and residual effects respectively. In addition, the composition and residual effects can be further decomposed into median and shape changes. We note here that this approach does *not* provide a single estimate of the "effect" of the covariate. This is because in the fully distributional setting there is not a single effect, such as the mean wage gain per additional year of education, but rather a conditional distribution of returns at each covariate level. When the covariate is discrete, one can also define a simple additive version of the decomposition that provides a useful summary of the distributional effects.

Where the methods in this section are similar in principle to those presented in previous section, they are not repeated in detail here.

5.1 Compositional adjustment

The age profile for the 14–22 year old population changed from 1966 to 1979, so the age composition of the two NLS cohorts is somewhat different. We can use the covariate adjustment technique to determine whether the differences in the age composition between the two cohorts explains some of the change we observe in relative wage gains.

Let Y_A be the adjusted random variable that describes the wage growth of a population with the marginal age distribution of the original cohort and the conditional wage gain distributions from the recent cohort. Y_A can be though of as the hypothetical wage gain distribution that would have prevailed in the recent cohort if it had retained the same age profile as the original cohort. [4]

We can now determine the compositional effects of the covariate (age) by comparing Y to Y_A. Both distributions are defined to have the same conditional distributions of wage gains given age, so all differences between the distributions are due to differences in the marginal age profiles. We define $R_A^1 = F_A(Y)$ to be the relative distribution of Y to Y_A. In the example here, we can interpret R_A^1 as the random variable describing the effect on wage gains of changes in the age distribution between the two cohorts. It isolates the direct compositional effect of age: the relative distribution if only the age profile of the two populations had changed. Note that R_A^1 will have a uniform distribution when marginal distribution of the covariate is the same in the two distributions; in this example, when the relative distribution of ages in 1966 and 1979 is uniform.

Having isolated the composition effect, we can examine what remains after adjusting for it by comparing Y_A to Y_0. Both distributions now have the same covariate profile, so all differences between them are due to changes in the conditional distribution of of the outcome attribute. We define $R_0^A = F_0(Y_A)$ to be the relative distribution of Y_A to Y_0, and can interpret it in this example as the random variable describing the age–adjusted relative distribution of wage gains. It represents the effect of changes in the distribution of age–specific wage gains between the two cohorts – changes in the "returns" to age. R_0^A will have a uniform distribution when the only difference between the two distributions is the covariate composition. Here, a uniform R_0^A would indicate that net of age–profile shifts, the wage gains are the same for both cohorts.

These two effects form a decomposition of the relative distribution of Y to Y_0 in the same sense as the median/shape decomposition above: R_0^A is the relative distribution of R_0^1 to R_A^1. The graphical display can be obtained by plotting the equivalent terms in 3.3. Here the first term in the decomposition represents the compositional effect of the covariate, and the second represents the residual changes.

5.2 Example

Figure 5 displays the relative distribution of age in the two cohorts. The recent

[4] Note that we can also define Y_A to have the marginal age distribution of the recent cohort and the conditional wage gains of the original. This changes the order of the comparisons below, and slightly changes the interpretation of the effects.

cohort has about 40% more respondents in the lowest age (14 at enrollment, 29 at the end of the observation period), and 20 to 30% more respondents in the upper ages (21–22 at enrollment, 35–36 at end). As a result, there are 10–20% fewer respondents in the middle of the age range.

Figure 5 about here

Figure 6 graphically represents the adjustment of the relative distribution for age composition changes (3.3). Panel (a) is the (unadjusted) relative density of wage gains that is being decomposed (same as Figure 3b). Panel (b) represents the age composition effects – that is, the component of (a) that is attributable to the effect of changes in the relative distribution of age. Panel (c) then represents the *age–adjusted* relative density of wage gains – that is, the expected relative density of wage gains had the age profiles of the two cohorts been identical.

Figure 6 about here

Figure 6(b) shows something very close to a uniform distribution. The implication is that the difference in age composition between the two cohorts had very little effect on the observed relative distribution of wage growth. There is a small but significant drop in the lower decile, however, indicating that age composition differences would be expected to produce a 20% drop in the fraction of low wage–gain workers in the recent cohort. Recall from Figure 5 that the recent cohort had relatively fewer cases in the middle of the age range. These groups appear to have experienced somewhat fewer low wage gains.

Figure 6(c) represents the age–adjusted relative wage gain distribution. Given the absence of major compositional effects, the adjusted distribution is not much different than the original distribution. The graphical perception is confirmed by the decomposition of the Kullback–Leiber divergence. The divergence due to the change in age composition is 0.001, only 1% of the overall divergence between the two cohorts. This divergence is not statistically significantly different from zero. One can interpret this as showing that the changing "returns" to age, rather than the changing age composition of the population, was the dominant cohort effect. Confidence intervals and p–values for all the Kullback–Leiber divergences are given in Table 3.

Table 3 about here

Once the relative density has been covariate adjusted, one can examine both the composition and residual components for median and shape changes. As a hypothetical example, suppose the age composition effect in Figure 6b had been large. A median/shape analysis could then go on to show that the composition effect was primarily a median shift, while the residual became more polarized. This kind of analysis can provide a rich description of the interrelated distributional changes, which in turn can help to inform and focus a theoretical debate by clearly identifying what needs to be explained. Combining covariate adjustment with median/shape decomposition is straight–forward, and given the absence of any significant effect found in the age–adjustment analysis above, we will not pursue this example further.

5.3 Discrete Level Contrasts

When the covariate is discrete, adjustment can also proceed as described above. In this context, however, it is often of interest to compare the groups defined by the covariate directly, rather than treating the covariate as a control variable and adjusting to eliminate its compositional effects. For example, if gender were the covariate, we might be interested in comparing the changes in the relative distributions of wage gains *by gender* over time, rather than adjusting the wage–gain distribution for the rising fraction of women in the workforce. While adjustment identifies the impact of women's growing labor force participation on the distribution of wage gains, the group comparison approach makes it possible to analyze the changes within and between the women's and men's wage gain distributions (an extended example can be found in Bernhardt, et al. 1995). In this section we will work through an example, using educational attainment as our discrete covariate.

In the literature on the growth in cross–sectional inequality, a consistent finding is that the wage premium for a college education has risen substantially (Juhn, et al. 1993; Katz and Murphy 1992). While the college educated are not doing uniformly better than they had in previous cohorts, those with low education are doing relatively worse on almost all measures – wages, job stability (Farber 1996a), benefits (Farber 1996b), and employment (DiPrete 1993). A natural question to ask is whether this penalty can also be found in wage growth profiles, and what kinds of location and shape shifts are at work.

We will define two educational groups and compare them here: the low education group is defined as those with at most a high–school diploma, the high education group as those completing one or more years of college education. Figure 7 compares the distributions of wage gains for these two education groups, as density overlays (a and c) and as relative densities, recent to original cohort (b and d). Panels (a) and (b) compare the wage gains for the high school educated across the two cohorts. The down shifting of wage gains for the recent cohort is quite apparent. Three times as many earners in the recent cohort experience wage gains that would have put them in the bottom decile of the original cohort, and there are 20 – 50% fewer wage gains in any of the higher deciles. The relative distribution for this group is dominated by the location shift. Panels (c) and (d) are the corresponding distributions for the college educated. For this group, the change between the two cohorts is less pronounced, and takes a different form. The relative frequency of both low and high wage gains increases for the recent cohort, though low wage gains still predominate. About twice as many recent wage gains fell in the bottom decile of the original distribution, but the fraction falling the highest decile rose by nearly 20%. Overall, the relative distribution for the college educated exhibits modest polarization and little evidence of a location shift.

Figure 7 about here

The summary statistics for these patterns are presented in Table 4. The median ratio shows a 23% loss in real wage gains for the high–school group, while the college group held steady. The entropy summary suggests that the overall change experienced by the high–school group was three to four times as large as that experienced by the college group. And while the median shift explained about 60% of the total change for the high–school group, virtually all of the change for the college group was due to changes in distributional shape. For both groups, the shape change took the form of growing inequality – as the MRP is significantly greater than 0 – and the polarization is greater in the lower tail of the distribution. It is notable, however, that for the college group, the polarization in the lower tail was much more extreme, as the index for the lower tail is about four times as large as that for the upper. This pattern is visible in the relative density panel in Figure 7d. The median shift for this group is so small that median–adjustment would have a negligible effect on this graph, so the panel is effectively displaying the shape shift.

Table 4 about here

To compare the two groups directly, we can form the relative distributions of high school to college educated wage gains, and compare these between cohorts. These patterns are summarized in Table 5 below. The second and third columns of the table represent the relative deciles (high school:college) for the original and recent cohorts respectively. The last column shows the difference, and represents how the high school group fared relative to the college group from one cohort to the next. For example, 19% of the high school wage gains fell into the bottom decile of the college group distribution in the original cohort, while in the recent cohort this rose to 27%, for a gain of nearly 8%. The relative fraction of the wage gains for the high school educated increased in every decile below the median, and decreased in every decile above.

Table 5 about here

5.4 Additive Decomposition

What impact did the relative median and shape shifts have on these changes in relative position? That is, what percent of workers were moved into or out of a decile by the changes in the median and shape of the wage gain distributions? A natural way to answer this question would be to compare the observed changes in column 3 to the changes that would have occurred if only the medians (or shapes) had changed from the original cohort. [5] This suggests a decomposition into the "marginal effects" of each change. Let H_s^m and C_s^m denote the distribution of wage gains for the high school and college groups respectively, with the median adjusted to the m^{th} cohort, and the shape (or conditional distribution of returns) from the s^{th} cohort. For example, H_o^r is H_o^o adjusted to have the same median as the high school wage gain distribution of the recent cohort. Let $g(H : C)$ denote the relative density of wage gains for high school to college educated workers. Then the marginal effects of

[5] The additive decompositions presented here follow the spirit of that presented in Bernhardt (1995), but differ in the details.

the median shift from the original relative density can be defined as

$$g(H_o^r : C_o^r) \; - \; g(H_o^o : C_o^o)$$

the marginal effect of the shape change in the high school distribution as:

$$g(H_r^o : C_o^o) \; - \; g(H_o^o : C_o^o)$$

and the marginal effect of the shape change in the college distribution as:

$$g(H_o^o : C_r^o) \; - \; g(H_o^o : C_o^o) \tag{5.1}$$

Each of these effects compares the original relative density, $g(H_o^o : C_o^o)$, to the hypothetical density that would have been produced by a change in the specific distributional component alone.

The effects do not sum to the total difference shown in the third column of Table 5 because they do not occur independently. The difference between the sum of these effects and the total change can be interpreted as an interaction effect. The effect of each distributional change depends on the others: if the median shift has moved a substantial fraction of workers out of a decile, then the shape effects will be operating on a smaller base, and will move a correspondingly smaller fraction of workers out of that decile (or into another) than they would have in the absence of a median effect. This interaction among the effects makes a unique decomposition of the total change into the three components problematic, unless one is willing to specify the order in which the effects are applied. The principle is the same as that involved in decomposing the explained variance in the linear model context when the covariates are correlated. Here, as there, it is possible to define sequential effects that uniquely decompose the sum, but, in general, the order of the sequence will matter.

For comparison, we will obtain an exhaustive decomposition by defining the effects sequentially, specifying first the relative median shift:

$$g(H_o^r : C_o^r) \; - \; g(H_o^o : C_o^o)$$

then the shape change in the college distribution:

$$g(H_o^r : C_r^r) \; - \; g(H_o^r : C_o^r)$$

and the shape change in the high school distribution:

$$g(H_r^r : C_r^r) \; - \; g(H_o^r : C_r^r) \tag{5.2}$$

These effects do sum to the total change shown in Table 5. The sequential order seems reasonable in this case: median effects are in some sense more fundamental than shape effects, and changes in the reference distribution (the college educated here) can be seen as more fundamental than changes in the comparison distribution.

Figure 8 presents the two decompositions side by side. Panel (a) represents the marginal effects as defined by (5.1), and panel (b) the sequential effects as defined by (5.2)[6]. The solid bars show the total change from Table 5, and each of the lines represents one of the three components in the decomposition.

Figure 8 about here

Qualitatively, the two decompositions show the same picture. The relative median shift tends to be the largest and most consistent contributor to the overall pattern, but its effect is clearly modified in some cases by the shape changes. Particularly in the bottom decile, the relative median shift alone would have produced nearly double the number of low wage gains for the high school educated, but the strong polarization of the lower tail of the college group's distribution virtually nullified that shift. The net result looks much like the effect due to the shape shift in the high school distribution: 5–7% more high school educated workers in the lowest decile of the college earners' wage gains. Most of the growth in the deciles below the median appears to have been generated by the college shape change: as college earners moved out of this section of the wage gain distribution, the relative fraction of high school earners increased. For the upper deciles most of the losses appear to be driven by the median shift. Here the shape effects for both groups are relatively smaller, and sometimes work in the opposite direction from the median. Recalling the summary polarization indices shown in Table 2, we can see that the modest polarization in the upper tail of the high school educated distribution, and the lack of polarization in the upper tail of the college educated distribution, helped to moderate the high school group's losses in the upper deciles.

[6] The interaction effect is not plotted in the first panel, but it is quite large in some deciles: { -5.0, -1.1, 10.9, -7.6, 0.2, 3.6, 1.5, -3.1, -1,2, and 1.6 }

6. Discussion

The relative distribution is a general non–parametric framework for the analysis of distributional difference and change. It builds on earlier techniques such as p–p plots and comparison densities, and expands the central insight of these techniques into a comprehensive and flexible framework for distributional analysis. The relative distribution can be used as the basis for exploratory, descriptive and analytical techniques. With the increase in information and complexity that accompanies the shift from simple means–based to fully distributional analyses, visualizing the data is an important aid to understanding. The graph of the relative density provides a powerful tool here; identifying differences between distributions, remaining close to the original data and retaining a simple intuitive meaning. The effects of location shape and covariate changes can be explored using these graphs, and summarized using the Kullback–Leiber divergence (entropy) and polarization indices. Entropy measures play a role similar to partial R^2 measures, and the polarization indices play the role of the Gini coefficient in the relative distribution context. Statistical inference for the relative PDF, CDF, entropy and polarization indices make it possible to formalize quantitative comparisons. Covariate decomposition techniques can be used to measure and adjust for the marginal impact of changes in population composition, or to create discrete level contrasts. Taken together, the integration of these different components into a single analytic framework provides a powerful approach to the analysis of full distributional information.

While relative distribution methods have much to offer, they have some limitations that should be kept in mind. Relative distribution methods are not for small data sets. While estimation requires a minimum of 20 observations, realistically, the displays and methods are not well behaved with less than 100 observations, and the decomposition techniques become fully functional with 1000 or more. This is the tradeoff for the absence of parametric assumptions: full distributional information requires data support for each quantile. With small–to–moderate data sets the variation swamps the distributional information, so the uncertainty of the distributional estimates make interpretation difficult. With more traditional parametric methods, we trade off uncertainty about the distribution for bias in the way the parametric distribution represents the distribution. For example, when we use means and variances to summarize the distribution, the implicit assumption is that these two parameters capture all of the information in the distribution – which is equivalent to the assumption of normality. Parameter estimates based on small samples can be grossly misleading if the actual distribution is far from normal.

These methods are robust to many of the problems that haunt other methods, like problem cases and parametric assumptions, but they are not robust to others, such as the common data problem of "heaping." The heaping problem arises in the survey context when respondents report in round numbers rather than exact values. Classic examples can be found in self–reported data on income, age, and lifetime numbers of sex partners (Handcock, et al. 1994; Heitjan and Rubin 1990; Morris 1993). Heaping can fundamentally change the quantile characteristics of a distribution, and the relative distribution graphical techniques in particular can be quite sensitive to this. Ironically, for all their non–robustness means and mean–based statistics are in fact quite robust to heaping.

Full distributional information can also become overwhelming in the context of multivariate decomposition. This again is the price one pays for not assuming that the conditional mean and variance provide an adequate summary of the relationships of interest. As noted above, summary measures based on the relative distribution can be developed for multivariate analyses. These measures need not be used blindly, as the graphical displays of the relative distribution extend to all forms of the covariate decomposition. Familiar summaries of covariate effects like regression coefficients, however, are notably absent in this approach. It is possible that links to quantile regression (Buchinsky 1994) will provide such summaries, and this is a topic for future research.

References

Absava, R. M. and Nadareishvili, M. M. (1985) Nonparametric Estimation Of Density Distribution Of Two Unknown Dimensions In D-Dimensional Euclidean Space (Russian) (Stma V27 947), *Sakharthvelos Ssr Mecnierebatha Akademis Moambe*, **117**, 257-260.

Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle, in *2nd International Symposium on Information Theory*, B. N. Petrov and F. Csaki (Ed.), 267-281.

Alexander, W. P. (1989) *Boundary Kernel Estimation Of The Two Sample Comparison Density Function* , Department Of Statistics, Texas A&M University.

Atkinson, A., et al. (1992) Empirical studies of earnings mobility, in *Fundamentals of Pure and Applied Economics*, J. Lesourne and H. Sonnenshein (Ed.).

Begg, C. B. (1991) Advances In Statistical Methodology For Diagnostic Medicine In The 1980's, *Statistics In Medicine* , **10**, 1887-1895.

Bernhardt, A. D., et al. (1995) Women's Gains Or Men's Losses? A Closer Look At The Shrinking Gender Gap In Earnings, *American Journal of Sociology*, **101**, 302-328.

Buchinsky, M. (1994) Changes in the US Wage structure 1963-1987: an application of quantile regression, *Econometrica*, **62**, 405-458.

Cambell, G. (1994) Advances In Statistical Methodology For Evaluation Of Diagnostic And Laboratory Tests, *Statistics In Medicine*, **13**, 499-508.

Cappelli, P. (1994) The Effect Of Restructuring On Employees, in *Looking Ahead: The Restructuring Of Employment*, ?? (Ed.), .

Chambers, J. M., et al. (1983) *Graphical Methods For Data Analysis*, Wadsworth, CA.

Cleveland, W. S. and Mcgill, R. (1984) Graphical Perception: Theory, Experimentation, And Application To The Development Of Graphical Methods, *Journal Of The American Statistical Association*, **79**, 531-554.

Cwik, J. and Mielniczuk, J. (1989) Estimating Density Ratio With Application To Discriminant Analysis, *Communications In Statistics*, **18**, 3057-3069.

Cwik, J. and Mielniczuk, J. (1993) Data-Dependent Bandwidth Choice For A Grade Density Kernel Estimate, *Statistics and Probability Letters*, **16**, 397-405.

Danziger, S. and Gottschalk, J. (1993) *Uneven Tides: Rising Inequality In America* , Russell Sage Foundation, New York.

Danziger, S. and Gottschalk, J. (1996) *America Unequal* , Russell Sage Foundation, New York.

DiPrete, T. (1993) Industrial restructuring and the mobility response of American workers in the 1980s, *Am Soc Rev*, **58**, 74-96.

Eubank, R. L., et al. (1987) Test Statistics Derived As Components Of Pearson's Phi-Squared Distance Measure, *Journal Of The American Statistical Association*, **82**, 816-825.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling And Its Applications* , Chapman And Hall, London.

Farber, H. S. (1996a) *The changing face of job loss in the united states: 1981-93* , Industrial Relations Section, Princeton University.

Farber, H. S. (1996b) *Job creation in the United States: Good or bad?* , Industrial Relations Section, Princeton University.

Gijbels, I. and Mielniczuk, J. (1995) Asymptotic Properties Of Kernel Estimators Of The Radon-Nikodym Derivative With Applications To Discriminant Analysis, *Statistica Sinica*, **5**, 261-278.

Good, I. J. (1950) *Probability and the Weighting of Evidence* , Griffin, London.

Gottschalk, P. and Moffit, R. (1994) The growth of earnings instability in the US labor market, *Brookings Papers on Economic Activity*, **2**, 217-272.

Grove, D. J. and Hannum, R. (1986) On Measuring Intergroup Inequality, *Sociological Methods And Research*, **15**, 142-159.

Haider, S. J. (1997) Earnings Instability and Earnings Inequality in the United States: 1967-1991, *Journal of Labor Economics*, **(forthcoming)**, .

Hand, D. J. (1982) *Kernel Discriminant Analysis* , Wiley, New York.

Handcock, M. S. and Janssen, P. (1996a) *Statistical Inference For The Relative Distribution* , Department of Statistics and Operations Research, New York University.

Handcock, M. S. and Janssen, P. (1996b) *Statistical Properties Of The Relative Polarization Indices* , Department of Statistics and Operations Research, New York University.

Handcock, M. S., et al. (1994) *Imputation methods for coarse income data* , Department of Statistics and Operations Research, New York University.

Harrison, T. (1994) *Lean And Mean: The Changing Landscape Of Corporate Power In The Age Of Flexibility* , Basic Books, New York.

Heitjan, D. F. and Rubin, D. B. (1990) Inference from coarse data via multiple imputation with application to age heaping, *Journal Of The American Statistical Association*, **85**, 304-314.

Holmgren, E. B. (1995.) The P-P Plot As A Method For Comparing Treatment Effects, *Journal Of The American Statistical Association*, **90**, 360-365.

Hsieh, F. (1995) The Empirical Process Approach For Semiparametric Two-Sample Models With Heterogeneous Treatment Effect, *Journal Of The Royal Statistical Society, Series B,* , **57**, 735-748.

Juhn, C., et al. (1993) U.S. Wage Inequality And The Rise In Returns To Skill, *The Journal Of Political Economy*, **101**, 410-422.

Karoly, L. A. (1993) The Trend In Inequality Among Families, Individuals, And Workers In The United States: A Twenty-Five Year Perspective, in *Uneven Tides: Rising Inequality In America*, S. Danziger and P. Gottschalk (Ed.), 19-97.

Katz, L. F. and Murphy, K. M. (1992) Changes In Relative Wages, 1963-1987: Supply And Demand Factors, *Quarterly Journal Of Economic*, **107**, 35-78.

Lehmann, E. L. (1975) *Nonparametrics: Statistical Methods Based On Ranks* , Holden Day, Oakland California.

Lehmann, E. L. (1983) *The Theory Of Point Estimation* , Wiley, New York.

Levy, F. and Murname, R. (1992) U.S. Earnings Levels And Earnings Inequality: A Review Of Recent Trends And Proposed Explanations, *Journal Of Economic Literature*, **30**, 1333-1381.

Li, G., et al. (1996) Quantile Comparison Functions In Two-Sample Problems, With Application To Comparisons Of Diagnostic Markers, *Journal Of The American Statistical Association*, **91**, 689-698.

Lorenz, M. O. (1905) Methods Of Measuring The Concentration Of Wealth, *Journal Of The American Statistical Association*, **9**, 209-219.

McCulloch, R. E. (1989) Local model influence, *Journal Of The American Statistical Association*, **84**, 473- 478.

Mielniczuk, J. (1992) Grade estimation of Kullback-Leibler information number, *Prob*

Math Stat, **13**, 139- 147.

Morris, M. (1993) Telling tails explain the discrepancy in sexual partner reports, *Nature*, **365**, 437-440.

Morris, M. (1996) Vive la Difference: Continuity and Change in the Gender Wage Gap, 1967-1987, in *Social Differentiation and Social Inequality*, J. Baron, D. Treiman and D. Grusky (Ed.).

Morris, M., et al. (1994) Economic Inequality: New Methods for New Trends, *American Sociological Review*, **59**, 205-219.

Parzen, E. (1977) *Nonparametric Statistical Data Science: A Unified Approach Based On Density Estimation And Testing For "White Noise"* , Statistical Sciences Division, State University of New York At Buffalo.

Parzen, E. (1992) Comparison Change Analysis, in *Nonparametric Statistics And Related Topics*, A. Saleh (Ed.), 3-15.

Prihoda, T. (1981) *A Generalized Approach To The Two Sample Problem: The Quantile Approach* , Department Of Statistics, Texas A&M University.

Shannon, C. E. (1948) A Mathematical Theory of Communication, *Bell System Technical Journal*, **27**, 379-423.

Silverman, B. W. (1978) Density Ratios, Empirical Likelihood And Cot Death, *Applied Statistics*, **X**, 26-33.

Simonoff, J. S. (1996) *Smoothing Methods In Statistics* , Springer-Verlag, New York.

Soofi, E. S. (1994) Capturing the intangible concept of information, *Journal Of The American Statistical Association*, **89**, 1243-1254.

Stevens, A. H. (1996) *Changes in earnings instability and job loss* , Rutgers University.

Theil, H. and Laitinen, K. (1980) Singular moment matrices in applied econometrics, in *Multivariate Analysis V*, P. R. Krishnaiah (Ed.), 629- 649.

Wilk, M. B. and Gnanadesikan, R. (1968) Probability Plotting Methods For The Analysis Of Data, *Biometrika* , **55**, 1-17.

Table 1: Kullback–Leiber dispersions for the Location/Shape Decomposition of Wage Gains: Recent to Original NLS Cohort

K–L Dispersion	Estimate	95% CI	p–value
overall change in wage growth	0.156	0.111 – 0.200	0.000
effect of location change	0.066	0.037 – 0.095	0.000
effect of shape change	0.090	0.050 – 0.129	0.000
percent due to location	42.3%	26.9 – 57.8	0.000
percent due to shape	57.7	42.2 – 73.1	0.000

Table 2: Relative Polarization Indices for Wage Gains: Recent to Original NLS Cohort

Polarization Index	Estimate	95% CI	p–value
Median Index	0.183	0.148 – 0.218	0.000
Lower Index	0.190	0.111 – 0.270	0.000
Upper Index	0.176	0.097 – 0.256	0.000

Table 3: Kullback–Leiber dispersions for the Age Decomposition of Wage Gains: Recent to Original NLS Cohort

K–L Dispersion	Estimate	95% CI	p–value
overall change wage growth	0.156	0.111 – 0.200	0.000
age compositional effect	0.001	-0.001 – 0.004	0.175
age–adjusted effect	0.154	0.112 – 0.196	0.000
percent due to age	0.9%	-0.5 – 2.3	0.175
residual	99.1	97.7 – 100.5	0.000

Table 4: Summary Statistics for Cohort Relative Distributions by Education

Measure	High–School	College
median ratio	0.77	0.99
K–L divergence	0.27	0.08
location divergence	0.15	0.00
shape divergence	0.12	0.08
Polarization (MRP)	0.16	0.19
Lower tail (LRP)	0.20	0.29
Upper tail (URP)	0.12	0.08

Table 5: Decile Relative Distributions of High School to College Educated Wage Gains by Cohort

Decile	Original Cohort	Recent Cohort	Change in decile
1	19.0	26.9	7.9
2	15.0	17.4	2.4
3	17.0	18.5	1.5
4	9.1	12.9	3.8
5	9.7	8.7	-1.0
6	11.0	6.2	-4.8
7	8.2	5.9	-2.3
8	5.1	2.1	-3.0
9	3.3	0.6	-2.7
10	2.5	1.2	-1.3

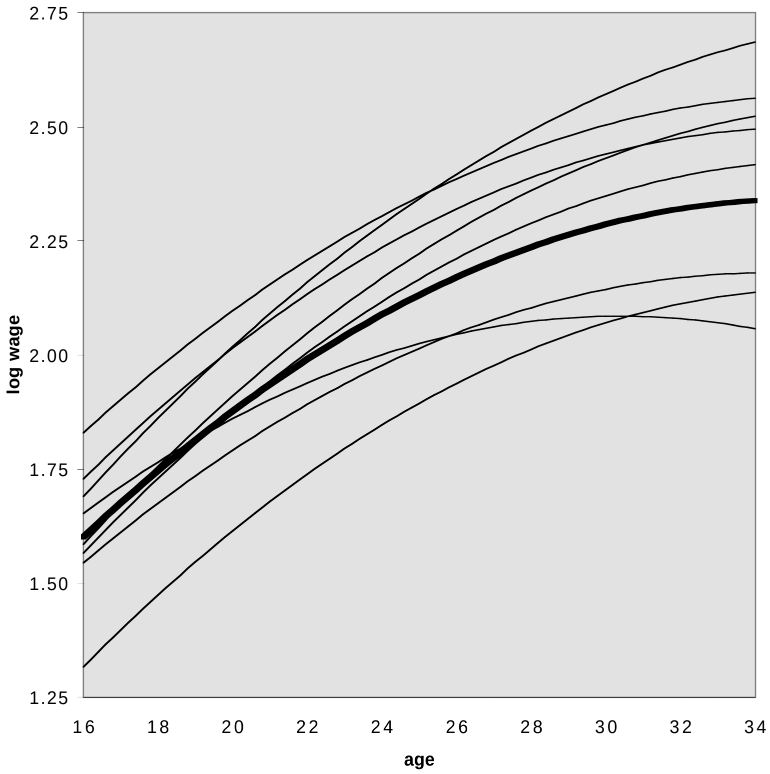

Fig. 1. Idealized wage profiles. The mixed effects model specified in equation 1.1 fits a fixed effect wage profile as a quadratic in age for the sample (shown in bold) but allows individual random effects for the intercept, linear and quadratic terms to generate profile heterogeneity. A typical distribution of fitted profiles might look like this.

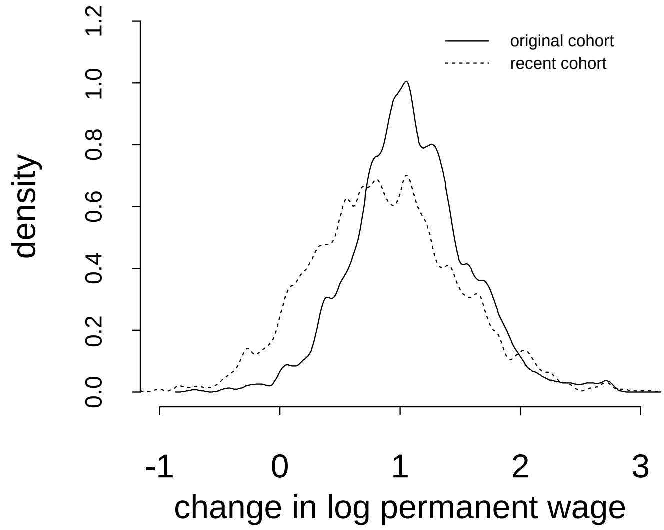

Fig. 2. The distributions of permanent wage growth in the original and recent NLS cohorts.

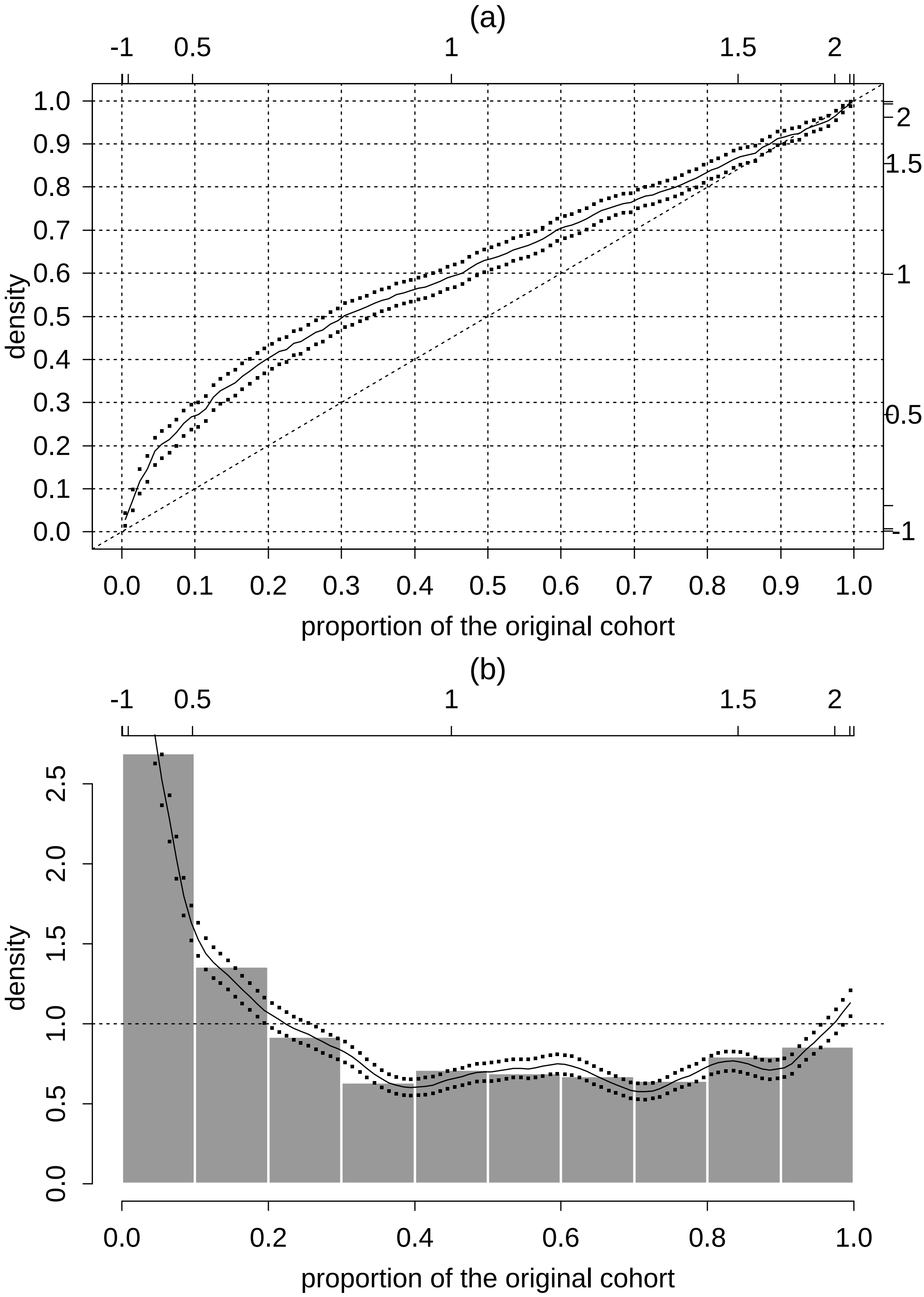

Fig. 3. The relative distribution of permanent wage growth in the original and recent NLS cohorts: (a) the relative CDF; (b) the relative PDF. A decile bar chart is superimposed on the density estimate. The upper and right axes are labeled in permanent log wage gains. The dotted lines are 95% pointwise confidence bounds.

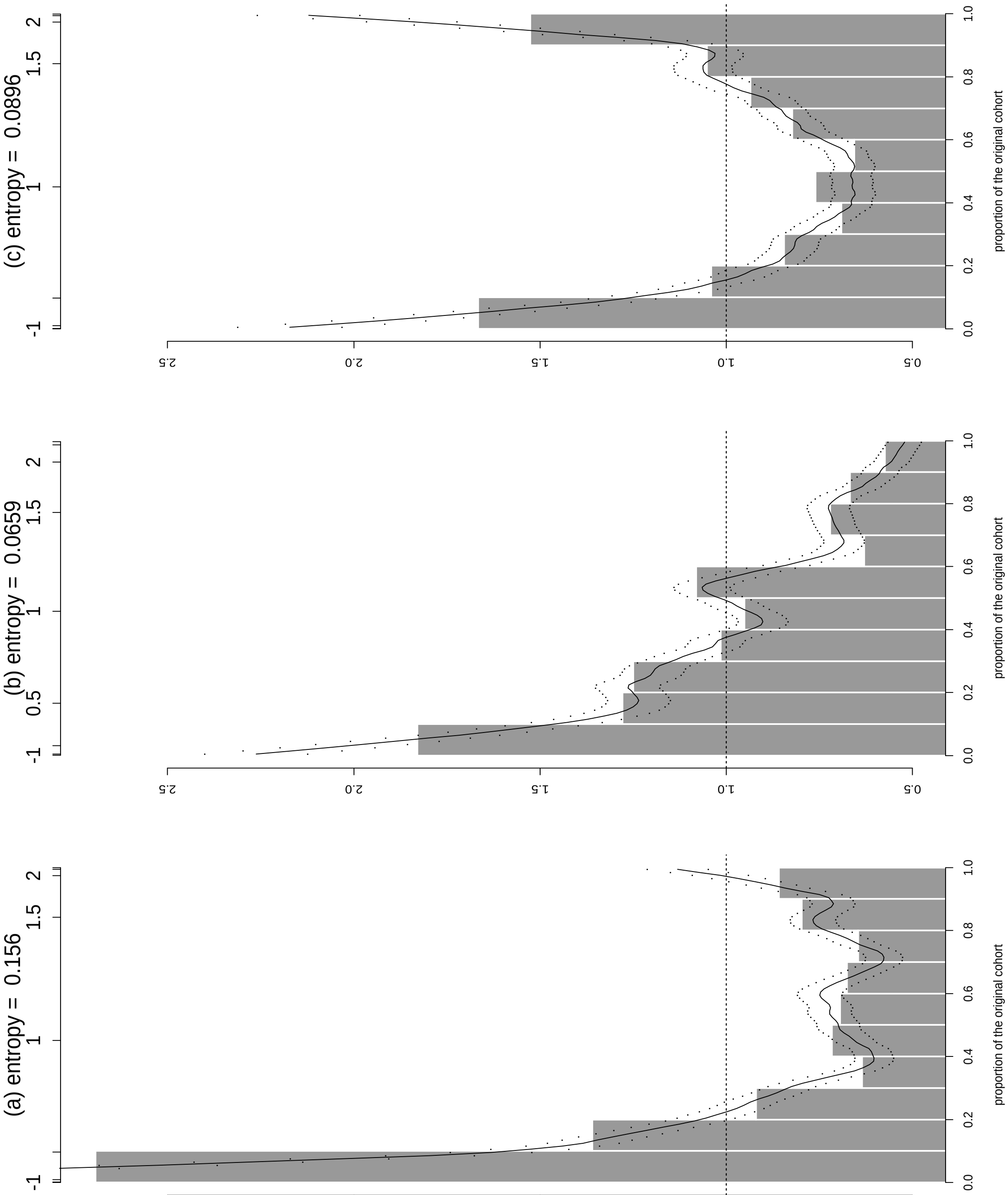

Fig. 4. Decomposing the relative distribution of permanent wage growth in the recent and original cohorts into the impact of changes in medians and changes in shape. (a) The (unadjusted) relative density of wage growth; (b) The effect of the median difference in wage growth between the cohorts; (c) The median–adjusted relative density of wage growth (the effect of changes in distributional shape)

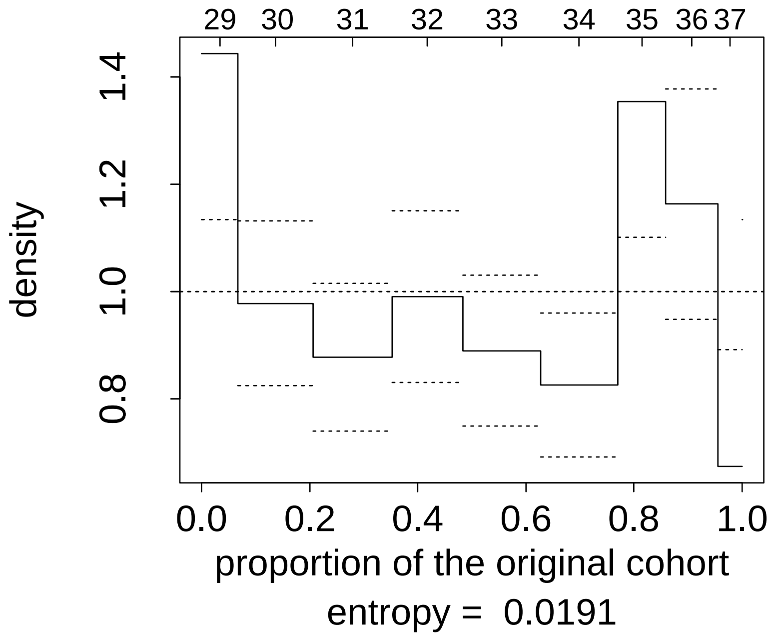

Fig. 5. The relative distribution of age for the recent to the original cohort. The upper axis is labeled by final age. The dotted lines are 95% pointwise confidence bounds.

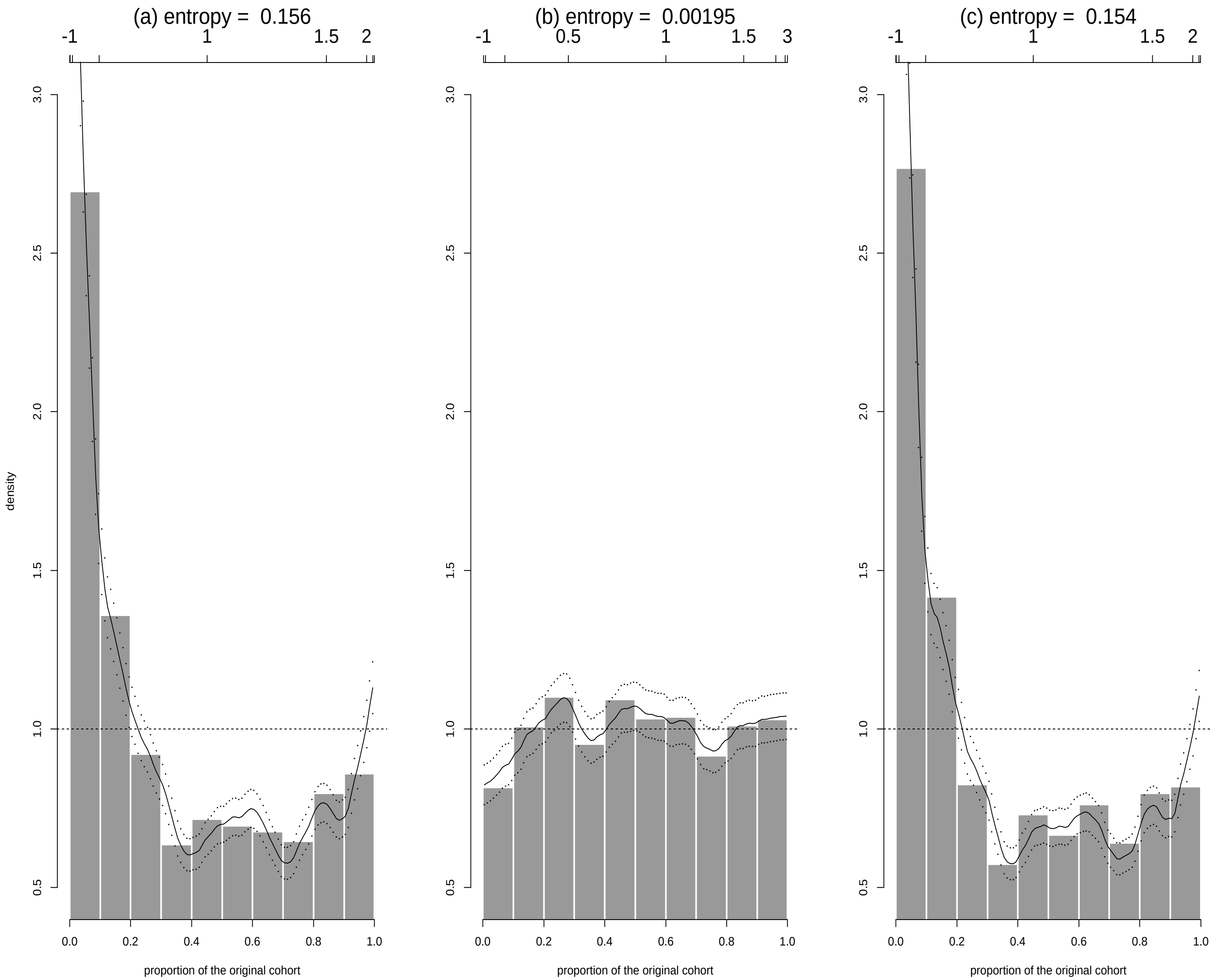

Fig. 6. Decomposing the relative distribution of permanent wage growth in the recent and original cohorts into the impact of changes in age composition and residual changes. (a) The (unadjusted) relative density of wage growth; (b) The effect of changes in the age profile between the cohorts; (c) The age–adjusted relative density of wage growth

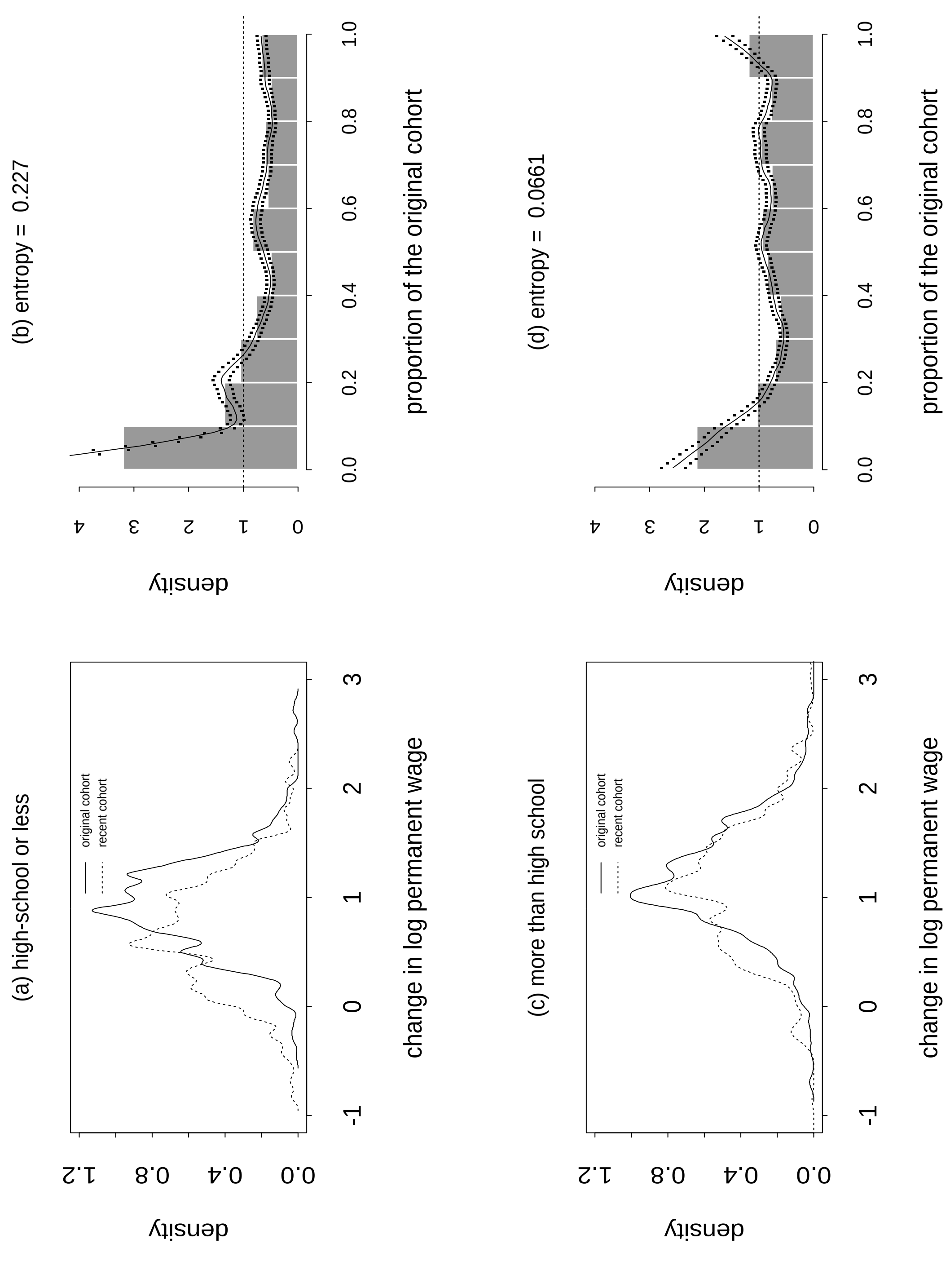

Fig. 7. The PDF overlays and cohort relative distributions of permanent wage growth for high school and college educated workers in the original and recent NLS cohorts. (a) Wage gain PDFs for workers with high–school or less education in each cohort; (b) Cohort relative distribution (R:O) for those with high–school or less; (c) Wage gain PDFs for workers with some college in each cohort; (d) Cohort relative distribution (R:O) for those with some college. A decile bar chart is superimposed on the relative density estimates. The upper and right axes are labeled in permanent log wage gains. The dotted lines are 95% pointwise confidence bounds. The distributions have been adjusted for age compositional differences using the procedure in Section 4.

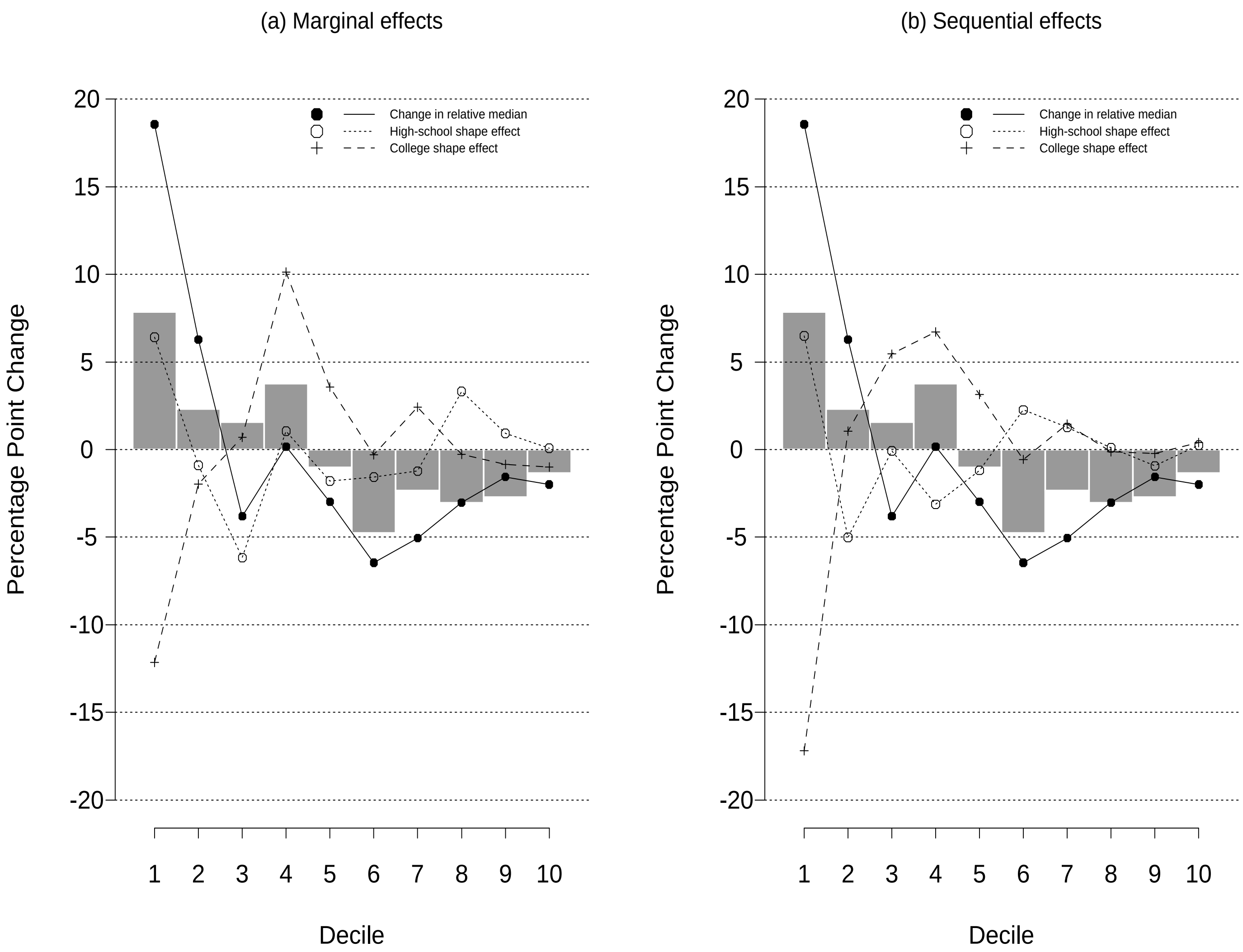

Fig. 8. Decompositions of the change in the cohort relative distribution of wage gains by education level. (a) the marginal effects as defined by (5.1), and (b) the sequential effects as defined by (5.2).