

Applying Relative Distribution Methods in R ¹

Mark S. Handcock
and
Eric M. Aldrich
University of Washington

Working Paper no. 27
Center for Statistics and the Social Sciences
University of Washington
Box 354322
Seattle, WA 98195-4322, USA

December 2002

¹Mark S. Handcock is Professor of Statistics and Sociology, Department of Statistics, University of Washington, Seattle, WA 98195-4320 (E-mail: handcock@stat.washington.edu) and; Eric M. Aldrich is a graduate student, Department of Statistics, University of Washington, Seattle, WA 98195-4320 (E-mail: eric@stat.washington.edu). This research was partially supported by the Russell Sage Foundation and the Rockefeller Foundation. The authors wish to thank Michele L. Shaffer, Annette D. Bernhardt, and Martina Morris.

Abstract

Relative distribution methods are a nonparametric statistical approach to the comparison of distributions. These methods combine the graphical tools of exploratory data analysis with statistical summaries, decomposition, and inference.

This report demonstrates software for implementing relative distribution methods within the R statistical package. It describes how to download and install the software, and use it to redo the analysis in the paper “Relative Distribution Methods” by Mark S. Handcock and Martina Morris, *Sociological Methodology*, Vol 28, July 1998. The full code, references and links to further resources are provided.

1 Introduction

In social science research, differences among groups or changes over time are a common focus of study. While means and variances are typically the basis for statistical methods used in this research, the underlying social theory often implies properties of distributions that are not well captured by these summary measures. Consider some of the current controversies regarding growing inequality in earnings, racial differences in test scores, socio-economic correlates of birth outcomes, and the impact of smoking on survival and health. The distributional differences 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. Survey and census data on such attributes contain a wealth of distributional information, but traditional methods of data analysis leave much of this information untapped.

Handcock and Morris (1998) (HM) present methods for full comparative distributional analysis. The methods are based on the relative distribution, a non-parametric complete summary of the information required for scale-invariant comparisons between two distributions. The relative distribution 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. A book length treatment of the methodology is given in Handcock and Morris (1999).

The data used in the report are from two cohorts of the National Longitudinal Survey (NLS), one initiated in 1966 and the other in 1979. These cohorts are referred to as the original and recent cohorts, respectively. The distributions of wage growth in the two cohorts are examined. Specifically, the growth profile of “permanent wages” is analyzed to study the question of wage mobility. A development of the estimation of these permanent wages and their relevance to the study of wage mobility is given in HM. For the purposes of this report, we can regard the permanent wages as measurements on two groups that we wish to compare.

The software described here can be obtained from the Relative Distribution Methods webpage at <http://www.stat.washington.edu/handcock/ReIDist>. Additional reports of interest can also be found here. The software is available in a number of forms. Here we will focus on the version for the statistical program R.

In the following sections we will reconstruct the analyses given in HM without discussing their substantive interpretation. To fully understand the use of the software it would be useful to have a copy of HM available for reference. The appropriate function calls and associated code are given for the methods described by section. In Section 2, directions are provided for the installation of the R statistical software and the `reldist` package, which is used to create graphical representations of relative distributions. In Section 3, the standard approach to comparing the two distributions is presented. In Section 4, the relative CDF and PDF of permanent wage growth in the original and recent NLS cohorts are constructed. In Section 5, the relative distribution of permanent wage growth in the two cohorts is decomposed into the impact of changes in medians and changes in shape. In Section 6, summary statistics for the location/shape decomposition of the relative distribution of wage gains are computed. In

Section 7, an example of covariate adjustment is provided, adjusting the relative distribution of permanent wage growth for changes in educational composition between the two cohorts. And finally in Section 8, the code necessary for reproducing a discrete level contrast example and an additive decomposition example is given.

2 Installing the software

2.1 Installing R

R is a language and environment for statistical computing and graphics. R is often referred to “GNU S” as it was inspired by the well-known and powerful software that has also inspired “S-Plus”. R is available as Free Software under the terms of the Free Software Foundation’s GNU General Public License in source code form. It can be easily installed over the web on a wide variety of LINUX and UNIX platforms. It also runs on Windows 9x/NT/2000/XP and MacOS. Official information about the R Project can be found at <http://www.r-project.org>. To download the most recent version of the software, go to <http://cran.r-project.org>. Alternatively, follow the “CRAN” link from the R Project website to find a list of URL’s that comprise the Comprehensive R Archive Network (CRAN). All of the CRAN servers maintain current versions of code and documentation for the R statistical package. From the CRAN, the R source code can be downloaded, or precompiled binary distributions can be downloaded for particular platforms. For platform specific installation instructions, follow the appropriate links.

2.2 Installing the R package `reldist`

To install the `reldist` package, simply open R and type

```
install.packages("reldist",contriburl="http://www.csde.washington.edu/~handcock")
```

at the prompt. Once the package has been downloaded and installed, it can be used by typing the call

```
library(reldist).
```

For instructions on usage, read the man page by typing

```
help(package="reldist").
```

Finally, to access the data that was used to create the figures in this paper, type

```
library(reldist)
data(nls).
```

Using the `nls` data with the code provided below, it is possible to recreate the figures at the end of this paper. This should be an optimal way to check the installation of the `reldist` package.

The data set contains two `data.frames`, one for the “original” and one for “recent” cohort (called `original` and `recent`, respectively). Each `data.frame` has three columns: the change

in permanent wages (in log-dollars), the final achieved educational level (in years), and the sample weight. The columns names are `chpermwage`, `endeduc`, and `wgt`, respectively.

The commands listed in this report are also available for download from the Relative Distribution Methods website.

3 Density Estimation

The standard approach to comparing the original and recent NLS cohort distributions involves looking at the summary statistics and plotting the probability density functions (PDFs) and Lorenz curves. The built in R function used for estimating the PDFs is `density`.

Because our data has weights we create a (weighted) sample from it for the purposes of creating these graphs. The commands are:

```
schpermwage1<-sample(original$chpermwage,size=1000,
                    prob=original$wgt/sum(original$wgt),replace=TRUE)
schpermwage2<-sample(recent$chpermwage,size=1000,
                    prob=recent$wgt/sum(recent$wgt),replace=TRUE)
```

The plots of the PDFs given in Figure 1 (a) are generated by

```
kwidth <- 0.2597
dens1 <- density(schpermwage1, n = 500, width=kwidth)
plot(x = (dens1$x), y = dens1$y, type = "l",
     xlab = "change in log permanent wage", ylab = "density",
     axes = FALSE,
     xlim = c(-1, 3),
     ylim=c(0,1.2))
title(main="(a)",cex=0.6)
axis(side = 1)
axis(side = 2)
fig1legend <- list(x=c(1.2,1.2),y=c(1.2,1.2))
legend(fig1legend,lty=1:2,cex=0.5, bty="n",
       legend=c("original cohort","recent cohort"))
dens2 <- density(schpermwage2, n = 500, width=kwidth)
lines(x = (dens2$x), y = dens2$y, type = "l",lty=2)
```

From the plot of the PDFs, we see that the recent cohort experienced smaller average wage gains, these gains were more variable, and the frequency of low wage gains was much greater for the recent cohort.

Lorenz curves are a standard method used for inequality comparison. The plots of the Lorenz curves given in Figure 1 (b) are generated by

```
swage1 <- sort(recent$chpermwage)
swage2 <- sort(original$chpermwage)
xout <- (0:1000)/1000
alpha <- seq(along=swage2)/length(swage2)
```

```

galpha <- cumsum(swage2)/sum(swage2)
fn1 <- approx(x=alpha,y=galpha,xout=xout)
plot(x = alpha, y = galpha, type = "l",
     xlab = "proportion of population",
     ylab = "proportion of wages",
     ylim=c(0,1.0))
legend(x=c(0,0),y=c(1.03,1.03),lty=1:2,cex=0.5, bty="n",
      legend=c("original cohort","recent cohort"))

```

Here we see that the Lorenz curve for the recent cohort lies uniformly below that of the original cohort which indicates that there is more inequality in the distribution of recent wage gains.

4 The Relative Distribution

To do a full distributional comparison based on the relative distribution, we look at the PDF and CDF of the relative distribution. Both preserve all of the information necessary to compare the two distributions. If the two distributions are identical, then the CDF of the relative distribution is a 45° line and the PDF of the relative distribution is that of the uniform on $[0, 1]$.

To obtain the relative CDF in Figure 2 (a) we use

```

fig2a <- reldist(y=recent$chpermwage,yo=original$chpermwage,
               ci=FALSE,smooth=0.4,
               yowgt=original$wgt,ywgt=recent$wgt,
               cdfplot=TRUE,
               yolabs=seq(-1,3,by=0.5),
               ylabs=seq(-1,3,by=0.5),
               cex=0.8,
               ylab="proportion of the recent cohort",
               xlab="proportion of the original cohort")
title(main="(a)",cex=0.6)

```

The option `cdfplot = TRUE` is used to obtain a plot of the CDF rather than the (default) density. The options `yowgt` and `ywgt` are used to assign a vector of weights to the reference distribution and comparison distribution, respectively. The `smooth` option identifies the degree of smoothness required in the fit. Specifying higher values of `smooth` leads to smoother curves, while specifying lower values leads to closer fits to the observed data. The relative PDF in Figure 2 (b) is produced with

```

fig2b <- reldist(y=recent$chpermwage,yo=original$chpermwage,
               ci=FALSE,smooth=0.4,
               yowgt=original$wgt,ywgt=recent$wgt,bar=TRUE,
               yolabs=seq(-1,3,by=0.5),
               ylim=c(0,2.5),cex=0.8,
               ylab="Relative Density",

```

```

                                xlab="Proportion of the Original Cohort")
title(main="(b)",cex=0.6)

```

Here the option `bar=TRUE` is used to superimpose a barplot on the relative density estimate. Figures 3 (a) and (b) show the effects of increasing the `smooth` option used in Figures 2 (a) and (b) by specifying `smooth = 1.2`.

5 Decomposing the Relative Distribution

In this section we decompose the overall relative distribution into two component relative distributions which depict differences in location and shape. Figure 4 displays the median and shape decomposition of the relative distribution of weight gains and is generated by

```

par(mfrow=c(1,3))
g10 <- reldist(y=recent$chpermwage, yo=original$chpermwage,
              smooth=0.4, ci=FALSE,
              ywgt=recent$wgt, yowgt=original$wgt,
              yolabs=seq(-1,3,by=0.5),
              ylim=c(0.5,3.0),
              bar=TRUE, quiet=FALSE,
              xlab="proportion of the original cohort")
title(main=paste("(a) entropy = ",format(g10$entropy,digits=3)),cex=0.6)
abline(h=1,lty=2)
g1A <- reldist(y=recent$chpermwage, yo=original$chpermwage,
              ywgt=recent$wgt, yowgt=original$wgt,
              show="effect",
              bar=TRUE, quiet=FALSE,
              ylim=c(0.5,3.0), ylab="",
              smooth=0.4, ci=FALSE,
              yolabs=seq(-1,3,by=0.5),
              xlab="proportion of the original cohort")
title(main=paste("(b) entropy = ",format(entropy(g1A,g10),digits=3)),cex=0.6)
abline(h=1,lty=2)
gA0 <- reldist(y=recent$chpermwage, yo=original$chpermwage,
              smooth=0.4, ci=FALSE,
              ywgt=recent$wgt, yowgt=original$wgt,
              show="residual",
              bar=TRUE, quiet=FALSE,
              ylim=c(0.5,3.0), ylab="",
              yolabs=seq(-1,3,by=0.5),
              xlab="proportion of the original cohort")
title(main=paste("(c) entropy = ",format(gA0$entropy,digits=3)),cex=0.6)
abline(h=1,lty=2)

```

Panel (a) shows the overall relative density (and is the same as Figure 2 (b)). Panel (b) represents the effect of the median shift in the wage gains between the two cohorts – displaying

what the relative density would have looked like if there had been no change in distributional shape. The option `show="effect"` is used to additively shift the reference sample median to the comparison sample median before comparing the two distributions. Panel (c) displays the effect of changes in distributional shape. The option `show="residual"` is used to additively scale the reference sample to the comparison sample before comparing the two distributions.

6 Summary Measures

To complement the graphical displays of the preceding sections, we compute summary measures based on the relative distribution which can be used for the comparison of distributional change. In particular, we calculate entropy, which is a widely used measure of the dispersion of the distribution, and the median relative polarization index, which provides a means to measure distributional polarization. Both summary measures have useful decompositions. The overall entropy may be decomposed into a median effect and a shape effect. The median relative polarization index may be decomposed into upper and lower polarization indices, representing the contributions made by components above and below the median of the relative distribution, respectively. In Table 1 the full set of summary statistics is presented. Note also that entropy summaries are given on the top of Figure 4.

The summary statistics may be reproduced with

```
format(rpy(y=recent$chpermwage,yo=original$chpermwage,
          ywgt=recent$wgt,yowgt=original$wgt,pvalue=TRUE),
       digits=3)
format(rply(y=recent$chpermwage,yo=original$chpermwage,
           ywgt=recent$wgt,yowgt=original$wgt,pvalue=TRUE),
       digits=3)
format(rply(y=recent$chpermwage,yo=original$chpermwage,
           ywgt=recent$wgt,yowgt=original$wgt,pvalue=TRUE,
           upper=TRUE),
       digits=3)
```

7 The Relative Distribution for Discrete Variables

The educational composition of the NLS cohorts may be different and this will likely effect the wage outcomes. We will consider the effect in the next section, but here consider the distributional difference between the two cohorts.

Figure 5 shows the relative distribution of final observed education in the two cohorts and is generated by

```
e1 <- original$endeduc
e1[e1 < 8] <- 8
e1[e1 > 18] <- 18
e2 <- recent$endeduc
```


		Entropy		
overall change in wage growth		0.125		
median effect		0.078		
shape effect		0.047		
percent due to median		62.4%		
percent due to shape		37.6%		
Polarization Index	Estimate	95% CI	<i>p</i> -value	
Median Index	0.183	0.148 – 0.219	0.000	
Lower Index	0.190	0.118 – 0.262	0.000	
Upper Index	0.176	0.104 – 0.249	0.000	

Table 1: Summary Statistics for the Location/Shape Decomposition of the Relative Distribution of Wage Gains: Recent to Original NLS Cohort

```
e2[e2 < 8] <- 8
e2[e2 > 18] <- 18
pdf("Fig5.pdf", width=4.5,height=4.5,horiz=FALSE)
g10 <- rddist(y=e2, yo=e1, pool=1, ci=FALSE, quiet=FALSE,
             ywgt=recent$wgt,yowgt=original$wgt,
             yolabs=sort(unique(e1)),
             ylab="relative density",
             xlab="proportion of the original cohort")
title(sub=paste("entropy = ",format(entropy(g10),digits=3)))
abline(h=1,lty=2)
```

8 Covariate Adjustment

One can separate the impacts of changes in population composition from changes in the covariate-outcome relationship by adjusting the relative distribution for changes in the distribution of other covariates. This method decomposes the relative distribution into the composition effect or the component that represents the effect of changes in the marginal distribution of the covariate, and a component that represents residual changes.

Based on Figure 5, the educational composition of the NLS cohorts has changed. The covariate adjustment technique can be used to determine whether differences in the educational profile between the two cohorts explain some of the changes in relative wage gains.

Figure 6 is a graphical representation of the adjustment of the relative distribution for education composition changes and is produced with

```
par(mfrow=c(1,3))
i3x <- sample(seq(along=original$chpermwage),
             size = 10*length(original$chpermwage),
             prob=rdsamp(e2,e1,recent$wgt,original$wgt),
             replace = TRUE)
```

```

schpermwage1 <- original$chpermwage[i3x]
wschpermwage1 <- original$wgt[i3x]
g10 <- reldist(y=recent$chpermwage, yo=original$chpermwage,
              smooth=0.4, ci=FALSE,
              ywgt=recent$wgt, yowgt=original$wgt,
              yolabs=seq(-1,3,by=0.5),
              ylim=c(0.5,3.0),
              bar=TRUE, quiet=FALSE,
              xlab="proportion of the original cohort")
title(main=paste("(a) entropy = ",format(g10$entropy,digits=3)),cex=0.6)
abline(h=1,lty=2)
g1A <- reldist(y=schpermwage1, yo=original$chpermwage,
              yowgt=original$wgt, ywgt=wschpermwage1,
              bar=TRUE, quiet=FALSE,
              ylim=c(0.5,3.0), ylab="",
              smooth=0.4, ci=FALSE,
              yolabs=seq(-1,3,by=0.5),
              xlab="proportion of the original cohort")
title(main=paste("(b) entropy = ",format(entropy(g1A,g10),digits=3)),cex=0.6)
abline(h=1,lty=2)
gA0 <- reldist(y=recent$chpermwage, yo=schpermwage1,
              smooth=0.4, ci=FALSE,
              ywgt=recent$wgt, yowgt=wschpermwage1,
              bar=TRUE, quiet=FALSE,
              ylim=c(0.5,3.0), ylab="",
              yolabs=seq(-1,3,by=0.5),
              xlab="proportion of the original cohort")
title(main=paste("(c) entropy = ",format(gA0$entropy,digits=3)),cex=0.6)
abline(h=1,lty=2)

```

Panel (a) is the (unadjusted) relative density of wage gains (same as Figure 2b), panel (b) represents the education composition effects, and panel (c) represents the *education-adjusted* relative density of wage gains. Thus panel (c) represents the expected relative density of wage gains had the education profiles of the two cohorts been identical.

9 Additional Topics

For a discrete covariate, we may adjust for this covariate as in Section 8, or we may compare the groups defined by the covariate directly. To demonstrate this technique, education is again used as a covariate, but now it is defined in discrete form. In particular education is divided into the categories of those with a high school degree or less and those with one or more years of college.

First we show the code for creating the variables:

```
# e11 is the final education for the Original
```

```

# cohort (with nob > 2 and non-attrited)
# 1= < HS 2= HS 3=HS+ 4= College+
#
# First generate the samples
#
#
e11 <- e1
e11[e11 < 12] <- 1
e11[e11 == 12] <- 2
e11[e11 > 12 & e11 < 16] <- 3
e11[e11 >= 16] <- 4
#
e12 <- e2
e12[e12 < 12] <- 1
e12[e12 == 12] <- 2
e12[e12 > 12 & e12 < 16] <- 3
e12[e12 >= 16] <- 4
#
i3x <- sample(seq(along=original$chpermwage),
              size = 10*length(original$chpermwage),
              prob=rdsamp(e12,e11,recent$wgt,original$wgt),
              replace = TRUE)
schpermwage1 <- original$chpermwage[i3x]
wschpermwage1 <- original$wgt[i3x]
sel1 <- e11[i3x]
#
pwhso <- schpermwage1[sel1 <= 2 & !is.na(sel1)]
pwsco <- schpermwage1[sel1 > 2 & !is.na(sel1)]
wgthso <- wschpermwage1[sel1 <= 2 & !is.na(sel1)]
wgtsco <- wschpermwage1[sel1 > 2 & !is.na(sel1)]
#
pwhsr <- recent$chpermwage[e12 <= 2 & !is.na(e12)]
pwsr <- recent$chpermwage[e12 > 2 & !is.na(e12)]
wgthsr <- recent$wgt[e12 <= 2 & !is.na(e12)]
wgtsr <- recent$wgt[e12 > 2 & !is.na(e12)]

```

Figure 7 compares the distributions of wage gains for the 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. Panels (c) and (d) compare the wage gains for the the college educated across the two cohorts.

The plots are generated by

```

par(mfrow=c(2,2))
spwhso<-sample(pwhso,size=(100000),prob=wgthso/sum(wgthso),replace=TRUE)
spwsco<-sample(pwsco,size=(100000),prob=wgtsco/sum(wgtsco),replace=TRUE)
spwhsr<-sample(pwhsr,size=(100000),prob=wgthsr/sum(wgthsr),replace=TRUE)

```

```

spwscr<-sample(pwscr,size=(100000),prob=wgtsr/sum(wgtsr),replace=TRUE)
nbar <- log(length(spwhso), base = 2) + 1
kwidth <- diff(range(spwhso))/nbar * 0.5
kwidth <- 1.4*kwidth
dens1 <- density(spwhso, n = 500, width=2.0*kwidth)
plot(x = (dens1$x), y = dens1$y, type = "l",
      xlab = "change in log permanent wage", ylab = "density",
      xlim = c(-1, 3), ylim=c(0,1.2))
fig7legend <- list(x=c(0.9,0.9),y=c(1.25,1.25))
legend(fig7legend,lty=1:2,cex=0.5, bty="n",
       legend=c("original cohort","recent cohort"))
title(main=paste("(a) high-school or less"),cex=0.6)
dens2 <- density(spwhsr, n = 500, width=1.7*kwidth)
lines(x = (dens2$x), y = dens2$y, type = "l",lty=2)
g10hs <- reldist(y=pwhsr, yo=pwhso, ci=FALSE, smooth=0.4,
                ywgt=wgtsr, yowgt=wgthso,
                bar=TRUE, quiet=FALSE,
                ylim=c(0,4),
                xlab="proportion of the original cohort")
title(main=paste("(b) entropy = ",format(g10hs$entropy,digits=3)),cex=0.6)
abline(h=1,lty=2)
nbar <- log(length(spwscr), base = 2) + 1
kwidth <- diff(range(spwscr))/nbar * 0.5
kwidth <- 1.2*kwidth
dens1 <- density(spwsco, n = 500, width=1.5*kwidth)
plot(x = (dens1$x), y = dens1$y, type = "l",
      xlab = "change in log permanent wage", ylab = "density",
      xlim = c(-1, 3), ylim=c(0,1.2))
fig1legend <- list(x=c(0.9,0.9),y=c(1.25,1.25))
legend(fig1legend,lty=1:2,cex=0.5, bty="n",
       legend=c("original cohort","recent cohort"))
title(main=paste("(c) more than high school"),cex=0.6)
dens2 <- density(spwscr, n = 500, width=2*kwidth)
lines(x = (dens2$x), y = dens2$y, type = "l",lty=2)
g10sc <- reldist(y=pwscr, yo=pwsco, ci=FALSE, smooth=0.4,
                ywgt=wgtsr, yowgt=wgtsco,
                bar=TRUE, quiet=FALSE,
                ylim=c(0,4),
                xlab="proportion of the original cohort")
title(main=paste("(d) entropy = ",format(g10sc$entropy,digits=3)),cex=0.6)
abline(h=1,lty=2)

```

To assess how much the location and shape shifts in each groups' distribution contributes to the overall change in their relative positions, we make a decomposition into the “marginal effects” of each change. It is also possible to obtain a unique decomposition by defining the effects sequentially. Figure 8 presents the two compositions side by side and is produced

with

```
par(mfrow=c(1,2))
rdhsrscr <- rdeciles(y=pwhsr, yo=pwscr, ywgt=wgthsr, yowgt=wgtsr,
  binn=binn)
rdhsosco <- rdeciles(y=pwhso, yo=pwsco, ywgt=wgthso, yowgt=wgtsco,
  binn=binn)
mscrdhsrscr <- rdeciles(y=pwhsr - wtd.median(pwhsr, weight=wgthsr) +
  wtd.median(pwhso, weight=wgthso), yo=pwscr -
  wtd.median(pwscr, weight=wgtsr) +
  wtd.median(pwsco, weight=wgtsco),
  ywgt=wgthsr, yowgt=wgtsr, binn=binn)
mhsrdhsrscr <- rdeciles(y=pwhso - wtd.median(pwhso, weight=wgthso) +
  wtd.median(pwhsr, weight=wgthsr), yo=pwsco -
  wtd.median(pwsco, weight=wgtsco) +
  wtd.median(pwscr, weight=wgtsr),
  ywgt=wgthso, yowgt=wgtsco, binn=binn)
m1rdhsrscr <- rdeciles(yo=pwsco,
  y=pwhsr - wtd.median(pwhsr, weight=wgthsr)
  + wtd.median(pwhso, weight=wgthso),
  yowgt=wgtsco, ywgt=wgthsr, binn=binn)
m2rdhsrscr <- rdeciles(y=pwhso,
  yo=pwscr - wtd.median(pwscr, weight=wgtsr)
  + wtd.median(pwsco, weight=wgtsco),
  ywgt=wgthso, yowgt=wgtsr, binn=binn)
m3rdhsrscr <- rdeciles(y=pwhsr,
  yo=pwsco - wtd.median(pwsco, weight=wgtsco)
  + wtd.median(pwscr, weight=wgtsr),
  yowgt=wgtsco, ywgt=wgtsr, binn=binn)
achange <- binn*(rdhsrscr$x - rdhsosco$x)
armeff <- binn*(mhsrdhsrscr$x - rdhsosco$x)
ahseff <- binn*(m1rdhsrscr$x - rdhsosco$x)
asceff <- binn*(m2rdhsrscr$x - rdhsosco$x)
ainteff <- achange - armeff - ahseff - asceff
barplot(height=achange, histo=TRUE, width=(1:binn)-0.5, axes=FALSE,
  xlab="Decile", ylab="Percentage Point Change",
  ylim=c(-20.0, 25))
axis(1, labels=TRUE, at=(1:binn))
axis(2, labels=TRUE, at=seq(-20.0, 25, length=10))
title(main="(a) Marginal effects", cex=0.6)
lines(y=(armeff), x=(1:binn), lty=1)
lines(y=(asceff), x=(1:binn), lty=3)
lines(y=(ahseff), x=(1:binn), lty=2)
abline(h=seq(-20, 25, length=10), lty=2)
points(y=(armeff), x=(1:binn), mark=16, cex=0.7)
points(y=(asceff), x=(1:binn), mark=3, cex=0.7)
```

```

points(y=(ahseff),x=(1:bin),mark=1,cex=0.7)
fig8legend <- list(x=c(4,4),y=c(25,25))
legend(fig8legend,pch=c(16,1,3),lty=c(1:3),cex=0.5, bty="n",
legend=c("Change in relative median",
"High-school shape effect","College shape effect"))
armeff <- binn*(mhsrdhsrscr$x - rdhsosco$x)
ahseff <- binn*(m3rdhsrscr$x - mhsrdhsrscr$x)
asceff <- binn*(rdhsrscr$x - m3rdhsrscr$x)
barplot(height=achange,histo=TRUE,width=(1:bin)-0.5,axes=FALSE,
xlab="Decile",ylab="Percentage Point Change",
ylim=c(-20.0,25))
axis(1,labels=TRUE,at=(1:bin))
axis(2,labels=TRUE,at=seq(-20.0,25,length=10))
title(main="(b) Sequential effects",cex=0.6)
lines(y=(armeff),x=(1:bin),lty=1)
lines(y=(asceff),x=(1:bin),lty=3)
lines(y=(ahseff),x=(1:bin),lty=2)
abline(h=seq(-20,25,length=10),lty=2)
points(y=(armeff),x=(1:bin),mark=16,cex=0.7)
points(y=(asceff),x=(1:bin),mark=3,cex=0.7)
points(y=(ahseff),x=(1:bin),mark=1,cex=0.7)
fig8legend <- list(x=c(4,4),y=c(20,25))
legend(fig8legend,pch=c(16,1,3),lty=c(1:3),cex=0.5, bty="n",
legend=c("Change in relative median",
"High-school shape effect","College shape effect"))

```

Panel (a) represents the marginal effects of the median shift from the original density, the marginal effect of the shape change in the high school distribution, and the marginal effect of the shape change in the college distribution. Panel (b) represents the sequential effects of the relative median shift from the original relative distribution, then the shape change in the college distribution from the median shifted original relative distribution, and finally the shape change in the high school distribution from the median shifted, college shape changed relative distribution.

References

- Handcock, Mark. S., and Morris, Martina (1998). Relative Distribution Methods. *Sociological Methodology*, Vol 28, p. 53-97.
- Handcock, Mark. S., and Morris, Martina (1999). *Relative Distribution Methods in the Social Sciences*. Springer-Verlag, New York.
- Shaffer, Michele. L. and Handcock, Mark. S. (1998). *Using Relative Distribution Software*, Technical Report, Department of Statistics, The Pennsylvania State University.

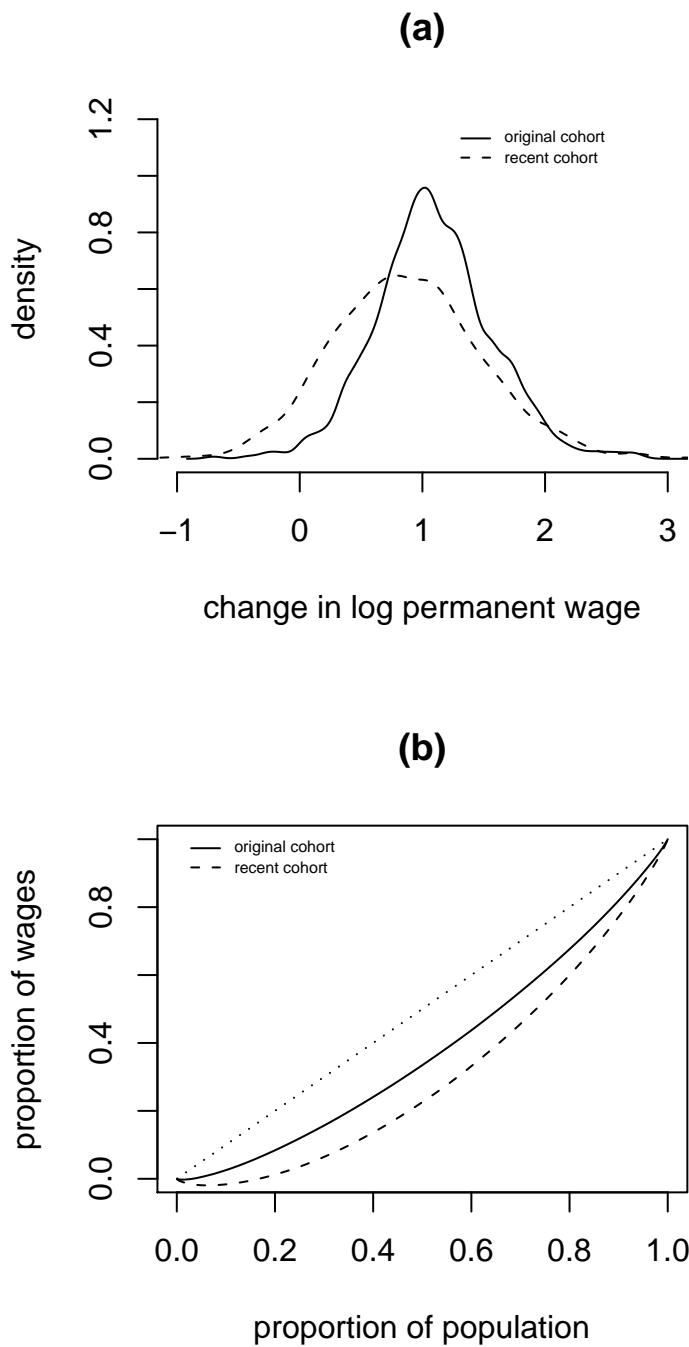


Figure 1: The distributions of permanent wage growth in the original and recent NLS cohorts. (a) PDF overlays for each cohort; (b) Lorenz curves for the PDFs.

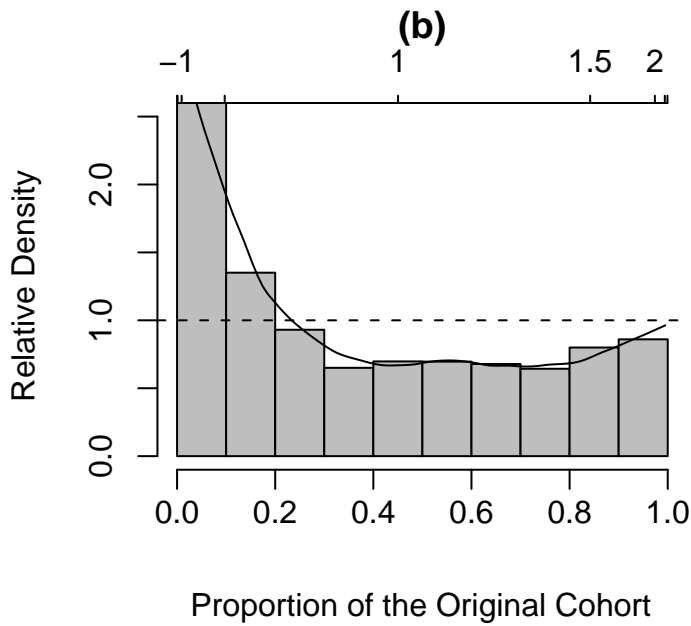
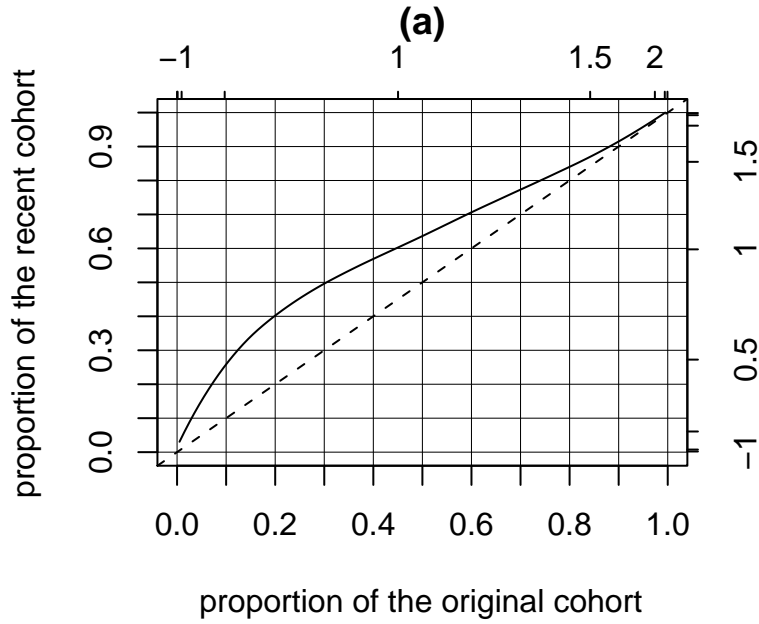


Figure 2: The relative distribution of permanent wage growth in the original and recent NLS cohorts: (a) the relative CDF; (b) the relative PDF. A decile bar chart is superimposed on the density estimate. The upper and right axes are labeled in permanent differences in log wages. The smoothing parameter is 0.4.

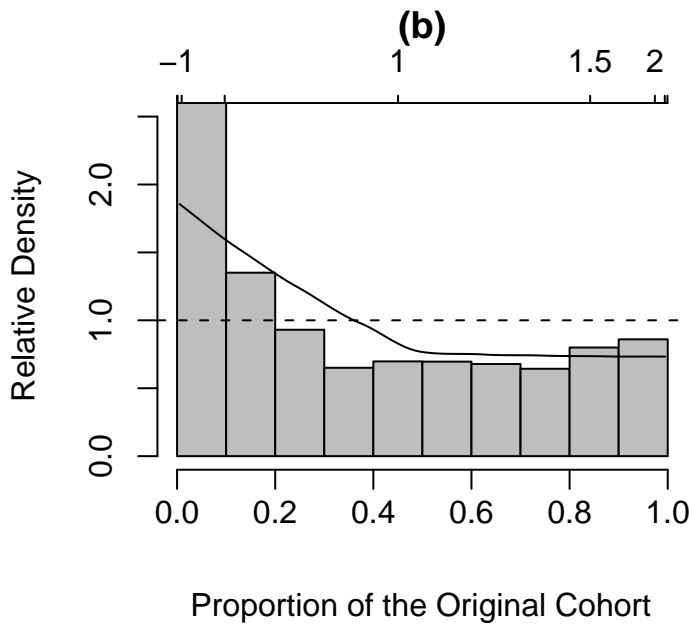
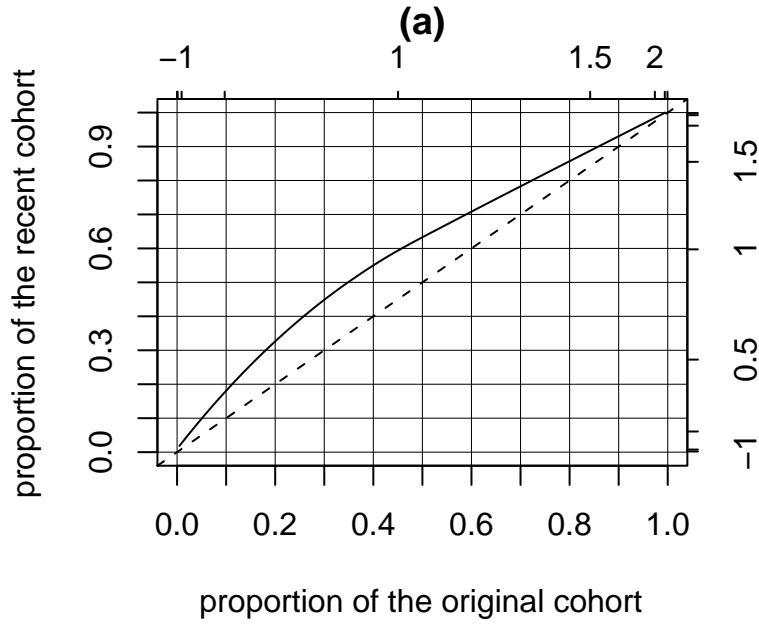


Figure 3: The relative distribution of permanent wage growth in the original and recent NLS cohorts: (a) the relative CDF; (b) the relative PDF. A decile bar chart is superimposed on the density estimate. The upper and right axes are labeled in permanent differences in log wages. The smoothing parameter is 1.2.

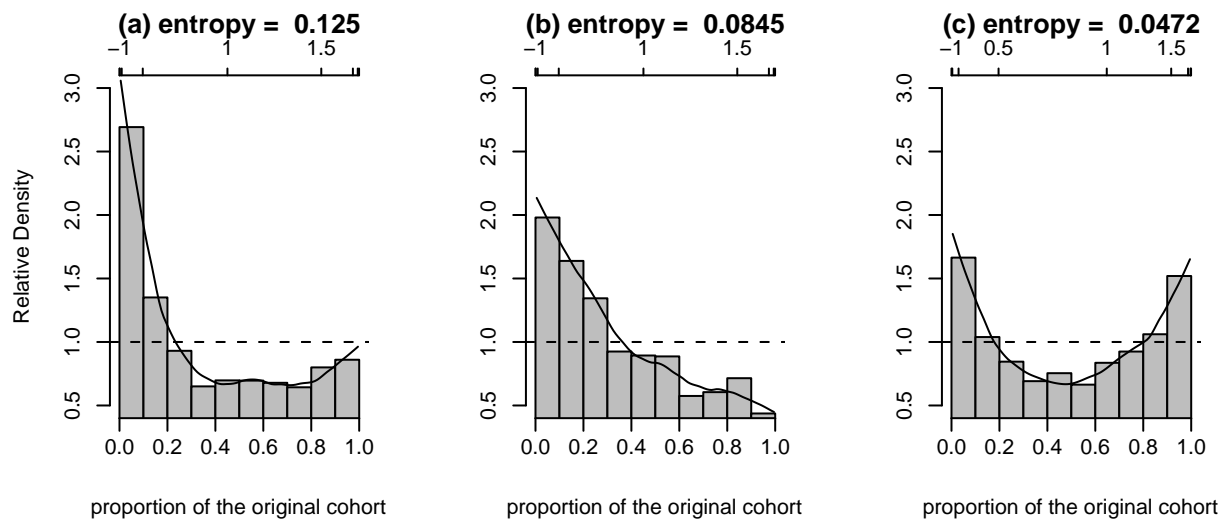


Figure 4: Decomposing the relative distribution of permanent wage growth in the recent and original NLS cohorts into the impact of changes in medians and changes in shape. (a) The (unadjusted) relative density of wage growth; (b) the effect of the median difference in wage growth between the cohorts;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

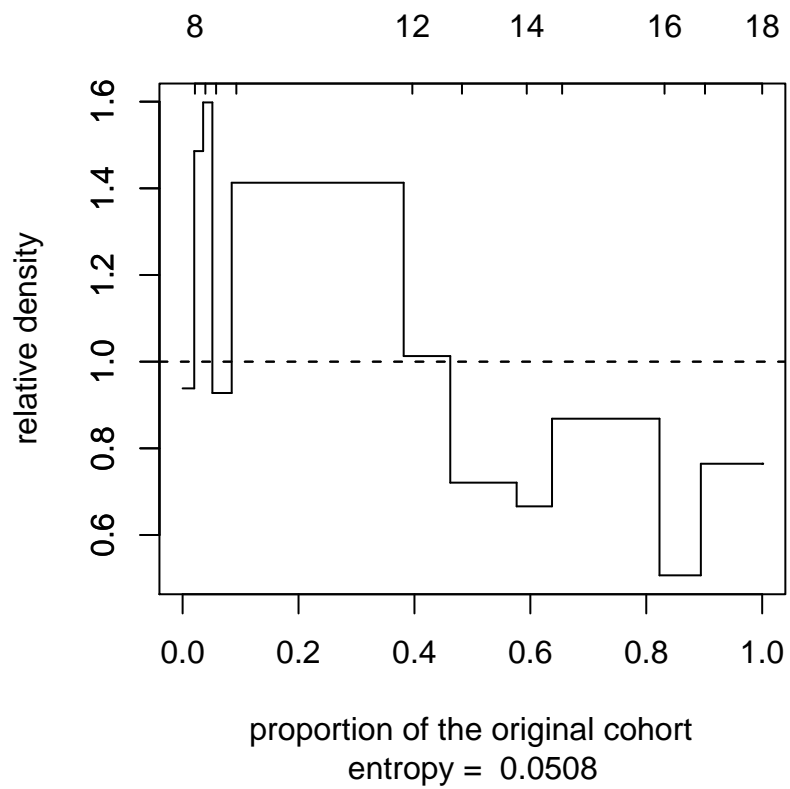


Figure 5: The relative distribution of education for the recent to the original cohort. The upper axis indicates the final number of years of schooling completed.

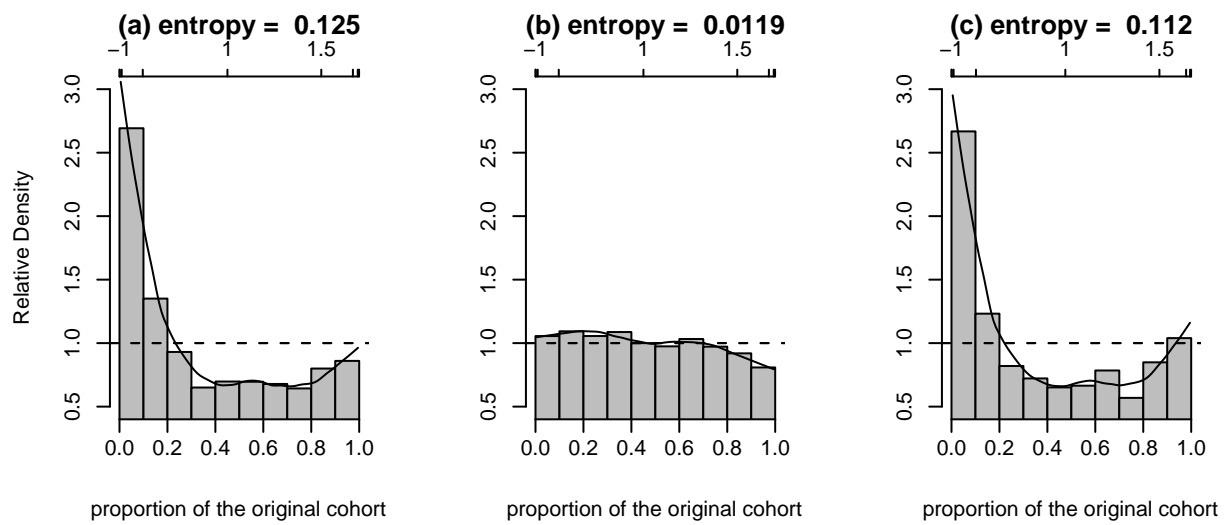


Figure 6: Adjusting the relative distribution of permanent wage growth for changes in the education composition between the two cohorts. (a) The (unadjusted) relative density of wage growth; (b) the effect of changes in the education profile between the cohorts; (c) the education-adjusted relative density of wage growth.

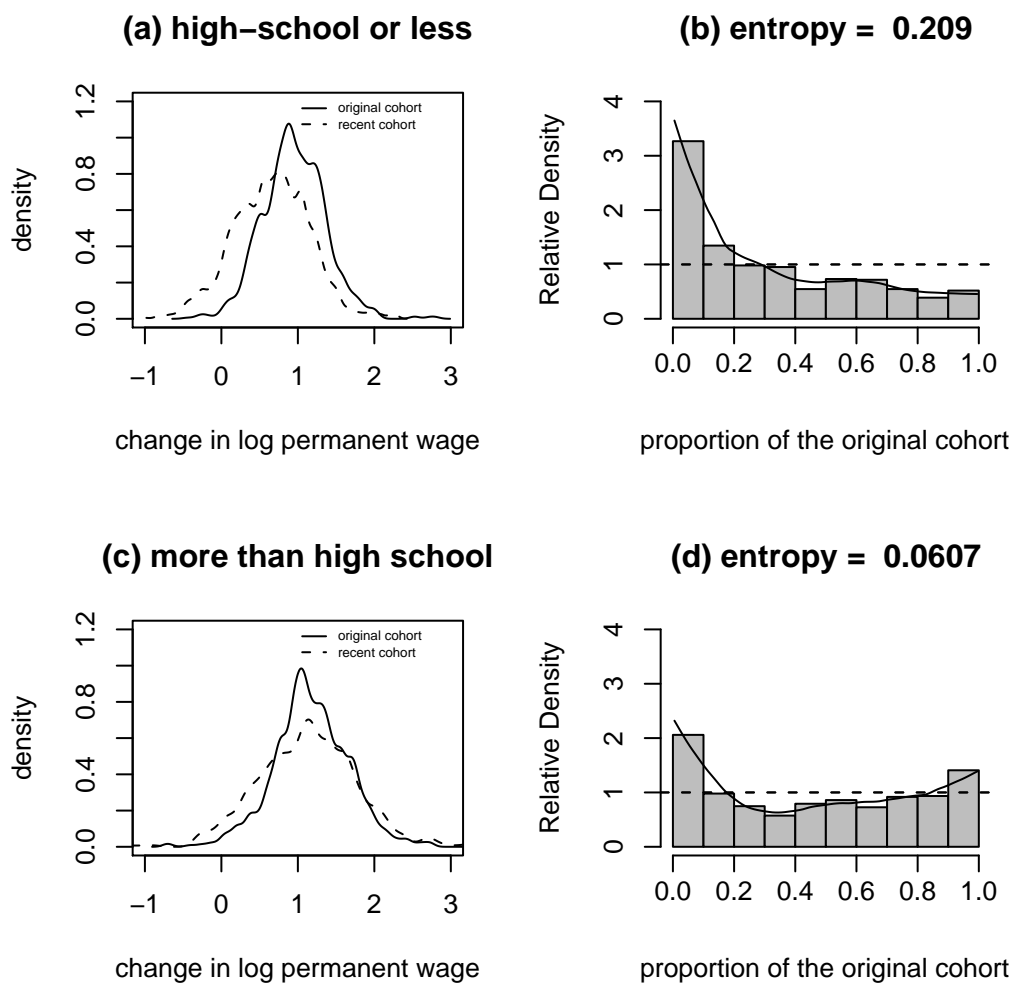


Figure 7: The PDF overlays and cohort relative distributions of permanent wage growth for high school and college-educated workers in the NLS. (a) wage gain PDFs for workers with high school or less education in each cohort; (b) cohort relative distribution (R:O) for those with high school or less; (c) 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.

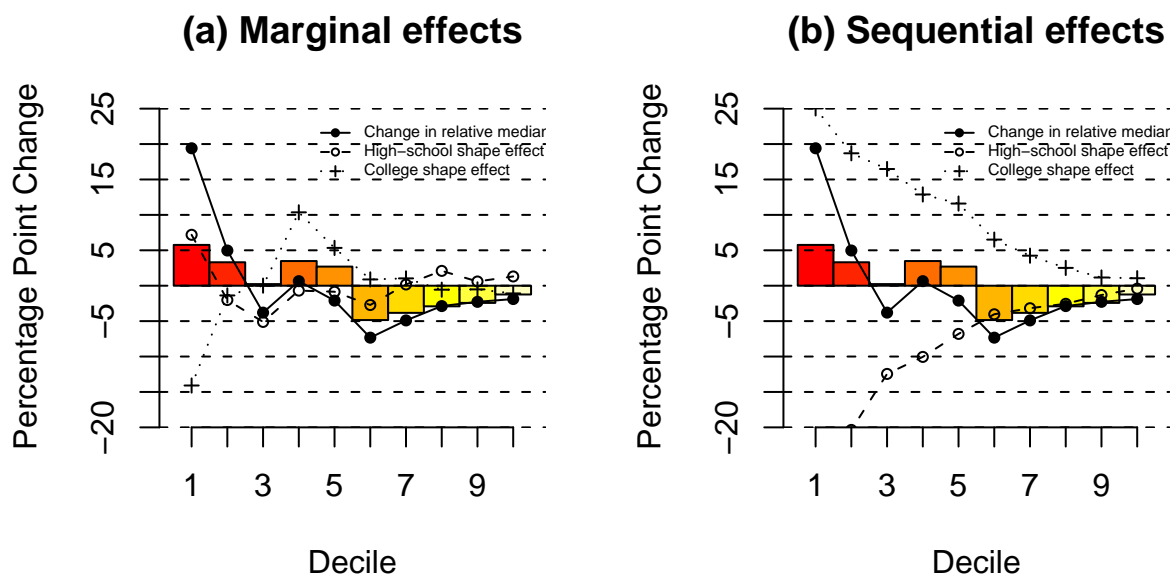


Figure 8: Sources of the change in the cohort relative distribution of wage gains by education level. (a) Marginal effects. (b) Sequential effects.