ISOLATING THE ROLES OF INDIVIDUAL COVARIATES IN REWEIGHTING ESTIMATION

TODD E. ELDER*, JOHN H. GODDEERIS AND STEVEN J. HAIDER
Department of Economics, Michigan State University, East Lansing, MI, USA

SUMMARY
A host of recent research has used reweighting methods to analyze the extent to which observable characteristics predict between-group differences in the distribution of an outcome. Less attention has been paid to using reweighting methods to isolate the roles of individual covariates. We analyze two approaches that have been used in previous studies, and we propose a new approach that examines the role of one covariate while holding the marginal distribution of the other covariates constant. We illustrate the differences between the methods with a numerical example and an empirical analysis of black–white wage differentials among males. Copyright © 2015 John Wiley & Sons, Ltd.

Received 29 July 2012; Revised 29 October 2014;

1. INTRODUCTION

Many studies have adopted the use of reweighting methods to analyze the role of observable covariates for ‘predicting’ or ‘explaining’ outcome differences across groups or over time. DiNardo et al.’s seminal study (1996; DFL hereafter) developed and applied a reweighting method to assess how changes in the distribution of observed worker characteristics contributed to increases in wage inequality during the 1980s. Subsequently, researchers have applied reweighting estimators to analyze between-group outcome differences in a variety of contexts (see, for example, Biewen, 2001; Barsky et al., 2002; Chiquiar and Hanson, 2005; Machado and Mata, 2005; Melly, 2005; Elder et al., 2011; Fortin et al., 2011; Altonji et al., 2012).1

In these applications, a primary focus has been the estimation of overall predicted group differences, which involves assessing how much of the difference in the distributions of outcomes across groups can be predicted by differences in the distributions of all covariates. Somewhat less attention has been paid to methods for isolating the roles of individual covariates (or other subsets of the full covariate set), although this question is often of interest.2 For example, one of the key questions raised in DFL was ‘How much of the increase in wage inequality during the 1980s can be accounted for by changes in the prevalence of unionization?’

In this paper, we examine three approaches to isolating the roles of individual covariates in reweighting estimation; one used by Machado and Mata (2005), one proposed by Fortin et al. (2011;
FLF hereafter) and an approach we propose. Our approach can be viewed as a generalization of the methods of Oaxaca (1973) and Blinder (1973) in which the role of one covariate is examined while holding the other covariates ‘constant’. Specifically, our method is designed to consider the effect of changing the distribution of one set of covariates, while holding constant the distribution of the other covariates and the dependence structure among all covariates. When the underlying model of outcomes is linear, the implied roles of individual covariates for mean outcome differences based on our approach are equivalent to those found using least-squares-based methods, which is not generally true of the other two reweighting approaches. Our method assumes that all covariates take discrete values. We demonstrate our methods with an empirical example based on wage differences between black and white men.

2. REWEIGHTING BASED ON A FULL SET OF COVARIATES

In this section, we first introduce our notation for reweighting estimation, and then we introduce an empirical example to focus ideas.

2.1. A reweighting framework

Assume that there are two groups, denoted by \( j \in \{ A, B \} \), and suppose that a researcher is interested in examining the difference between these groups in the density of an outcome \( y \). Consider a model in which the density of \( y \) in each group is related to a vector of covariates \( w \). To use a reweighting approach to assess how the group-level difference in the distribution of \( w \) contributes to the group-level difference in the density of \( y \), one could reweight group \( A \) to have the distribution of \( w \) found in group \( B \). As DFL show, the resulting reweighted density of \( y \) corresponds to the counterfactual density that would hold if group \( A \) had group \( B \)’s distribution of \( w \) but its own mapping from \( w \) to \( y \). Letting \( j_{yw} \) denote the group whose mapping from \( w \) to \( y \) is used and \( j_w \) the group whose distribution of \( w \) is used, the counterfactual density can be written as

\[
f(y|j_{yw} = A, j_w = B) = \int_w f(y|j_{yw} = A, w) dF(w|j = B)
\]

\[
= \int_w f(y|j_{yw} = A, x) \psi(w) dF(w|j = A)
\]

where \( \psi(w) = \frac{dF(w|j = B)}{dF(w|j = A)} \).

The equality in the third line of equation (1) assumes a ‘common support’ condition that requires all values of \( w \) observed in group \( B \) to also be observed in group \( A \); if this condition does not hold, it is impossible to reweight group \( A \) to have the distribution of \( w \) found in group \( B \). All reweighting methods require some form of support assumption, an issue we consider in more detail below. The third line of equation (1) shows that the counterfactual density can be obtained by reweighting group \( A \), with weights \( \psi(w) \) that depend only on the values of the covariates.

---

3 One might instead reweight group \( B \) to match the distribution of \( w \) found in group \( A \), and this alternative choice of reference group would typically lead to different inferences because of differences across groups in the mapping from covariates to outcomes. Because we focus on issues related to differences in the distribution of covariates, for simplicity we maintain that group \( A \) is the reference group throughout.

4 See Nopo (2008) for a detailed analysis of support issues and a constructive approach to dealing with a lack of support in the case of reweighting using the full set of covariates.
In practice, the weights \( w \) are usually constructed using a substitution that follows from Bayes’ rule:

\[
\psi(w) = \frac{\frac{dF(y|j=B)}{dF(y|j=A)}}{Pr(j=B|w)/Pr(j=A|w)} = \frac{Pr(j=B|w)/Pr(j=A|w)}{Pr(j=B)/Pr(j=A)} \tag{2}
\]

This expression implies that the weights can be calculated from estimated probabilities of group membership conditional on \( w \). Once these weights are obtained, the calculated contribution of between-group differences in all observable characteristics, sometimes referred to as the ‘aggregate decomposition’, is the difference between the actual density of outcome \( y \) for group \( A, f(y|j=A) \), and the counterfactual density \( f(y; j=yw=A, jw=B) \). Hereafter, we denote the reweighted density of \( y \) using weights \( \omega \) as \( f(y; \omega | j=A) \), so \( f(y; j=yw=A, jw=B) \) may be equivalently rewritten as \( f(y; \omega | j=A) \).

One may calculate any statistic based on this counterfactual distribution and compare it to the analogous statistic based on the factual group \( A \) distribution.

### 2.2. An application to the black–white male wage gap

To focus these ideas, consider the well-known wage gap between black and white males. A large literature has established that some of the gap is predictable based on differences in characteristics such as education, labor market experience, marital status, industry and occupation.\(^5\) We analyze the black–white wage gap among males aged 25–59 employed in the civilian labor force, using the 1% PUMS file of the 2000 Census data distributed by IPUMS USA (Ruggles et al., 2010). We define group \( A \) to be white males and group \( B \) to be black males. The outcome measure \( y \) is the logarithm of average hourly wages (annual earnings divided by usual hours worked per week and by weeks worked last year), and we analyze the roles of five covariates: education, experience, marital status, industry and occupation. Here and elsewhere in the paper, each covariate is represented by one or more indicator variables, e.g. education is represented by indicators for <12 years, 12 years, 13–15 years and \( \geq 16 \) years.\(^6\) Given the large sample sizes available in the 2000 PUMS data, we use a 10% random sample and exclude observations with missing data on any of the five covariates or wages. We also exclude observations with hourly wages below $1 and above $3000. Our final analysis sample consists of 213,908 observations for whites and 23,945 observations for blacks.

Table I shows our basic sample information. The first two columns in the table show sample means of log wages and each of the covariates, separately by race. Average log wages are 0.240 higher among whites than among blacks. Whites are also more highly educated, more likely to be married, and more likely to work in relatively high-wage industries and occupations.

To illustrate how reweighting methods work, the third column presents means of the counterfactual white sample in which whites are reweighted to have the black distribution of all covariates. As is readily apparent, full reweighting gives whites very similar characteristics to blacks.\(^7\) Taken together, these characteristics predict about 59% of the overall mean log wage gap (= (2.822 – 2.681) / (2.822 – 2.582)).

Of course, if we were only interested in mean differences, then flexibly specified linear regression models would suffice. An advantage of reweighting methods is that differences in the entire

---

\(^5\) Altonji and Blank (1999) provide a detailed review of this literature.

\(^6\) Potential experience is defined as age minus 6 minus the number of years of education (using the algorithm of Angrist et al., 2011, to define years of education) and then grouped into categories of <10 years, 10–14 years, 15–19 years, 20–24 years, 25–29, years, 30–34 years and \( \geq 35 \) years. Marital status is represented by a binary indicator. Occupation is represented by the six broad categories used in the 2000 Census. Industry is represented by eight categories, which are taken from the 20 sectors defined in the 1997 North American Industry Classification System (see Table I).

\(^7\) The reweighted white means are not identical to the black means because the model of group membership used to generate weights is not fully saturated, i.e. it includes all of the covariates in Table I but not interactions among them. We estimate this group membership model as a logit.

---

Table I. Unweighted and weighted sample means, 2000 Census

<table>
<thead>
<tr>
<th></th>
<th>Unweighted</th>
<th>Reweighting method</th>
<th>(z = education)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whites</td>
<td>Blacks</td>
<td>Full</td>
</tr>
<tr>
<td>Log hourly wage</td>
<td>2.822</td>
<td>2.582</td>
<td>2.681</td>
</tr>
<tr>
<td>Years of education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>0.085</td>
<td>0.127</td>
<td>0.128</td>
</tr>
<tr>
<td>12</td>
<td>0.318</td>
<td>0.405</td>
<td>0.404</td>
</tr>
<tr>
<td>13-15</td>
<td>0.299</td>
<td>0.310</td>
<td>0.310</td>
</tr>
<tr>
<td>16+</td>
<td>0.298</td>
<td>0.158</td>
<td>0.158</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management/professional</td>
<td>0.335</td>
<td>0.193</td>
<td>0.192</td>
</tr>
<tr>
<td>Service</td>
<td>0.092</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>Sales/office</td>
<td>0.160</td>
<td>0.160</td>
<td>0.159</td>
</tr>
<tr>
<td>Farming/fishing/forestry</td>
<td>0.011</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>Construction/maintenance</td>
<td>0.185</td>
<td>0.150</td>
<td>0.150</td>
</tr>
<tr>
<td>Production/transportation</td>
<td>0.218</td>
<td>0.314</td>
<td>0.315</td>
</tr>
<tr>
<td>Married</td>
<td>0.713</td>
<td>0.582</td>
<td>0.582</td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture/mining/construction</td>
<td>0.165</td>
<td>0.117</td>
<td>0.117</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.215</td>
<td>0.202</td>
<td>0.202</td>
</tr>
<tr>
<td>Trade</td>
<td>0.147</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td>Transportation and warehousing</td>
<td>0.063</td>
<td>0.102</td>
<td>0.103</td>
</tr>
<tr>
<td>FIRE</td>
<td>0.083</td>
<td>0.075</td>
<td>0.075</td>
</tr>
<tr>
<td>Management/administration</td>
<td>0.087</td>
<td>0.084</td>
<td>0.083</td>
</tr>
<tr>
<td>Education/health care</td>
<td>0.152</td>
<td>0.192</td>
<td>0.193</td>
</tr>
<tr>
<td>Arts and other services</td>
<td>0.086</td>
<td>0.105</td>
<td>0.105</td>
</tr>
<tr>
<td>Potential experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>0.126</td>
<td>0.128</td>
<td>0.125</td>
</tr>
<tr>
<td>10–14</td>
<td>0.147</td>
<td>0.164</td>
<td>0.162</td>
</tr>
<tr>
<td>15–19</td>
<td>0.163</td>
<td>0.181</td>
<td>0.180</td>
</tr>
<tr>
<td>20–24</td>
<td>0.174</td>
<td>0.174</td>
<td>0.175</td>
</tr>
<tr>
<td>25–29</td>
<td>0.163</td>
<td>0.149</td>
<td>0.151</td>
</tr>
<tr>
<td>30–34</td>
<td>0.126</td>
<td>0.110</td>
<td>0.112</td>
</tr>
<tr>
<td>35+</td>
<td>0.101</td>
<td>0.093</td>
<td>0.094</td>
</tr>
</tbody>
</table>

Note: The means in the column labeled ‘Full’ are based on weights contructed from a logit model of group membership as a linear function of all of the covariates. N = 213,908 for whites and reweighted whites; N = 23,945 for blacks.

Figure 1. Log wage densities: whites, blacks and reweighted whites
distribution of $y$ can be examined. As an illustration, Figure 1 plots the densities of log wages for whites, blacks and reweighted whites. Reweighting shifts the white density to the left towards the black density, but the magnitude of the shift is not uniform across the support. Specifically, the density appears to shift more in the upper tail than in the lower tail. To see this, note that the reweighted white density essentially matches the black density at values above roughly 3.75, but the reweighted white density lies far from the black density (and relatively close to the unweighted white density) below roughly 1.5. These patterns imply that the covariates can explain more of the black–white differences in the upper tail of the distribution than they can in the lower tail.

Up to this point, we have been considering the joint role of all covariates. However, given that much of the black–white wage gap can be explained by these five variables, an obvious additional question arises: What roles do each of these covariates play in the closing of the wage gap? The rest of the paper focuses on approaches to answering this question.

### 3. ISOLATING THE ROLES OF INDIVIDUAL COVARIATES

Our goal is to isolate the role of a particular covariate (or set of covariates) in the context of reweighting estimators. Specifically, we consider the extent to which one set of observable characteristics predict between-group differences in the distribution of an outcome, while holding the rest of the observable characteristics ‘constant’.

#### 3.1. A linear model

Before considering how to isolate the role of an individual covariate with reweighting methods, we first illustrate what this means in the context of a linear model. Consider a simple data-generating process in which $y$ is linearly related to two sets of covariates, denoted $z$ and $x$ (both of which are assumed to be scalars for this example), and that between-group differences in the mean of $y$ can arise due to differences in the means of those covariates and to an intercept shift. Specifically, suppose

$$y_i = \beta_0 + \beta_d d_i + \beta_z z_i + \beta_x x_i + u_i$$

(3)

where $d_i$ equals 1 if individual $i$ is a member of group $A$ and zero otherwise, and $u_i$ is an error term that is orthogonal to $d_i$, $z_i$ and $x_i$.\(^8\) In this case, the group difference in the expectation of $y$ can be written as

$$E(y \mid j = A) - E(y \mid j = B) = \beta_d + \beta_z [E(z \mid j = A) - E(z \mid j = B)] + \beta_x [E(x \mid j = A) - E(x \mid j = B)]$$

(4)

where $\beta_d$ represents the component that is unrelated to the two covariates, and the role of each covariate is defined in an intuitive way. Standard ordinary least squares (OLS)-based methods estimate the roles of covariates in this way. For example, Altonji and Blank’s (1999) influential study includes such a decomposition to examine the ‘differences due to characteristics’ between black and white log wages (p. 3159).

---

\(^8\) When the coefficients $\beta_z$ and $\beta_x$ vary across groups, decompositions of between-group mean differences into the roles attributable to individual covariates are not unique. See Oaxaca (1973) and Blinder (1973) for early discussions of this issue and Elder et al. (2010) for a more recent contribution. The central findings presented below are unaffected if $\beta_z$ and $\beta_x$ vary across groups, so we impose that they are group invariant for simplicity.
3.2. Isolating the role of covariates in reweighting estimation

Again, let $z$ denote a particular covariate (or set of covariates) from the vector $w$, and let $x$ denote the vector of remaining covariates. We consider three methods for isolating the role of $z$, two used previously in the literature and a third we propose here.

Method 1 (‘First In’)
Machado and Mata (2005) use an approach that reweights using only the covariate $z$. This approach answers the question: ‘What would be group A’s density of $y$ if it had group B’s distribution of $z$ but its own distribution of $x$ conditional on $z$?’

In the context of reweighting, this approach identifies the counterfactual density:

$$f(y; j; x, z) = A, j \neq A; j = B)$$

$$= \int \int f(y \mid j = A; z, x) dF(x \mid z, j = A) dF(z \mid j = B)$$

$$= \int \int f(y \mid j = A; z, x) dF(x \mid z, j = A) \psi_{\text{rel}}(z, x) dF(z \mid j = A)$$

where $\psi_{\text{rel}}(z, x) = \frac{dF(z \mid j = B)}{dF(z \mid j = A)}$.

As is apparent from the weighting function $\psi_{\text{rel}}(z, x)$, this approach involves reweighting group A in order to match the marginal distribution of $z$ observed in group $B$. One then assesses the role of $z$ by comparing $f(y; j = A)$ to $f(y; \psi_{\text{rel}}; j = A)$.

As FLF point out, however, this approach has a serious drawback: to the extent that $z$ is correlated with elements of $x$, the method attributes to $z$ both the effect of differences between groups in $z$ and the effect of differences in $x$ that are correlated with $z$. As an illustration, assume that $z$ represents marital status, that married workers have greater educational attainment in both groups than do non-married workers, and that workers in group $A$ are both more likely to be married and more highly educated than workers in group $B$. Then, reweighting workers in group $A$ to match group $B$’s marriage rate—by assigning relatively large weights to non-married workers and relatively small weights to married workers—also shifts the educational distribution of group $A$. The resulting counterfactual population of workers is both less likely to be married and less educated than workers in group $A$, so the estimated effect of marital status will also capture some of the effect of education.

Method 2 (‘Last In’)
FLF propose an approach that is also a variant of standard reweighting methods, but instead is based on excluding the covariate $z$. This approach answers the question: ‘What would be group A’s density of $y$ if it had group B’s distribution of $z$ conditional on $x$ but its own distribution of $x$?’,

which corresponds to the counterfactual density

---

9 Machado and Mata’s approach differs from DFL’s in other ways, including the use of quantile regression to estimate parametric models of the mapping from covariates to outcomes. We adapt their approach to a reweighting context for ease of comparison.

10 From this point forward, we leave implicit that the counterfactual population retains group A’s mapping from $\{z, x\}$ to $y$, which will be the case for all of the counterfactuals described below.

11 FLF write that this method is invalid as a way of performing the decomposition for the same reason that [an OLS regression-based] decomposition would be invalid if the coefficient used for one covariate was estimated without controlling for the other covariates’ (p. 61).

12 FLF also compare a reweighting approach to the recentered influence function-based approach they develop in their earlier work (Firpo et al., 2009), noting that both approaches can be used to isolate the role of individual covariates. Antecol et al. (2008) also use FLF’s reweighting approach in assessing the role of occupational sorting in the sexual orientation wage gap.
The weights in equation (6) can be calculated by dividing weights produced using the full \( w \) vector by weights that use all covariates except \( z \). One then assesses the role of \( z \) by comparing \( f(y|j=A) \) to \( f(y; \psi^z|j=A) \).

In our example, in which \( z \) represents marital status, the resulting counterfactual population has the same distribution of education (and all other covariates in \( x \)) found in group \( A \), so this approach does not mistakenly attribute to marital status an effect that is instead due to other characteristics that are correlated with marital status. However, this method has its own limitation: although the counterfactual distribution of marital status does not match group \( B \)’s distribution of marital status conditional on the other covariates, its marginal distribution of marital status does not match group \( B \)’s (its marriage rate is higher than that found in group \( B \)). More generally, the counterfactual population’s marginal distribution of \( z \) will match that of group \( B \) in only two cases: when the two groups have identical marginal distributions of \( x \), or when \( z \) and \( x \) are independent in group \( B \). This is an unappealing feature of Method 2 if the underlying goal is to assess what would happen to the distribution of outcomes if group \( A \) had the same distribution of \( z \) found in group \( B \).

**Method 3**

We propose a method that is constructed to replicate the marginal distribution of \( z \) found in group \( B \)  while retaining the marginal distribution of \( x \) found in group \( A \). This approach corresponds to the following question: “What would be group \( A \)’s density of \( y \) if it had group \( B \)’s marginal distribution of \( z \) but its own marginal distribution of \( x \)?”

For now, we assume the satisfaction of a support condition: every value of \( z \) that is observed in the population is also observed in group \( A \) in combination with every value of \( x \) observed in group \( A \):

\[
dF(z, x \mid j = A) > 0 \quad \forall z, x \text{ for which } dF(z) > 0 \text{ and } dF(x \mid j = A) > 0
\]

(7)

We propose weights of the following form:

\[
\psi^3(z, x) = \frac{dF(z \mid j = B) - dF(z \mid j = A) + dF(z \mid x, j = A)}{dF(z \mid x, j = A)}
\]

(8)

which produces the counterfactual density

\[
f(y; \psi^3|j = A) = \int \int f(y \mid j = A, z, x)\psi^3(z, x)dF(z \mid x, j = A)dF(x \mid j = A)
\]

(9)

One can then assess the role of \( z \) by comparing \( f(y|j=A) \) to \( f(y; \psi^3|j=A) \). As we show in Appendix A, the \( \psi^3(z, x) \) weights produce a counterfactual distribution that matches group \( B \)’s marginal distribution of \( z \) and group \( A \)’s marginal distribution of \( x \), as was our goal.
We make four additional remarks concerning Method 3. First, because the marginal distributions of \( z \) and \( x \) do not uniquely define their joint distribution, the question we posed in defining Method 3 has multiple answers. The \( \psi^3(z, x) \) weights are designed to maintain other features of group \( A \)'s distribution of the covariates, namely, both the marginal distribution of \( x \) and the dependence structure between \( z \) and \( x \). For example, if \( z \) and \( x \) are scalar random variables, then the \( \psi^3(z, x) \) weights maintain group \( A \)'s covariance between \( z \) and \( x \). More generally, a natural way to 'preserve a dependence structure' would involve preserving \( dF(z \mid x) \) for all \( x \), but this is infeasible if \( dF(z) \) changes while \( dF(x) \) does not. Instead, our counterfactual preserves the relative values of \( dF(z \mid x) \) across different values of \( x \); specifically, reweighting group \( A \) using weights defined by \( \psi^3(z, x) \) produces a counterfactual that matches group \( A \)'s value of \( [dF(z \mid X = x) - dF(z \mid X = x')] \) for all values of \( z \), \( x \) and \( x' \) (see Appendix B).

We describe the implications of this property in Sections 3.3 and 3.4 below. Thus we view our proposed method as one that most closely approximates the typical 'all else equal' goal of empirical research when examining the effect of a change in one covariate.

Second, unlike in Methods 1 and 2, applying Bayes' rule to rewrite \( \psi^3(z, x) \) does not yield a simple function of only the probabilities of group membership. Specifically, the weights involve marginal and conditional densities of \( z \) that must be estimated directly (even after applying Bayes' rule). Our method assumes that all covariates take discrete values, which allows us to use cell frequencies to estimate these densities. Despite this restriction, we believe our methods are useful because many economic applications involve only discrete-valued covariates.

Third, because \( [dF(z \mid j = B) - dF(z \mid j = A) + dF(z \mid x, j = A)] \) is not guaranteed to be non-negative for every \( \{z, x\} \) combination, it is possible for some weights to be negative. If there are negative weights, the marginal distributions of \( z \) and \( x \) will still be appropriate, but the counterfactual joint distribution will not be well defined. We discuss this point in Section 4.

Fourth, we note a similarity between Method 3 and that of Rothe (2012), who is also interested in the effect on an outcome of a 'ceteris paribus' change in the distribution of one covariate. Rothe operationalizes the idea of maintaining the dependence structure by considering counterfactuals in which the distribution of the other covariates is held constant and the joint distribution of covariate ranks is maintained. However, when covariates are discrete, as in our case, their ranks are not uniquely defined. As a result, Rothe's approach applied to discrete covariates places bounds on the outcome change of interest rather than providing a unique estimate. Whenever our \( \psi^3(z, x) \) weights produce a valid counterfactual, our estimate of the outcome change will be within Rothe's bounds; see Appendix B for further details.

Finally, it is useful to consider an alternative set of weights:

\[
\psi^{3\ast}(z, x) = \frac{dF(z \mid j = B)}{dF(z \mid x, j = A)}
\]  

These weights produce marginal distributions of covariates that are identical to those produced by the weights defined in equation (8). However, they do not retain the dependence structure between \( z \) and \( x \),

---

13 In principle, our approach could be extended to deal with continuous covariates. However, violations of condition (7) are likely to be more prevalent in the case of continuous covariates.

14 If \( z \) and \( x \) were continuous, negative weights would be guaranteed to arise in many applications. For example, suppose that \( z \) and \( x \) are bivariate normal and that the entire distribution of \( z \) is shifted by some constant in group \( B \) relative to group \( A \). Then there will always be values of \( z \) such that \( dF(z j = B) - dF(z j = A) < 0 \), and if \( z \) and \( x \) are not independent in \( A \), corresponding values of \( x \) such that \( dF(z x, j = A) \) is arbitrarily small.

15 We thank the anonymous referees for encouraging us to explore the connections between Rothe’s approach and ours.

16 As a demonstration of this result, consider Rothe’s empirical example in which he examines the effect of the change in union status on the distribution of log hourly wages between 1983–85 and 2003–05. He reports the following bounds: \([-0.038, -0.004] \) at the mean, \([-0.037, -0.009] \) at the 10th percentile, \([-0.062, -0.024] \) at the 50th percentile and \([-0.006, 0.027] \) at the 90th percentile. Using Rothe’s data, Method 3 delivers the following estimates, all of which are contained within those bounds: \(-0.026 \) at the mean, \(-0.023 \) at the 10th percentile, \(-0.043 \) at the 50th percentile and 0.007 at the 90th percentile.
but instead produce counterfactual distributions in which \( z \) and \( x \) are independent. In models in which the covariates are separable in their effects on \( y \), such as in the linear model given by equation (3), both sets of weights will deliver identical results regarding the roles of covariates for differences in the mean of \( y \). More generally, however, the dependence structure between \( z \) and \( x \) matters, and a counterfactual that imposes independence of \( z \) and \( x \) may not be suitable to isolate a ceteris paribus change.

### 3.3. An illustrative example

While the motivation for using reweighting methods is strongest when one is interested in the full distribution of an outcome in a nonlinear environment, it is helpful to illustrate the differences between the methods in the context of the simple linear model given by equation (3). Consider a case in which \( z \) and \( x \) are both binary, with the joint distributions of \( z \) and \( x \) given by the probabilities in the top two panels in Table II, \( \beta_0 = \beta_d = 0 \), and \( \beta_z = \beta_x = 1 \). The two covariates are positively correlated in both groups, but the mean of each covariate is 0.2 higher in group \( A \). It is readily verified that the mean of the outcome variable \( y \) is 1.2 in group \( A \) and 0.8 in group \( B \). The application of equation (4) then implies that the role of \( z \) for the group difference in \( E(y) \) is 0.2 \[= 1 \times (0.6 - 0.4)\], as is the role of \( x \).

In the bottom three panels of Table II, we apply the three reweighting methods to estimate the role of \( z \). Each cell entry in these panels is simply the corresponding group \( A \) probability multiplied by the relevant weight for that cell, calculated using the equations in Section 3.2. For example, the Method 1 weight applied to the \((z = 0, x = 0)\) cell equals \( P(z = 0 | j = B) \times P(z = 0 | j = A) = 0.6 / 0.4 \), so the resulting cell entry equals 0.375 \((= 0.25 \times 0.6 / 0.4)\). The role of \( z \) is then calculated by subtracting the mean of \( y \) in the reweighted group \( A \) from the mean of \( y \) in the actual group \( A \).

Turning to the results for Method 1, the implied role of \( z \) is 0.275 \((= 1.2 - 0.925)\), rather than the 0.2 estimate based on equation (4). To see why, note that the marginal distribution of group \( z \) in the reweighted distribution matches that found in group \( B \), as was the goal, but the marginal distribution

---

**Table II. An illustrative example of applying the three reweighting methods**

<table>
<thead>
<tr>
<th></th>
<th>( \Pr[z = 0] )</th>
<th>( \Pr[z = 1] )</th>
<th>( \Pr[z = 0 \text{ or } z = 1] )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actual Group A joint distribution</strong> (( E(y) = 1.2 ))</td>
<td>( 0.25 )</td>
<td>0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>( \Pr[x = 0] )</td>
<td>0.15</td>
<td>0.45</td>
<td>0.60</td>
</tr>
<tr>
<td>( \Pr[x = 1] )</td>
<td>0.40</td>
<td>0.60</td>
<td>1</td>
</tr>
<tr>
<td><strong>Actual Group B joint distribution</strong> (( E(y) = 0.8 ))</td>
<td>( 0.45 )</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>( \Pr[x = 0] )</td>
<td>0.15</td>
<td>0.25</td>
<td>0.40</td>
</tr>
<tr>
<td>( \Pr[x = 1] )</td>
<td>0.60</td>
<td>0.40</td>
<td>1</td>
</tr>
<tr>
<td><strong>Reweighted Group A joint distribution using Method 1</strong> (( E(y) = 0.925 ))</td>
<td>( 3/2 \times 0.25 = 0.375 )</td>
<td>( 2/3 \times 0.15 = 0.10 )</td>
<td>0.475</td>
</tr>
<tr>
<td>( \Pr[x = 0] )</td>
<td>( 3/2 \times 0.15 = 0.225 )</td>
<td>( 2/3 \times 0.45 = 0.30 )</td>
<td>0.525</td>
</tr>
<tr>
<td>( \Pr[x = 0 \text{ or } x = 1] )</td>
<td>0.60</td>
<td>0.40</td>
<td>1</td>
</tr>
<tr>
<td><strong>Reweighted Group A joint distribution using Method 2</strong> (( E(y) = 1.075 ))</td>
<td>( 6/5 \times 0.25 = 0.30 )</td>
<td>( 2/3 \times 0.15 = 0.10 )</td>
<td>0.40</td>
</tr>
<tr>
<td>( \Pr[x = 0] )</td>
<td>( 3/2 \times 0.15 = 0.225 )</td>
<td>( 5/6 \times 0.45 = 0.375 )</td>
<td>0.60</td>
</tr>
<tr>
<td>( \Pr[x = 0 \text{ or } x = 1] )</td>
<td>0.525</td>
<td>0.475</td>
<td>1</td>
</tr>
<tr>
<td><strong>Reweighted Group A joint distribution using Method 3</strong> (( E(y) = 1.0 ))</td>
<td>( 33/25 \times 0.25 = 0.33 )</td>
<td>( 7/15 \times 0.15 = 0.07 )</td>
<td>0.40</td>
</tr>
<tr>
<td>( \Pr[x = 0] )</td>
<td>( 45/25 \times 0.15 = 0.27 )</td>
<td>( 11/15 \times 0.45 = 0.33 )</td>
<td>0.60</td>
</tr>
<tr>
<td>( \Pr[x = 0 \text{ or } x = 1] )</td>
<td>0.60</td>
<td>0.40</td>
<td>1</td>
</tr>
</tbody>
</table>

**Note:** For the reweighted joint distributions, we additionally list the weight that is computed using the appropriate formula in Section 3.1 and the relevant actual group \( A \) joint probability; multiplying these quantities together gives the reweighted joint probability.
of \( x \) has shifted relative to group \( A \): \( P(x = 1) \) has fallen by 0.075, from 0.6 to 0.525. This shift in \( P(x = 1) \) is then incorrectly attributed to the role of \( z \) (because \( \beta_1 = \beta_2 = 1 \), shifts in \( P(x = 1) \) translate directly into differences in the mean of \( y \)).

Turning to the results for Method 2, the implied role of \( z \) is now 0.125 (= 1.2 - 1.075), which is less than the 0.2 estimate based on equation (4). To see why, note that the reweighted group \( A \) now preserves the actual group \( A \) distribution of \( x \), but it does not produce group \( B \)'s marginal distribution of \( z \): \( P(z = 1) \) in the reweighted group \( A \) distribution is 0.475 rather than the target value of 0.40, and this difference of 0.075 is the same as the difference between the implied role of \( z \) based on equation (4) and that based on Method 2 (i.e., 0.2 vs. 0.125).

Finally, we turn to the results for Method 3 in the bottom panel of Table II. The implied role of \( z \) is now 0.2 (= 1.2 - 1.0), the same as the implied role based on equation (4). This equivalence follows from the fact that the reweighted group \( A \) distribution now matches group \( B \)'s distribution of \( z \) and group \( A \)'s distribution of \( x \). Moreover, this counterfactual distribution maintains group \( A \)'s dependence structure between \( z \) and \( x \), in that it preserves group \( A \)'s value of \( [dF(z \mid X = x) - dF(z \mid X = x')] \) for all \( z \), \( x \) and \( x' \). Specifically, because \( P(z = 0 \mid x) \) increases by the same amount for all values of \( x \) (by 0.2, from 0.625 to 0.825 for \( x = 0 \) and from 0.25 to 0.45 for \( x = 1 \)), \( P(z = 0 \mid x = 1) - P(z = 0 \mid x = 0) \) remains constant at 0.375. Moreover, the covariance between \( z \) and \( x \) equals 0.09 in both distributions.

In summary, only Method 3 yields correct estimates of the roles of \( z \) and \( x \) in this linear framework. This property is not merely a property of the DGP used here, but is true of all linear models. Moreover, as we show in the following section, Method 3 produces appealing estimates of the role of \( z \) even when the effects of \( z \) and \( x \) on the distribution of \( y \) are non-separable.

### 3.4. Preserving the dependence structure between \( z \) and \( x \)

As discussed above, the \( \psi^3(z, x) \) weights preserve the dependence structure between \( z \) and \( x \) by producing a counterfactual that matches group \( A \)'s value of \( [dF(z \mid X = x) - dF(z \mid X = x')] \) for all values of \( z \), \( x \) and \( x' \). To illustrate the implications of doing so, consider again a model involving an outcome \( y \) and explanatory variables \( z \) and \( x \), but one that includes a full set of interactions of indicator variables for each combination of values of \( z \) and \( x \). In addition, assume that condition (7) holds throughout and that all weights are non-negative.

Recognizing explicitly that the supports of \( z \) and \( x \) are discrete, the role of \( z \) in explaining the between-group difference of the density of \( y \) that is estimated by Method 3 is given by the expression \( f(y \mid j = A) - f(y; \psi^3 \mid j = A) \), where \( f(y; \psi^3 \mid j = A) \) is the counterfactual density of \( y \):

\[
    f(y; \psi^3 \mid j = A) = \sum_z \sum_x \left[ f(y \mid z, x, j = A)\psi^3(z, x)P(z, x \mid j = A) \right] (11)
\]

In Appendix C we show that

\[
    f(y \mid j = A) - f(y; \psi^3 \mid j = A) \\
    = \sum_x \left[ P(x \mid j = A)\sum_z f(y \mid z, x, j = A)(P(z \mid j = A) - P(z \mid j = B)) \right] (12)
\]

Thus the role of \( z \) delivered by Method 3 is a weighted sum of the difference in the probabilities of \( z \) in the two groups (i.e. \( P(z \mid j = A) - P(z \mid j = B) \)), where these probability differences are weighted by the conditional densities in group \( A \) (i.e. \( f(y \mid z, x, j = A) \)), which are themselves averaged across the marginal distribution of \( X \) in group \( A \) (i.e. \( P(x \mid j = A) \)).

The expression in equation (12) extends immediately to expectations, and it further simplifies when \( z \) is binary (see Appendix C):
Thus the role of $z$ in explaining the between-group difference in the expectation of $y$ is the product of the between-group difference in the mean of $z$ and the expected value of the marginal effect of $z$ on $y$ in group $A$, with that expectation evaluated over the marginal distribution of $x$. In other words, because we are now considering a model with a full set of interactions between $z$ and $x$, ‘the’ marginal effect of $z$ on $y$ no longer has a unique answer, but instead depends on the distribution of $x$. Method 3 resolves this ambiguity by using the average marginal effect over the distribution of $x$ observed in group $A$, which is arguably the most natural choice. Weights that produce the appropriate marginal distributions but do not maintain the dependence structure, such as those in equation (10), generally lead to implied ‘roles of $z$’ that are not based on average marginal effects.

4. EMPIRICAL ISSUES WHEN IMPLEMENTING METHOD 3

There are two primary empirical issues that can arise when implementing Method 3. The first is violations of the support condition given by equation (7), which we have assumed holds to this point. For simplicity, rather than working with our preferred $\psi^{3*}(z,x)$ weights, we illustrate the importance of the support condition using the simpler weights $\psi^{3*}(z,x)$ described in equation (10), and we again adopt notation that explicitly recognizes that the supports of $z$ and $x$ are discrete. Extension to the $\psi^{3*}(z,x)$ weights is discussed in Appendix D.

If condition (7) does not hold, so that there exist values of $z$ and $x$ for which $P(z) > 0$ and $P(x | j = A) > 0$ but $P(z, x | j = A) = 0$, then $\psi^{3*}(z,x)$ may be undefined for some $\{z, x\}$ combinations. Even though these combinations do not appear in group $A$, implying that the undefined weights do not need to be computed, they still present problems. Moreover, this support condition is stronger than the analogous support condition needed for full reweighting.17

To describe the problem, we note that the marginal density of $z$ in group $j$ is given by

$$P(z | j) = \sum_x P(z, x | j)$$

(14)

and the marginal density of $x$ in group $j$ is given by

$$P(x | j) = \sum_z P(z, x | j)$$

(15)

Now consider the reweighted version of the right-hand side of equation (15) for group $A$. For any value of $x$ we may confine the sum to values of $z$ for which $P(z, x | j = A) > 0$; at these values, the weights are always defined, and it is straightforward to show that

$$\sum_z \psi^{3*}(z,x)P(z, x | j = A)$$

$$= \sum_{z: P(z, x | j = A) > 0} \frac{P(z | j = B)}{P(z, x | j = A)} \frac{P(z, x | j = A)}{P(x | j = A)}$$

$$= P(x | j = A) \times \sum_z \{ P(z | j = B) \times 1(P(z, x | j = A) > 0) \}$$

(16)

17 The support condition for full reweighting is weaker in the sense that it requires only common support, i.e. that a configuration of characteristics that appears in group $B$ (the target population) also appears in group $A$ (the population being reweighted). In contrast, the support condition in equation (7) requires that every value of $z$ must appear in combination with every value of $x$ in the population being reweighted. This condition can fail even if support is the same in both populations.
When condition (7) holds, so that \( P(z, x | j = A) > 0 \) for all values of \( z \) and \( x \), this expression trivially equals \( P(x | j = A) \) because \( P(z | j = B) \) sums to 1 across all values of \( z \). When condition (7) does not hold, however, the value of the sum in the last line of equation (16) is less than 1, so equation (16) does not produce the appropriate counterfactual density \( P(x | j = A) \). By similar reasoning, when equation (7) does not hold, using the \( \psi^{-3} \) weights also does not produce the appropriate counterfactual density \( P(z | j = B) \).

Diagnosing support issues is straightforward because it can be examined directly. The support condition in equation (7) requires that every value of \( z \) must appear in combination with every value of \( x \) in the group being reweighted. Not only will appropriate cross-tabulations for group A (the group that is to be reweighted) help one assess whether the support condition is met, but it will also directly show what variable categories are problematic.

Fortunately, two potential solutions exist. One solution is to redefine the covariates into coarser groupings so that condition (7) is satisfied. Another solution follows from considering the last line of equation (16). For each value of \( x \) for which at least one value of \( z \) does not appear, if one multiplies all available weights \( \psi^{-3} (z, x) \) associated with that value of \( x \) by \( \left\{ \frac{\sum P(z | j = B) \times 1(P(z, x | j = A) > 0)}{1} \right\}^{-1} \), then \( P(x) \) in the reweighted population will once again equal \( P(x | j = A) \). Ensuring that all of the target marginal probabilities are matched becomes more complicated when multiple combinations of values of \( z \) and \( x \) are not observed in group A. In Appendix D we describe a more complicated algorithm that builds on these ideas to accomplish this task.

The second empirical issue is that some of the weights can be negative. As noted above, \( \psi^{-3} (z, x) \) weights are not guaranteed to be non-negative, even if the support condition is always satisfied. It is also possible that our algorithm to correct for violations of the support condition (i.e. the second solution above) can produce negative weights. A negative weight implies that the probability of the associated \( \{z, x\} \) combination is negative in the counterfactual distribution, so that the counterfactual is not a proper probability distribution.

Once again, diagnosing this empirical issue is straightforward: one can simply examine the distribution of weights to determine whether any are negative. In our empirical example below, the share of weights that are negative is low (see footnote 21). And once again, two potential solutions exist, although each has its own problems. The first possible solution is to set negative weights to zero and rescale the remaining weights so that they have a mean of one. The second is to use all of the weights, including negative ones. For the first solution, the counterfactual marginal distributions of \( z \) and \( x \) will not match their targets exactly, but the weights produce a well-defined probability distribution and can be used for any counterfactual calculation. For the second solution, the counterfactual marginal distributions of \( z \) and \( x \) will match their targets exactly, but it is possible that the counterfactual distribution of \( y \) is not well defined. The possibility of negative weights is a potentially significant drawback to our approach, and we urge readers to use caution in applying these methods when negative weights arise.

Lastly, one omnibus diagnostic that can alert the practitioner that something is amiss is to inspect the summary statistics for the reweighted group. If the summary statistics do not match those of the target group, then something has gone wrong with the reweighting. The diagnostics listed above can then help the practitioner isolate the exact problems that may exist.

---

18 Because theory is often silent about how covariates should be defined, this solution could be attractive in many settings. In our empirical example below, we were able to eliminate the support problem by reducing the number of categories of occupation from 6 to 5, of industry from 8 to 4, and of potential experience from 7 to 6.

19 The existence of a negative conditional probability could also lead to the existence of conditional probabilities greater than one. Either situation, of course, implies that the counterfactual is not a proper probability distribution. Because it is sufficient to diagnose such problems by searching for negative weights/implied negative probabilities, we focus on this part of the problem.

20 In a related context, Kline (2011) uses negative weights in calculating a counterfactual mean based on a reweighting estimator.
To demonstrate how the various methods work in practice, we return to the black–white wage gap example. The last three columns of Table I present sample means of counterfactual white populations generated by using Methods 1, 2 and 3 to isolate the role of education, i.e. education plays the role of $z$. Method 3 is implemented using the $\psi_3(z, x)$ weights with cell frequencies to estimate all densities. The patterns across the three methods are what we expect to find. Specifically, Method 3 approximately produces the desired distributions of characteristics: the black distribution of education and the white distributions of experience, occupation, industry and marital status. Method 1 matches the desired distribution for education, but not for the other covariates. Method 2 matches the desired distribution for the other covariates, but not for education.

Table III presents the estimates for the roles of the five covariates in predicting the mean black–white log wage gap based on OLS regressions and each of the reweighting methods. For all four methods, the most important covariates are education, marital status and occupation. However, as in our illustrative example, Method 1 produces estimates of contributions that are notably larger than those produced by OLS, and Method 2 produces estimates that are notably smaller. In contrast, Method 3 produces estimates that are most similar to OLS for all covariates (the largest differences from OLS are for education and occupation). Similarly, Method 3 produces a sum of contributions that is quite close to that of OLS, while Method 1’s is considerably larger and Method 2’s is considerably smaller.

Turning to questions that are less well suited to an OLS framework, Table IV shows estimates of various percentiles of the white and black log wage distributions, along with three counterfactual distributions (labeled ‘Reweighted whites’) that isolate the roles of the three variables that have the largest effects at the mean: education, marital status and occupation. All estimates in the three ‘Reweighted whites’ columns are based on Method 3. The column labeled ‘$z =$ education’, for example, shows that black–white educational differences produce a 0.058 difference ($= 2.813 - 2.755$) in the medians of the black and white log wage distributions. The effect of education varies only slightly across the distribution of log wages, judging by the differences between the ‘Unweighted whites’ and

Table III. Contribution of covariates to the mean black–white log wage gap

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Reweighting method</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Education</td>
<td>0.068</td>
<td>0.084</td>
<td>0.039</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.005</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Marital status</td>
<td>0.022</td>
<td>0.035</td>
<td>0.016</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Occupation</td>
<td>0.045</td>
<td>0.083</td>
<td>0.018</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Industry</td>
<td>0.005</td>
<td>-0.001</td>
<td>0.002</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Sum of contributions</td>
<td>0.143</td>
<td>0.206</td>
<td>0.080</td>
<td>0.140</td>
</tr>
</tbody>
</table>

Note: Reweighting methods are described in the text. Each cell lists the estimated contribution of the covariate shown in the first column, with standard errors in parentheses. For the reweighting methods, standard errors are estimated from 200 bootstrap replications. In all cases, $N = 213,908$ for whites and reweighted whites; $N = 23,945$ for blacks.

5. APPLYING THE METHODS TO THE BLACK–WHITE WAGE GAP

To demonstrate how the various methods work in practice, we return to the black–white wage gap example. The last three columns of Table I present sample means of counterfactual white populations generated by using Methods 1, 2 and 3 to isolate the role of education, i.e. education plays the role of $z$. Method 3 is implemented using the $\psi_3(z, x)$ weights with cell frequencies to estimate all densities. The patterns across the three methods are what we expect to find. Specifically, Method 3 approximately produces the desired distributions of characteristics: the black distribution of education and the white distributions of experience, occupation, industry and marital status. Method 1 matches the desired distribution for education, but not for the other covariates. Method 2 matches the desired distribution for the other covariates, but not for education.

Table III presents the estimates for the roles of the five covariates in predicting the mean black–white log wage gap based on OLS regressions and each of the reweighting methods. For all four methods, the most important covariates are education, marital status and occupation. However, as in our illustrative example, Method 1 produces estimates of contributions that are notably larger than those produced by OLS, and Method 2 produces estimates that are notably smaller. In contrast, Method 3 produces estimates that are most similar to OLS for all covariates (the largest differences from OLS are for education and occupation). Similarly, Method 3 produces a sum of contributions that is quite close to that of OLS, while Method 1’s is considerably larger and Method 2’s is considerably smaller.

Turning to questions that are less well suited to an OLS framework, Table IV shows estimates of various percentiles of the white and black log wage distributions, along with three counterfactual distributions (labeled ‘Reweighted whites’) that isolate the roles of the three variables that have the largest effects at the mean: education, marital status and occupation. All estimates in the three ‘Reweighted whites’ columns are based on Method 3. The column labeled ‘$z =$ education’, for example, shows that black–white educational differences produce a 0.058 difference ($= 2.813 - 2.755$) in the medians of the black and white log wage distributions. The effect of education varies only slightly across the distribution of log wages, judging by the differences between the ‘Unweighted whites’ and

---

21 We emphasize that Method 3 only approximately matches the target marginal distributions of $z$ and $x$ because some weights were negative and were reset to zero. The percentages of weights that were negative when education, potential experience, marital status, occupation and industry played the role of $z$ were 2.6, 0.0, 0.0, 3.2 and 1.0, respectively.
In contrast, the table reveals large differences in the effects of occupation and education across the wage distribution. Recall that these two variables have roughly the same effect on the mean (0.055 vs. 0.052, both with standard errors of 0.01) but, as Table IV shows, occupation has only a 0.019 (= 2.813/2.794) effect on the median—much smaller than the education effect of 0.058. Occupation also has essentially no effect through the 25th percentile but a very large effect at the top end of the distribution, i.e. of the 0.257 difference between the 95th percentiles of the black and white log wage distributions, 0.103 (= 3.907/3.804) can be accounted for by black–white differences in occupation. In fact, occupation appears to be solely responsible for the pattern noted above in Figure 1: the five covariates jointly explain more of the black–white difference in the upper tail of the wage distribution than in the lower tail.

6. DISCUSSION AND CONCLUSION

We analyze three reweighting methods for isolating the roles of individual covariates in producing between-group differences in outcome distributions. We show that the method we propose more closely approximates the typical empirical thought experiment of considering the role of one covariate while holding ‘all else equal’, such as when least-squares methods are used to examine mean outcome differences. In contrast, we show that the two methods used in previous studies answer fundamentally different questions.

We demonstrate our approach using an empirical analysis of log wage differences between black and white adult males. Our proposed approach yields results that are similar to those produced by OLS, while the other methods yield substantially different results. Using the same empirical example, we also show that it is straightforward to analyze features other than means of the counterfactual log wage distributions that would result if whites had blacks’ marginal distribution of one covariate but their own distribution of all other covariates.

We note several caveats about our approach. First, in some circumstances, a counterfactual distribution with the properties we are seeking to capture may not exist. We suggest two possible responses to that situation, but neither is entirely satisfying. Second, we recommend our approach only when all covariates can be treated as discrete. Rothe (2012) suggests a related, non-reweighting approach that can be used with continuous covariates.

Note: All estimates in the ‘Reweighted whites’ column are based on Method 3, as described in the text. \( N = 213,908 \) for whites and reweighted whites; \( N = 23,945 \) for blacks.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>Unweighted Whites</th>
<th>Unweighted Blacks</th>
<th>Reweighted whites ( z = \text{education} )</th>
<th>Reweighted whites ( z = \text{marital status} )</th>
<th>Reweighted whites ( z = \text{occupation} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.056</td>
<td>0.749</td>
<td>1.022</td>
<td>1.019</td>
<td>1.040</td>
</tr>
<tr>
<td>5</td>
<td>1.753</td>
<td>1.489</td>
<td>1.718</td>
<td>1.727</td>
<td>1.753</td>
</tr>
<tr>
<td>10</td>
<td>2.015</td>
<td>1.753</td>
<td>1.976</td>
<td>1.987</td>
<td>2.030</td>
</tr>
<tr>
<td>25</td>
<td>2.412</td>
<td>2.163</td>
<td>2.367</td>
<td>2.392</td>
<td>2.417</td>
</tr>
<tr>
<td>50</td>
<td>2.813</td>
<td>2.571</td>
<td>2.755</td>
<td>2.788</td>
<td>2.794</td>
</tr>
<tr>
<td>75</td>
<td>3.200</td>
<td>2.969</td>
<td>3.146</td>
<td>3.180</td>
<td>3.178</td>
</tr>
<tr>
<td>90</td>
<td>3.607</td>
<td>3.373</td>
<td>3.530</td>
<td>3.585</td>
<td>3.532</td>
</tr>
<tr>
<td>95</td>
<td>3.907</td>
<td>3.650</td>
<td>3.825</td>
<td>3.873</td>
<td>3.804</td>
</tr>
<tr>
<td>99</td>
<td>4.915</td>
<td>4.709</td>
<td>4.861</td>
<td>4.914</td>
<td>4.828</td>
</tr>
</tbody>
</table>

Note: All estimates in the ‘Reweighted whites’ column are based on Method 3, as described in the text. \( N = 213,908 \) for whites and reweighted whites; \( N = 23,945 \) for blacks.

22 We obtain standard errors from 200 bootstrap replications in Table III. Note that we have not established formal inference properties for our estimator.
Third, although the counterfactuals generated by our approach are most similar to those produced by OLS, they may not always be the objects of interest. For example, suppose that one is interested in the effects of maternal age and education, both measured when an infant is born, on the black–white gap in infant mortality. If one wishes to estimate the effect on white infant mortality rates of a shift from the white distribution of maternal age to the black distribution, Method 3 does so in a way that holds the distribution of maternal education constant, as would an OLS regression that included both covariates. However, because very young mothers will necessarily have relatively low educational attainment, it is difficult to imagine a policy change that would shift the share of mothers who are teens but would not also shift the distribution of maternal education. Thus, considering the combined direct and indirect (through education) effects of maternal age may also be of interest for analyses of the potential effects of policies that target the age distribution of mothers.23

Despite these caveats, the methods developed here provide an attractive approach to further understand the contribution of each of the individual covariates to between-group differences in outcome distributions. Our methods are based on a simple and transparent idea, creating counterfactual distributions that shift the marginal distribution of one covariate at a time while holding the distributions of other covariates constant. Our analytical and empirical results demonstrate that our proposed methods capture the spirit of an approach based on linear models but allow for the flexibility of reweighting.

ACKNOWLEDGEMENTS

We thank Fabian Lange, Peter Schmidt, Gary Solon and numerous seminar participants for detailed comments on earlier drafts. A version of this paper was presented at the 8th World Congress of the International Health Economics Association, July 2011. Elder is an Associate Professor of Economics at Michigan State University, and Goddeeris and Haider are Professors of Economics at Michigan State University. Research reported in this paper was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R03HD073565.

REFERENCES


23 Altonji *et al.* (2012) reason along these lines in adopting a sequential approach to measuring the role of individual covariates on outcomes. To illustrate their approach, suppose that maternal education and child education are the two covariates of interest. Because maternal education is determined prior to and likely influences child education, it is arguably of interest to consider maternal education first in a sequential decomposition, recognizing that doing so will attribute to maternal education effects on outcomes that arise through correlated changes in child education.


**APPENDIX A: THE COUNTERFACTUAL MARGINAL DISTRIBUTIONS OF Z AND X**

**Proposition.** Assume the support condition (7) is satisfied. Reweighting Group A using weights given by $w_A(z, x) = \frac{dF(z | j = B) - dF(z | j = A) + dF(z | x, j = A)}{dF(z | x, j = A)}$ produces a counterfactual distribution that matches group B’s marginal distribution of z and Group A’s marginal distribution of x.

**Proof.** First note the identity:

$$dF(z | j = A) = \int_x dF(z | x, j = A)dF(x | j = A)$$  \hspace{1cm} (A.1)

In the counterfactual population, $dF(z)$ is the reweighted version of the right-hand side of equation (A.1):

$$\int_x dF(z | x, j = A)w_A(z, x)dF(x | j = A)$$

$$= \int_x dF(z | x, j = A)\left[\frac{dF(z | j = B) - dF(z | j = A) + dF(z | x, j = A)}{dF(z | x, j = A)}\right]dF(x | j = A)$$  \hspace{1cm} (A.2)

$$= dF(z | j = B) - dF(z | j = A) + \int_x dF(z | x, j = A)dF(x | j = A)$$

$$= dF(z | j = B)$$
Analogously, $dF(x)$ in the counterfactual population, which equals $\int dF(x|z,j = A)\psi^3(z,x)dF(z|j = A)$, matches group $A$’s distribution of $x$:

$$\int dF(x|z,j = A)\psi^3(z,x)dF(z|j = A)$$

$$= \int dF(x|z,j = A)\left[\frac{dF(z|j = B) - dF(z|j = A) + dF(z|x,j = A)}{dF(z|x,j = A)}\right]dF(z|j = A)$$

$$= \int dF(z|x,j = A)dF(x|j = A)\left[\frac{dF(z|j = B) - dF(z|j = A) + dF(z|x,j = A)}{dF(z|x,j = A)}\right] (A.3)$$

$$= \int dF(x|j = A)dF(z|j = B) - dF(z|j = A) + dF(z|x,j = A)$$

$$= dF(x|j = A)$$

APPENDIX B: THE EFFECT OF REWEIGHTING ON $dF(z|X = x) - dF(z|X = x')$

**Proposition.** Reweighting group $A$ using weights defined by $\psi^3(z,x)$ produces a counterfactual that matches group $A$’s value of $[dF(z|X = x) - dF(z|X = x')]$ for all $z$, $x$ and $x'$, assuming the weights always exist and are non-negative.

**Proof.** The weighting function $\psi^3(z,x) = \frac{dF(z|j = B) - dF(z|j = A) + dF(z|x,j = A)}{dF(z|x,j = A)}$ produces the following counterfactual conditional density:

$$dF^c(z|X = x) = dF(z|x,j = A)\frac{dF(z|j = B) - dF(z|j = A) + dF(z|x,j = A)}{dF(z|x,j = A)}$$

$$= dF(z|j = B) - dF(z|j = A) + dF(z|x,j = A)$$

$$= dF(z) - dF(z|j = A) + dF(z|x,j = A)$$

where the last line of follows because $dF^c(z)=dF(z|j=B)$. Using two different values of $x$, denoted $x$ and $x'$, in equation (A.4) yields

$$dF^c(z|X = x) - dF^c(z|X = x') = [dF^c(z) - dF(z|j = A) + dF(z|x,j = A)]$$

$$\quad - [dF^c(z) - dF(z|j = A) + dF(z|x',j = A)]$$

$$= dF(z|X = x,j = A) - dF(z|X = x',j = A)$$

$\square$

As a first implication of this proposition, note that when $x$ and $z$ are scalars the $\psi^3(z,x)$ weights preserve the covariance of $x$ and $z$. To see this, note that

$$\text{cov}(z,x|j = A) = E(zx|j = A) - E(z|j = A)E(x|j = A)$$

$$= \int z\int x dF(z|x,j = A)dF(x|j = A)\int dF(z|j = A)\int x dF(x|j = A)$$

$$= \int z\int x [dF(z|x,j = A) - dF(z|j = A)]dF(x|j = A)$$

Because the counterfactual preserves group $A$’s value of $[dF(z|X = x) - dF(z|X = x')]$ for all $z$, $x$, and $x'$, it also preserves group $A$’s value of $[dF(z|X) - dF(z)]$ for all $z$ and $x$. Moreover, the
counterfactual preserves group A’s value of \( dF(x) \), so the integral in the last line of equation (A.6) does not change upon reweighting, implying that \( \text{cov}(z, x) \) in the counterfactual equals \( \text{cov}(z, x \mid j = A) \).

As a second implication of this proposition, note that equation (A.4) can be rewritten in terms of the joint distribution of \( z \) and \( x \):

\[
dF^c(z, x) = dF(z \mid x, j = A) dF(x \mid j = A) \frac{dF(z \mid j = B) - dF(z \mid j = A)}{\text{dF}(z \mid x, j = A)}
\]

\[
= dF(x \mid j = A) \times [dF(z \mid j = B) - dF(z \mid j = A) + dF(z \mid x, j = A)]
\]

\[
= dF(z, x \mid j = A) + dF(x \mid j = A) \times [dF(z \mid j = B) - dF(z \mid j = A)]
\]

This expression is useful because, when \( z \) is discrete, Rothe’s (2012) counterfactual joint density \( dF^c(z, x \mid j = A) \) is bounded between \( dF(z, x \mid j = A) \) and \( \max[0, dF(z, x \mid j = A) + dF(z \mid j = B) - dF(z \mid j = A)] \). Clearly, \( dF(z, x \mid j = A) \) lies within the interval given by Rothe’s bounds as long as \( dF(z, x \mid j = A) \) lies between \( dF(z \mid j = B) - dF(z \mid j = A) \). Therefore, our counterfactual density of \( \{z, x\} \) will always lie within Rothe’s (2012) discrete-regressor bounds if the density produced using our preferred weights lies in the \([0, 1]\) interval. Additionally, because both our methods and Rothe’s (2012) hold the conditional density of \( y \) constant at \( dF(y \mid z, x, j = A) \), our estimate of any function of the counterfactual density of \( y \) must be within Rothe’s bounds.

APPENDIX C: DERIVING THE EXPRESSIONS IN SECTION 3.4

To obtain equation (12), first note that the counterfactual density \( f(y; \psi^{23} \mid j = A) \) can be written as

\[
f(y; \psi^{23} \mid j = A) = \sum_z \sum_x \left[ f(y \mid z, x, j = A) \psi^{23}(z, x) P(z, x \mid j = A) \right]
\]

\[
= \sum_z \sum_x \left[ f(y \mid z, x, j = A) \left( \frac{P(z \mid j = B) - P(z \mid j = A)}{P(z \mid x, j = A)} + 1 \right) P(z, x \mid j = A) \right]
\]

\[
= f(y \mid j = A) + \sum_z \sum_x \left[ f(y \mid z, x, j = A) \left( \frac{(P(z \mid j = B) - P(z \mid j = A))}{P(z \mid x, j = A)} \right) P(z, x \mid j = A) \right]
\]

Therefore

\[
f(y \mid j = A) - f(y; \psi^{23} \mid j = A) = \sum_z \sum_x \left[ f(y \mid z, x, j = A) \left( \frac{(P(z \mid j = A) - P(z \mid j = B))}{P(z \mid x, j = A)} \right) P(z, x \mid j = A) \right]
\]

\[
= \sum_x \left[ P(x \mid j = A) \sum_z f(y \mid z, x, j = A) (P(z \mid j = A) - P(z \mid j = B)) \right]
\]

(A.7)

where the second equality follows because \( P(z, x \mid j = A)/P(x \mid j = A) = P(x \mid j = A) \).

To obtain equation (13), note that equation (A.7) extends immediately to

\[
E(y \mid j = A) - E(y; \psi^{23} \mid j = A) = \sum_x \left[ P(x \mid j = A) \left( \sum_z E(y \mid z, x, j = A) (P(z \mid j = A) - P(z \mid j = B)) \right) \right]
\]

(A.8)

Copyright © 2015 John Wiley & Sons, Ltd.

*J. Appl. Econ.* (2015)

DOI: 10.1002/jae
Assuming that \( z \) is binary then delivers the desired result:

\[
E(y \mid j = A) - E(y; \psi^{3*} \mid j = A) = \sum_x \left\{ P(x \mid j = A) \left[ E(y \mid z = 1, x, j = A)(P(z = 1 \mid j = A) - P(z = 1 \mid j = B)) \right. \\
\left. + E(y \mid z = 0, x, j = A)(1 - P(z = 1 \mid j = A) - (1 - P(z = 1 \mid j = B))) \right] \right\}
\]

\[
= \{ E_x_{j=A} [E(y \mid z = 1, x, j = A) - E(y \mid z = 0, x, j = A)] \} \{ E(z \mid j = A) - E(z \mid j = B) \}
\]

\[(A.9)\]

**APPENDIX D: ADJUSTMENTS FOR VIOLATIONS OF THE SUPPORT CONDITION**

To describe our approach for addressing violations of the support condition given by equation (7), we first define some notation for various counts of observations. Let \( p(x) \) be an index of the possible values of \( x \) \((p = 1, \ldots, P)\), and let \( q(z) \) be an index of the possible values of \( z \) \((q = 1, \ldots, Q)\) (note that we now explicitly treat \( x \) and \( z \) as discrete and finite valued). For convenience, we will suppress the arguments of \( p \) and \( q \) and will express the weights as functions of \( q \) and \( p \) rather than of \( z \) and \( x \). Then, define

\[ N^j_{pq} \quad \text{as the number of observations with characteristics \((p, q)\) in group } j \]
\[ N^j_q \quad \text{as the number of observations with characteristics } q \text{ in group } j \left( \equiv \sum_p N^j_{pq} \right) \]
\[ N^j_p \quad \text{as the number of observations with characteristics } p \text{ in group } j \left( \equiv \sum_q N^j_{pq} \right) \]
\[ N^j \quad \text{as the number of observations in group } j \left( \equiv \sum_p \sum_q N^j_{pq} \right) \]

We first discuss adjustments to the \( \psi^{3*} \) weights and later how the results can be extended to adjust the \( \psi^3 \) weights. When the sample analog of condition (7) is satisfied, the weights given by equation (17) are constructed empirically as

\[
\psi^{3*}(q, p) = \frac{N^B_{pq}/N^B_p}{N^A_{pq}/N^A_p}.
\]

\[(A.10)\]

When condition (7) is not satisfied, we partition the values of \( x \) into two sets, denoting the set of values of \( p \) for which there is at least one observation in group \( A \) for each value of \( z \) (i.e. values of \( p \) such that \( \min_q \left( N^A_{pq} \right) > 0 \)) as \( 'p^* \) and the remaining values of \( p \) as \( 'p^{**} \) (i.e. those values of \( p \) such that \( \min_q \left( N^A_{pq} \right) = 0 \)). To create a counterfactual distribution with the appropriate marginal distributions, \( p^{**} \) must not be empty. Then, the adjusted weights are

\[
\psi^{3*}(q, p) = \frac{N^B_{pq}/N^B_p}{N^A_{pq}/N^A_p} \times f_1(p) \times f_2(q) - \left\{ f_3(p) \times f_4(q) \times N^A_{pq}/N^A_p \right\} \text{ if } p \in p^{**}
\]

\[
\frac{\sum_q \left[ f_4(p) \times 1 \left( N^A_{pq} > 0 \right) \right]}{\sum_q f_5(q)} \text{ if } p \in p^{*}
\]

\[(A.11)\]
where

\[
\begin{align*}
  f_1(p) &= \frac{N^B}{\sum_q [N^B_{pq} \times 1(N^A_{pq} > 0)]}, \\
  f_2(q) &= \frac{N^A}{\sum_p [N^A_{pq} \times 1(N^A_{pq} > 0)]}, \\
  f_3(p) &= \frac{N^A_p \times f_1(p)}{N^A} \times \sum_q \left[ \frac{N^B_q}{N^B} \times [f_2(q) - 1] \times 1(N^A_{pq} > 0) \right], \\
  f_4(q) &= \frac{N^B_q \times f_2(q)}{N^B} \times \sum_p \left[ \frac{N^A_p}{N^A} \times [f_1(p) - 1] \times 1(N^A_{pq} > 0) \right], \\
  f_5(q) &= f_4(q) \left[ 1 - \sum_p \left( \frac{f_1(p)}{\sum_q f_4(q) \times 1(N^A_{pq} > 0)} \right) \right].
\end{align*}
\]

(A.12)

When condition (7) is not satisfied, using the equation (A.10) weights multiplied by \(f_1(p)\) would produce a distribution with group A’s distribution of \(x\); similarly, using the equation (A.10) weights multiplied by \(f_2(q)\) would produce a distribution with group B’s distribution of \(z\). Multiplying by \(f_1(p) \times f_2(q)\), as we do in equation (A.11), produces weights that are on average too large, necessitating further adjustments involving \(f_3(p), f_4(q)\) and \(f_5(q)\). For all values of \(x\) in \(p^+\), the adjustments to the weights shown in equation (A.12) assure that group A’s marginal probabilities are matched. The adjustments for values of \(x\) in \(p^\cdot\) shown in equation (A.12) take account of what has been done for \(x\) in \(p^+\), and assure not only that group A’s marginal probabilities are matched for \(x\) in \(p^\cdot\), but also that group B’s marginal probabilities are matched for all values of \(z\).

To construct the \(\psi^5\) weights adjusted for violation of equation (7), start with the weights defined by equation (A.11), subtract weights defined similarly to those defined in equation (A.11) except that all ‘B’ superscripts in equations (A.11) and (A.12) have been changed to ‘A’ and add 1. In addition to the reason discussed at the end of Section 3.1, these adjustments for violation of condition (7) create another possibility for generating negative weights, but in our experience this scenario has not been empirically relevant. In the empirical example we describe in Section 5, the adjustments for violation of the support condition produce no negative weights. They also produce no negative weights if we reduce the sample size to 1% of the underlying data, so that we are left with 21,497 observations and 7560 possible unique values of the five covariates.

An Illustrative Example

To see how the adjustments described above work in practice, suppose that \(z\) and \(x\) can each take only two values and that the relative frequencies in group B are as follows:

<table>
<thead>
<tr>
<th>(z = 0)</th>
<th>(z = 1)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x = 0)</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>(x = 1)</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>Total</td>
<td>1/2</td>
<td>1/2</td>
</tr>
</tbody>
</table>
First consider a case in which condition (7) is satisfied, with the following relative frequencies in group A:

<table>
<thead>
<tr>
<th>Case 1</th>
<th>$z = 0$</th>
<th>$z = 1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/3</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>1/2</td>
<td>1/6</td>
<td>2/3</td>
</tr>
<tr>
<td>Total</td>
<td>2/3</td>
<td>1/3</td>
<td>1</td>
</tr>
</tbody>
</table>

These tables highlight in bold the marginal distribution of $z$ in group $B$ and the marginal distribution of $x$ in group $A$, which we want to match in the counterfactual distribution. Using the notation used in the main text, if we use $\psi_{3}^z (z, x) = \frac{dF(z \mid j = B)}{dF(z \mid j = A)}$ to reweight group $A$, the weight applied to the $(z = 0, x = 0)$ cell would be $\frac{Pr(z = 0 \mid j = B)}{Pr(z = 0 \mid j = A)} = \frac{1/2}{1/3} = 1$, so that the counterfactual relative frequency would be $1 \times 1/6 = 1/6$. Filling in the other entries (by multiplying the group $A$ relative frequency by the relevant weight), we find that all of the counterfactual marginal probabilities are correct:

<table>
<thead>
<tr>
<th>Case 1</th>
<th>$z = 0$</th>
<th>$z = 1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0$</td>
<td>$1/6 \times 1 = 1/6$</td>
<td>$1/6 \times 1 = 1/6$</td>
<td>1/3</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>$1/2 \times 2/3 = 1/3$</td>
<td>$1/6 \times 2 = 1/3$</td>
<td>2/3</td>
</tr>
<tr>
<td>Total</td>
<td>1/2</td>
<td>1/2</td>
<td>1</td>
</tr>
</tbody>
</table>

Note that $z$ and $x$ are independent in the counterfactual distribution: $Pr(x = 0 \mid z)$ does not vary by $z$, and $Pr(z = 0 \mid x)$ does not vary by $x$. This will always be true if condition (7) is satisfied and the $\psi_{3}^z (z, x)$ weights are used.

Now consider a situation in which condition (7) is not satisfied, which means that there is at least one ‘empty cell’. Suppose the group $B$ distribution of $\{z, x\}$ is given by

<table>
<thead>
<tr>
<th>Case 2</th>
<th>$z = 0$</th>
<th>$z = 1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 0$</td>
<td>1/3</td>
<td>1/3</td>
<td>2/3</td>
</tr>
<tr>
<td>$x = 1$</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>Total</td>
<td>2/3</td>
<td>1/3</td>
<td>1</td>
</tr>
</tbody>
</table>

and the group $A$ distribution is given by

24 Although we develop this example using the $\psi_{3}^z (z, x)$ weights, in any case in which $z$ and $x$ are binary and there is one empty cell in group $A$, the adjusted $\psi_{3}^z (z, x)$ and $\psi_{3}^x (z, x)$ weights are identical.
Thus each group has an empty cell at \((z = 1, x = 1)\). If we reweight the non-empty cells of group \(A\) using \(\psi_{z}^{3*}(z, x)\), we obtain the following counterfactual relative frequencies:

The \(z = 0\) column and \(x = 0\) row, with no empty cells, have the correct totals in that the frequency of the \(z = 0\) column matches that in Table A.IV and the frequency of the \(x = 0\) row matches that in Table A.V. However, the \(z = 1\) column and \(x = 1\) row totals are too small. Note that the sum of the joint probabilities is 4/5; the ‘missing probability’, 1/5, is what weighting by \(\psi_{z}^{3*}(z, x)\) would have placed in cell \((z = 1, x = 1)\) if there were any such observations in group \(A\). Comparing Tables A.V and A.VI, it is easy to see that one could ‘fix’ the \(x = 1\) row by multiplying the weight used for \((z = 0, x = 1)\) by 3/2, the ratio of the desired row total to the actual total. However, the \(z = 0\) column total would no longer be correct. Similarly, it is easy to fix the second column by multiplying the weight used for \((z = 1, x = 0)\) by 5/2, but the row 1 total would no longer be correct.

To get a sense of how our algorithm works by applying it to this example, we begin by performing both of the above adjustments (i.e. multiplying by both \(f_{1}(p)\) and \(f_{2}(q)\)), which yields the following matrix:

After this step, in general some row and column totals may be correct, but the remaining totals will be too large. In this case, only the \(z = 1\) column and \(x = 1\) row totals are correct: the \(z = 1\) column total matches that in Table A.IV, while the \(x = 1\) row total matches that in Table A.V. Our next step is to reduce weights if necessary in any rows with empty cells to correct the totals in those rows. In this case

As an informal way to see why reweighting by \(\psi_{z}^{3*}(z, x)\) fails in this case, note that this scheme can only be successful when implementing it will make the distributions of \(z\) and \(x\) independent. When some combinations of \(z\) and \(x\) do not appear in group \(A\), the reweighted distributions of \(z\) and \(x\) cannot be independent, regardless of the weights used.
there is only one such row \((x = 1)\), and no further adjustment to it is needed, but that will not typically be true in more complex cases.

Once the rows with empty cells have correct totals, our last step is to remove the excess weight from the row (or rows) with no empty cells, in such a way that the row total(s) and all column totals are correct. Because the other rows now have correct totals and the grand totals are the same across rows and columns, this is always possible. It is easy to see what must be done in this example: only cell \((z = 0, x = 0)\) must be changed, because the second column is already correct; this adjustment corresponds to the term involving \(f_3(p)f_4(q)\), and \(f_5(q)\) in equation \((A.12)\). The final result is as follows:

<table>
<thead>
<tr>
<th></th>
<th>(z = 0)</th>
<th>(z = 1)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x = 0)</td>
<td>(1/5 \times 1/3 = 1/15)</td>
<td>(1/5 \times 5/3 = 1/3)</td>
<td>(2/5)</td>
</tr>
<tr>
<td>(x = 1)</td>
<td>(3/5 \times 1 = 3/5)</td>
<td>0</td>
<td>(3/5)</td>
</tr>
<tr>
<td>Total</td>
<td>(2/3)</td>
<td>(1/3)</td>
<td>(1)</td>
</tr>
</tbody>
</table>

In this reweighted population, the marginal distribution of \(z\) matches that found in Table A.IV, while the marginal distribution of \(x\) matches that found in Table A.V, exactly as intended.