AVERAGE GAPS AND OAXACA–BLINDER DECOMPOSITIONS: A CAUTIONARY TALE ABOUT REGRESSION ESTIMATES OF RACIAL DIFFERENCES IN LABOR MARKET OUTCOMES

TYMON SŁOCZYŃSKI*

WINNER OF THE 2019 BEST PAPER COMPETITION
LERA / *ILR REVIEW* SPECIAL SERIES IN EMPLOYMENT RELATIONS

Using a recent result from the program evaluation literature, the author demonstrates that the interpretation of regression estimates of between-group differences in wages and other economic outcomes depends on the relative sizes of subpopulations under study. When the disadvantaged group is small, regression estimates are similar to the average loss for disadvantaged individuals. When this group is a numerical majority, regression estimates are similar to the average gain for advantaged individuals. The author analyzes racial test score gaps using ECLS-K data and racial wage gaps using CPS, NLSY79, and NSW data, and shows that the interpretation of regression estimates varies substantially across data sets. Methodologically, he develops a new version of the Oaxaca–Blinder decomposition, in which the unexplained component recovers a parameter referred to as *the average outcome gap*. Under additional assumptions, this estimand is equivalent to the average treatment effect. Finally, the author reinterprets the Reimers, Cotton, and Fortin decompositions in the context of the program evaluation literature, with attention to the limitations of these approaches.

*TYMON SŁOCZYŃSKI is an Assistant Professor at the Department of Economics and International Business School, Brandeis University.

I am grateful to Arun Advani, Joshua Angrist, Anna Baranowska-Rataj, Elizabeth Brainerd, Brantly Callaway, Thomas Crossley, Todd Elder, Steven Haider, Krzysztof Karbownik, Patrick Kline, Michał Myck, Mateusz Myśliwski, Ronald Oaxaca, Pedro Sant'Anna, Jörg Schwiebert, Gary Solon, Adam Szulc, Joanna Tyrowicz, Glen Waddell, Rudolf Winter-Ebmer, Jeffrey Wooldridge, seminar participants at Bank of Canada, Brandeis, CEPS/INSTEAD, and SGH, and participants at many conferences for helpful comments and discussions. I acknowledge financial support from the National Science Centre (grant DEC-2012/05/N/HS4/00395), the Foundation for Polish Science (a "Start" scholarship), the "Weź stypendium – dla rozwoju" scholarship program, and the Theodore and Jane Norman Fund. This article builds on ideas from and supersedes my papers, "Population Average Gender Effects" and "Average Wage Gaps and Oaxaca–Blinder Decompositions." Data and copies of the computer programs used to generate the results presented in the article are available from the author at tslocz@brandeis.edu, with the exception of CPS data, which are available from IPUMS.

KEYWORDS: black–white gaps, decomposition methods, test scores, treatment effects, wages

ILR Review, 73(3), May 2020, pp. 705–729
DOI: 10.1177/0019793919874063. © The Author(s) 2019
Journal website: journals.sagepub.com/home/ilr

Despite five decades of progress since the civil rights movement, black–white gaps in economic outcomes persist in the United States. Much research has focused on racial differences in wages (Neal and Johnson 1996; Lang and Manove 2011), labor force participation (Boustan and Collins 2014), unemployment (Ritter and Taylor 2011), home ownership (Collins and Margo 2001; Charles and Hurst 2002), wealth (Blau and Graham 1990; Barsky, Bound, Charles, and Lupton 2002), cognitive skills (Fryer and Levitt 2004, 2006, 2013; Bond and Lang 2013), non-cognitive skills (Elder and Zhou 2017), infant mortality (Elder, Goddeeris, and Haider 2016), and other outcomes. Recent surveys of this topic—and the related problem of racial discrimination—include Charles and Guryan (2011), Fryer (2011), and Lang and Lehmann (2012). Even after controlling for many observable characteristics of individuals, a typical study finds a significant black–white gap that remains unexplained.

Traditionally, unexplained gaps in mean outcomes have been examined using decomposition methods (see, e.g., Elder, Goddeeris, and Haider 2010; Fortin, Lemieux, and Firpo 2011; Firpo 2017). As noted by Charles and Guryan (2011), however, in recent empirical work researchers have typically focused on a simpler approach of estimating the following model using ordinary least squares (OLS):

$$Y_i = \alpha B_i + X_i\beta + \varepsilon_i, \tag{1}$$

where Y_i is the outcome under study, B_i is a binary variable that indicates race (1 if black, 0 if white), and X_i is a row vector of observed characteristics. Indeed, this simple method has been used in many important articles on black–white gaps, including Collins and Margo (2001), Charles and Hurst (2002), Fryer and Levitt (2004, 2006), Lang and Manove (2011), Bond and Lang (2013), Fryer and Levitt (2013), Fryer, Pager, and Spenkuch (2013), Rothstein and Wozny (2013), and Elder and Zhou (2017).

In this article, I borrow from the recent program evaluation literature to illustrate some important limitations of this approach. As discussed by, among others, Angrist (1998), Humphreys (2009), and Słoczyński (2018), OLS estimation of a model analogous to Equation (1) does not recover, in general, the average treatment effect (ATE), unless the effects of the treatment are homogeneous. These results extend to studies of between-group differences in economic outcomes. In particular, this article demonstrates that the interpretation of regression estimates of such differences depends on the relative sizes of the subpopulations under study (e.g., blacks and whites), which is a straightforward extension of a recent result in Słoczyński (2018). While the previous literature on the interpretation of the OLS estimand has focused on treatment effects, in this article I explicitly consider a framework in which the main variable of interest is an "attribute," in the sense of Holland (1986), and thus cannot possibly constitute a "treatment" in any actual experiment. Note, however, that this distinction between

causal inference and decomposition analysis has implications for how we label our parameters of interest but not for the algebra of least squares, which forms the basis of the results in Angrist (1998), Humphreys (2009), and Słoczyński (2018).

This article concentrates particularly on the following implication of the result in Słoczyński (2018). If we refer to one of the groups as "disadvantaged" (e.g., blacks) and to the other as "advantaged" (e.g., whites), then regression estimates will be similar to the average loss for disadvantaged individuals under the condition that they also constitute a numerical minority. When instead these individuals are a numerical majority—albeit disadvantaged—regression estimates will be similar to the average gain for advantaged individuals.

This relationship between the interpretation of regression estimates and the relative sizes of subpopulations under study is illustrated empirically in several applications to racial gaps in test scores (using data from the Early Childhood Longitudinal Study-Kindergarten [ECLS-K]) and in wages (using data from the Current Population Survey [CPS], the National Longitudinal Survey of Youth [NLSY79], and the National Supported Work [NSW] Demonstration). Compared with the ECLS-K, CPS, and NLSY79 data, in which the proportion of blacks is relatively low, the interpretation of regression estimates is very different in the NSW data, in which blacks constitute a numerical majority. Methodologically, I also develop a new version of the Oaxaca–Blinder decomposition (Oaxaca 1973; Blinder 1973) in which the "unexplained component" could be interpreted as the average treatment effect if we decided to invoke the potential outcome model (see, e.g., Holland 1986; Imbens and Wooldridge 2009). Since it is preferable to treat demographic characteristics as attributes, I usually refer to this object as *the average outcome gap*—an equivalent parameter that lacks a causal interpretation. Finally, I also provide treatment-effects reinterpretations of the Reimers (1983), Cotton (1988), and Fortin (2008) decompositions. Each of these procedures is easily shown to recover some generally uninteresting convex combination of conditional average outcome gaps.

Theory

Consider a population divided into two mutually exclusive categories, indexed by $W_i \in \{0, 1\}$ and referred to as the advantaged group ($W_i = 1$) and the disadvantaged group ($W_i = 0$). For each individual i, we also observe an outcome, Y_i, and a row vector of observed characteristics, X_i. In this case, $\mu_1(x) = \mathrm{E}(Y_i|X_i = x, W_i = 1)$ is the expected outcome of an advantaged individual with $X_i = x$ and $\mu_0(x) = \mathrm{E}(Y_i|X_i = x, W_i = 0)$ is the expected outcome of a disadvantaged individual with these characteristics. Moreover, define *the conditional average outcome gap* as $\delta(x) = \mu_1(x) - \mu_0(x)$, that is, the gap between the expected outcomes of an advantaged and a disadvantaged individual with $X_i = x$. This object is also referred to by Li,

Morgan, and Zaslavsky (2018) as the conditional average controlled difference. Dependent on the question we wish to answer, we may average $\delta(x)$ over the whole population, over the subpopulation of advantaged individuals, or over the subpopulation of disadvantaged individuals. Define *the average outcome gap* as

$$\delta_{gap} = \mathrm{E}[\delta(X_i)], \tag{2}$$

namely, the expected value of the conditional average outcome gap over X_i.[1] Within the framework of a potential outcome model, and under additional assumptions, this parameter is equivalent to the average treatment effect, τ_{ATE}. Moreover, define *the average gain for advantaged individuals* and *the average loss for disadvantaged individuals* as

$$\delta_{gain} = \mathrm{E}[\delta(X_i)|W_i = 1] \qquad \text{and} \qquad \delta_{loss} = \mathrm{E}[\delta(X_i)|W_i = 0] \tag{3}$$

respectively. Similarly, under certain conditions, these parameters can be regarded as equivalents of the average treatment effect on the treated, τ_{ATT}, and the average treatment effect on the controls, τ_{ATC}. It is also the case that

$$\delta_{gap} = \mathrm{P}(W_i = 1) \cdot \delta_{gain} + \mathrm{P}(W_i = 0) \cdot \delta_{loss}. \tag{4}$$

Thus, a particular weighted average of the average gain for advantaged individuals and the average loss for disadvantaged individuals is equal to the average outcome gap.

It should be noted that without further assumptions $\delta(x)$, δ_{gap}, δ_{gain}, and δ_{loss} cannot be interpreted as causal or counterfactual; they are also identified from the data. As demonstrated by Fortin et al. (2011), a counterfactual interpretation can be justified by a set of three additional assumptions: simple counterfactual treatment, overlapping support, and conditional independence/ignorability. These assumptions are elaborated as follows:

> **Assumption 1 (Simple Counterfactual Treatment):** The observed conditional mean of advantaged (disadvantaged) individuals represents a counterfactual conditional mean for disadvantaged (advantaged) individuals.

This assumption restricts the analysis to counterfactuals that are based on the observed conditional mean for the other group. In other words, the observed conditional mean of advantaged individuals provides a counterfactual for disadvantaged individuals, and vice versa. It should be noted that this assumption rules out the presence of general equilibrium effects, and this might be a substantial restriction in some empirical contexts.

[1]This notation intentionally mimics Imbens and Wooldridge (2009: 26–27) so that the analogy between conditional average treatment effects and conditional average outcome gaps becomes clear.

Assumption 2 (Overlapping Support): Let the support of observed characteristics X_i be χ. For all x *in* χ, $0 < \mathrm{P}(W_i = 1 | X_i = x) < 1$.

The overlapping support assumption ensures that no combination of observed characteristics can be used to identify group membership. This restriction might be somewhat controversial in the context of black–white differences in economic outcomes, as it is likely that many black or white individuals might have few counterparts in the other subpopulation; clearly, similar problems can also arise in other empirical contexts.

Assumption 3 (Conditional Independence/Ignorability): Denote the unobserved characteristics as ε_i. Let $(W_i, X_i, \varepsilon_i)$ have a joint distribution. Then, $W_i \perp \varepsilon_i | X_i$, i.e., the individual's unobserved characteristics are independent of group membership, conditional on observed covariates.

This assumption rules out the presence of unobserved characteristics, which would be correlated with both group membership and outcomes, conditional on observed covariates. For example, this requirement would be violated in the case of black–white differences in wages if school quality were correlated with both wages and race (conditional on X_i), while also being unobserved.[2] Indeed, on the one hand, Card and Krueger (1992) argued that omitting measures of school quality might affect estimates of black–white wage gaps; on the other hand, Grogger (1996) presented a different view.

Note that Assumptions 1, 2, and 3 guarantee identification of the aggregate decomposition (Fortin et al. 2011). If we maintain these assumptions, we can construct a counterfactual distribution, which would be observed if the outcomes of disadvantaged individuals were determined according to the conditional mean of advantaged individuals, and vice versa. This counterfactual experiment provides a meaningful interpretation of δ_{gap}, δ_{gain}, and δ_{loss}. The average outcome gap, δ_{gap}, is equal to the difference between mean outcomes in two counterfactual distributions. In the first distribution, the outcomes of all individuals are determined according to the conditional mean of advantaged individuals; in the second, the outcomes of all individuals are determined according to the conditional mean of disadvantaged individuals. Similarly, the average gain for advantaged individuals, δ_{gain}, is equal to the average gap between 1) actual outcomes of these individuals, and 2) their counterfactual outcomes, which would be observed if these outcomes were determined according to the conditional mean of disadvantaged individuals. Finally, the average loss for disadvantaged individuals, δ_{loss}, is equal to the average gap between 1) their counterfactual outcomes, which would be observed if these outcomes were determined according to

[2]Of course, some form of endogeneity might also arise if unobserved covariates with different correlation patterns exist. However, as demonstrated by Fortin et al. (2011), identification of the aggregate decomposition is not threatened unless the conditional independence assumption is violated.

the conditional mean of advantaged individuals, and 2) the actual outcomes of disadvantaged individuals.

Arguably, δ_{loss} might be the most intuitive estimand in many empirical contexts. For example, in a study of black–white differences in wages, it seems reasonable to focus on counterfactual wages of black workers, which would be observed if they were paid according to the wage structure of white workers. At the same time, the decomposition literature has often been concerned with both gains and losses (see, e.g., Fortin 2008), and therefore δ_{gap} and δ_{gain} might also be interesting. The average outcome gap—a noncausal equivalent of the average treatment effect—is especially likely to be the primary object of interest in many empirical studies. Comparing mean outcomes of *all individuals* in two counterfactual distributions that differ only in the choice of the conditional mean used to generate them is intuitively appealing.

Regression Estimates

As noted previously, researchers often analyze between-group differences in economic outcomes by means of OLS estimation of the simple linear model:

$$Y_i = X_i\gamma + \delta W_i + \varepsilon_i. \tag{5}$$

Now, unlike in Equation (1), the disadvantaged group is the omitted category.[3] This ensures that the sign of δ is consistent with the signs of δ_{gap}, δ_{gain}, and δ_{loss}.

Słoczyński (2018) studied the interpretation of regression estimates in a model analogous to Equation (5) in which W_i is instead a binary treatment variable (1 if treated, 0 if control). His main result is that

$$\hat{\delta}_{OLS} = (1 - \hat{\pi}) \cdot \tilde{\tau}_{ATT} + \hat{\pi} \cdot \tilde{\tau}_{ATC}, \tag{6}$$

where $\hat{\delta}_{OLS}$ is the OLS estimate of the coefficient on W_i in Equation (5), $\tilde{\tau}_{ATT} = (\hat{\iota}_1 - \hat{\iota}_0) + \left(\hat{\theta}_1 - \hat{\theta}_0\right) \cdot \mathrm{E}_n[\hat{p}(X_i)|W_i = 1]$ and $\tilde{\tau}_{ATC} = (\hat{\iota}_1 - \hat{\iota}_0) + \left(\hat{\theta}_1 - \hat{\theta}_0\right) \cdot \mathrm{E}_n[\hat{p}(X_i)|W_i = 0]$ are particular estimates of the average treatment effect on the treated (ATT) and the average treatment effect on the controls (ATC); $\hat{p}(X_i)$ is the estimated propensity score from the linear probability model; $\hat{\iota}_1$ and $\hat{\iota}_0$ are the estimated intercepts and $\hat{\theta}_1$ and $\hat{\theta}_0$ are the estimated slope coefficients from group-specific (i.e., conditional on W_i) regressions of Y_i on $\hat{p}(X_i)$; and $\hat{\pi} = \frac{\hat{\mathrm{P}}(W_i = 1) \cdot \mathrm{V}_n[\hat{p}(X_i)|W_i = 1]}{\hat{\mathrm{P}}(W_i = 1) \cdot \mathrm{V}_n[\hat{p}(X_i)|W_i = 1] + \hat{\mathrm{P}}(W_i = 0) \cdot \mathrm{V}_n[\hat{p}(X_i)|W_i = 0]}$ is increasing in $\hat{\mathrm{P}}(W_i = 1)$, the sample proportion of treated individuals.[4] Refer to

[3] In this case, of course, all elements of γ other than the intercept are equal to the corresponding elements of β in Equation (1). Also, $\gamma_0 = \alpha + \beta_0$, where β_0 denotes the intercept in Equation (1) and γ_0 denotes the intercept in Equation (5).

[4] Also, for a generic random variable Z, $\mathrm{E}_n[Z_i] = n^{-1} \sum_{i=1}^{n} Z_i$ and $\mathrm{V}_n[Z_i] = n^{-1} \sum_{i=1}^{n} (Z_i - \mathrm{E}_n[Z_i])^2$.

Słoczyński (2018) for additional detail, including the derivation of this result, the intuition behind it, and a number of further extensions and empirical applications.

In this article, I focus on a setting in which W_i is instead an attribute (e.g., race or gender). Since this only influences the labeling of various parameters of interest—but not the algebra of least squares—the result in Słoczyński (2018) also implies that in the current setting:

$$\hat{\delta}_{OLS} = (1 - \hat{\pi}) \cdot \tilde{\delta}_{gain} + \hat{\pi} \cdot \tilde{\delta}_{loss}, \tag{7}$$

where $\hat{\pi}$ is again increasing in $\hat{\mathrm{P}}(W_i = 1)$.[5]

In other words, if there are many disadvantaged individuals (e.g., blacks), the weight on the average loss for these individuals, $\hat{\pi}$, is relatively small. In a benchmark case where $\mathrm{V_n}[\hat{p}(X_i)|W_i = 1] = \mathrm{V_n}[\hat{p}(X_i)|W_i = 0]$, $\hat{\pi}$ is equal to $\hat{\mathrm{P}}(W_i = 1)$. What follows,

$$\hat{\delta}_{OLS} \simeq \hat{\mathrm{P}}(W_i = 0) \cdot \tilde{\delta}_{gain} + \hat{\mathrm{P}}(W_i = 1) \cdot \tilde{\delta}_{loss}. \tag{8}$$

This result has important implications for the interpretation of $\hat{\delta}_{OLS}$. Consider, for example, the problem of analyzing gender wage gaps. Intuitively, in a typical study, the proportions of male and female workers are roughly similar.[6] In this case, $\hat{\delta}_{OLS} \simeq \tilde{\delta}_{gap}$. If instead we are interested in the average wage loss for women, δ_{loss}, we need to use a different method.

Conversely, when we focus on black–white gaps in economic outcomes, the disadvantaged group (i.e., blacks) also constitutes a numerical minority, at least in the United States. In this case, $\hat{\delta}_{OLS} \simeq \tilde{\delta}_{loss}$, and hence the interpretation of regression estimates is substantially different. If we are interested in estimating the average outcome gap, δ_{gap}, a different method must be chosen.

Of course, blacks do not constitute a numerical minority in all studies of black–white differences in economic outcomes. Sometimes we might intentionally focus on a population that is predominantly black. For example, Stiefel, Schwartz, and Ellen (2006) analyzed test score gaps in a big-city school district. In some countries, such as South Africa, blacks are both disadvantaged and a numerical majority (Sherer 2000; Allanson and Atkins 2005). In either of these cases, regression estimates would be similar to an estimate of the average gain for whites, $\hat{\delta}_{OLS} \simeq \tilde{\delta}_{gain}$, although this parameter is less likely to be of direct interest.

---

[5]The exact expressions for $\tilde{\delta}_{gain}$ and $\tilde{\delta}_{loss}$ are identical to those for $\tilde{\tau}_{ATT}$ and $\tilde{\tau}_{ATC}$, respectively. Although it might be difficult to conceptualize the "propensity score" for race or other demographic characteristics, it does not matter for this definition.

[6]See, for example, Blau and Beller (1988), Weinberger and Kuhn (2010), and Blau and Kahn (2017). Note, however, that none of these three studies restricts its attention to such simple regression estimates.

Oaxaca–Blinder Decompositions

The simplest solution to this problem with regression estimates is to allow the regression coefficients to be different for both groups of interest:

$$Y_i = X_i\beta_1 + \upsilon_{1i} \text{ if } W_i = 1 \quad \text{and} \quad Y_i = X_i\beta_0 + \upsilon_{0i} \text{ if } W_i = 0. \tag{9}$$

Also, $\mathrm{E}(\upsilon_{1i}|X_i, W_i) = \mathrm{E}(\upsilon_{0i}|X_i, W_i) = 0$. In this case, the raw mean difference in outcomes, $\delta_{raw} = \mathrm{E}(Y_i|W_i = 1) - \mathrm{E}(Y_i|W_i = 0)$, can be decomposed as:

$$\delta_{raw} = \mathrm{E}(X_i|W_i = 1) \cdot (\beta_1 - \beta_0) + [\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_0, \tag{10}$$

where the first element, $\mathrm{E}(X_i|W_i = 1) \cdot (\beta_1 - \beta_0)$, reflects intergroup differences in regression coefficients and is often referred to as *the unexplained component*, while the second element, $[\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_0$, reflects intergroup differences in mean covariate values and is often referred to as *the explained component*. Similarly:

$$\delta_{raw} = \mathrm{E}(X_i|W_i = 0) \cdot (\beta_1 - \beta_0) + [\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_1. \tag{11}$$

The difference between Equations (10) and (11) rests on using alternate comparison coefficients to calculate the explained component as well as measuring the distance between the regression functions, $\beta_1 - \beta_0$, for a different set of covariate values. Moreover, Equations (10) and (11) recover the average gain for advantaged individuals and the average loss for disadvantaged individuals, respectively:

$$\delta_{gain} = \mathrm{E}(X_i|W_i = 1) \cdot (\beta_1 - \beta_0) \text{ and } \delta_{loss} = \mathrm{E}(X_i|W_i = 0) \cdot (\beta_1 - \beta_0). \tag{12}$$

Traditionally, the decomposition literature regards the choice of the comparison coefficients in this context—in other words, the choice between Equations (10) and (11)—as necessarily ambiguous. A number of studies have suggested alternative solutions to this comparison group choice problem. Such an approach is often referred to as "generalized" Oaxaca–Blinder, and it involves an alternative decomposition:

$$\begin{aligned}\delta_{raw} = {} & \mathrm{E}(X_i|W_i = 1) \cdot (\beta_1 - \beta_c) + \mathrm{E}(X_i|W_i = 0) \cdot (\beta_c - \beta_0) \\ & + [\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_c,\end{aligned} \tag{13}$$

where β_c is the set of comparison coefficients. In the context of decomposing differences in wages, these coefficients are typically referred to as the "nondiscriminatory" or "competitive" wage structure. Note that if $\beta_c = \beta_1 = \beta_0$, then there is no unexplained component, because $\beta_1 = \beta_0$ implies that both groups have the same conditional mean.

As noted previously, several papers have suggested alternative comparison coefficients for Equation (13). These coefficients are often of the form $\beta_c = \lambda \cdot \beta_1 + (1 - \lambda) \cdot \beta_0$, where $\lambda \in [0, 1]$ is a weighting factor. If $\lambda = 0$, then the disadvantaged group is used as reference, $\beta_c = \beta_0$, and Equation

(13) simplifies to Equation (10). Similarly, if $\lambda = 1$, then the advantaged group is used as reference, $\beta_c = \beta_1$, and Equation (13) simplifies to Equation (11). Alternatively, Reimers (1983) suggested $\lambda = \frac{1}{2}$ and Cotton (1988) suggested $\lambda = \mathrm{P}(W_i = 1)$, the proportion of advantaged individuals. Moreover, in the context of wage gaps, Neumark (1988) developed a simple model of Beckerian discrimination and showed that identification of the nondiscriminatory wage structure is ensured, for example, in a case in which the utility function of the representative producer is homogeneous of degree zero with respect to the labor inputs of advantaged and disadvantaged workers. Such a wage structure can be approximated by regression coefficients in a pooled model that excludes the indicator for group membership (Neumark 1988). Although this solution constitutes the most popular alternative to the basic decomposition (Weichselbaumer and Winter-Ebmer 2005), it has been criticized by both Fortin (2008) and Elder, Goddeeris, and Haider (2010). They argued that exclusion of the indicator for group membership can bias coefficients on other covariates, which also affects the unexplained component. Therefore, Fortin (2008) proposed using—as the comparison wage structure—the coefficients from a pooled model that includes this variable, such as β in Equation (1) or γ in Equation (5). By construction, the unexplained component in such a decomposition is equal to the coefficient on the indicator for group membership in the corresponding pooled model, such as α in Equation (1) or δ in Equation (5).

Recovering the Average Outcome Gap

A number of studies (Barsky et al. 2002; Black, Haviland, Sanders, and Taylor 2006, 2008; Melly 2006; Fortin et al. 2011; Kline 2011) have noted that the unexplained component in Equation (10) can be interpreted as τ_{ATT}, as long as a potential outcome model is invoked. In a noncausal framework, the basic decomposition recovers δ_{gain} or δ_{loss}, as in Equation (12). It is natural to ask whether there exists an alternative decomposition, perhaps a version of Equation (13), such that its unexplained component can be interpreted as τ_{ATE} or δ_{gap}. In other words, we wish to determine whether a particular choice of β_c, or maybe of λ, implies that

$$\begin{aligned}(14)\qquad \delta_{gap} &= \mathrm{E}(X_i)\cdot(\beta_1-\beta_0) = \mathrm{E}(X_i|W_i=1)\cdot(\beta_1-\beta_c)\\ &\quad + \mathrm{E}(X_i|W_i=0)\cdot(\beta_c-\beta_0).\end{aligned}$$

In fact, this result follows from the choice of $\lambda = \mathrm{P}(W_i = 0)$, as stated in Proposition 1.

Proposition 1 (Oaxaca–Blinder and the Average Outcome Gap): The unexplained component of the Oaxaca–Blinder decomposition in Equation (13) is equal to the average outcome gap, δ_{gap}, if $\beta_c = \mathrm{P}(W_i = 0)\cdot\beta_1 + \mathrm{P}(W_i = 1)\cdot\beta_0$. Then, Equation (13) takes the form

$$\delta_{raw} = \delta_{gap} + [\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_c.$$

A proof of Proposition 1 follows immediately from simple algebra. Perhaps surprisingly, the choice of $\lambda = \mathrm{P}(W_i = 0)$ implies that the proportion of advantaged individuals is used to weight the coefficients for disadvantaged individuals, and that the proportion of disadvantaged individuals is used to weight the coefficients for advantaged individuals. Although this weighting scheme may at first appear counterintuitive, both sets of coefficients play a clearly defined role in this decomposition—as the counterfactual for the other group (Assumption 1). This is why more weight should be put on the coefficients of the smaller group, which are used to provide the counterfactual for the larger one.[7]

Interestingly, this alternative decomposition is equivalent to a flexible linear regression model for the average treatment effect, discussed in Imbens and Wooldridge (2009) and Wooldridge (2010). If W_i now denotes the treatment indicator, τ_{ATE} can also be recovered as the coefficient on W_i in the regression of Y_i on 1, W_i, X_i, and $W_i \cdot [X_i - \mathrm{E}(X_i)]$. As noted by Imbens and Wooldridge (2009), this model implies that

$$\begin{aligned}(15)\qquad \tau_{ATE} = {} & \mathrm{E}(Y_i|W_i = 1) - \mathrm{E}(Y_i|W_i = 0) - [\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \\ & [\mathrm{P}(W_i = 0) \cdot \beta_1 + \mathrm{P}(W_i = 1) \cdot \beta_0],\end{aligned}$$

which is equivalent to the decomposition in Proposition 1. Similarly, the unexplained component of the decomposition in Equation (10) is equal to the coefficient on W_i in the regression of Y_i on 1, W_i, X_i, and $W_i \cdot [X_i - \mathrm{E}(X_i|W_i = 1)]$, and the unexplained component of the decomposition in Equation (11) is equal to the coefficient on W_i in the regression of Y_i on 1, W_i, X_i, and $W_i \cdot [X_i - \mathrm{E}(X_i|W_i = 0)]$.

Several recent articles have criticized the dependence of traditional decomposition methods on linear conditional means (Barsky et al. 2002; Frölich 2007; Ñopo 2008). Thus, it is useful to clarify that the main insight underlying Proposition 1 is unrelated to the linearity assumptions in Equation (9). If we write the counterfactual conditional mean as $\mu_c(x) = \lambda \cdot \mu_1(x) + (1 - \lambda) \cdot \mu_0(x)$, we can always decompose δ_{raw} as

$$\begin{aligned}(16)\qquad \delta_{raw} = {} & (1 - \lambda) \cdot \mathrm{E}[\delta(X_i)|W_i = 1] + \lambda \cdot \mathrm{E}[\delta(X_i)|W_i = 0] \\ & + \{\mathrm{E}[\mu_c(X_i)|W_i = 1] - \mathrm{E}[\mu_c(X_i)|W_i = 0]\}.\end{aligned}$$

As before, the choice of $\lambda = \mathrm{P}(W_i = 0)$ and, equivalently, $\mu_c(x) = \mathrm{P}(W_i = 0) \cdot \mu_1(x) + \mathrm{P}(W_i = 1) \cdot \mu_0(x)$ ensures that $\delta_{gap} = (1 - \lambda) \cdot \mathrm{E}[\delta(X_i)|W_i = 1] + \lambda \cdot \mathrm{E}[\delta(X_i)|W_i = 0] = \mathrm{P}(W_i = 1) \cdot \delta_{gain} + \mathrm{P}(W_i = 0) \cdot \delta_{loss}$. Clearly, if one group is "small" and the other is "large," we need to put a "large" weight

[7]Note that Duncan and Leigh (1985) used a similar decomposition in an application to union wage premiums. Oaxaca and Ransom (1988), however, criticized this approach as being "not a very intuitive procedure."

on the conditional mean of the "small" group, as it constitutes the counterfactual conditional mean for the "large" one.

Estimation of δ_{gap}, δ_{gain}, and δ_{loss} also does not require any linearity assumptions, even though they underlie Equations (12) and (14). In general, any of the standard estimators of τ_{ATE} and τ_{ATT} under conditional independence can be used to estimate δ_{gap} and $\delta_{gain}/\delta_{loss}$, respectively. We can probably assume that the better an estimator is for various average treatment effects, the better it also is for various parameters based on conditional average outcome gaps (see, e.g., Fortin et al. 2011). Indeed, several recent studies have used reweighting (Barsky et al. 2002), other methods based on the propensity score (Frölich 2007), matching on covariates (Black et al. 2006, 2008; Ñopo 2008), and regression trees (Mora 2008) to study between-group differences in various outcomes.

Interpreting the Explained Component

Traditionally, decomposition methods have been used to provide estimates of both the unexplained and explained components. The interpretation of the explained components in Equations (10) and (11) is well known. Similarly, it might be useful to clarify the interpretation of the explained component in Proposition 1, $[\mathrm{E}(X_i|W_i=1)-\mathrm{E}(X_i|W_i=0)]\cdot\beta_c$, and Equation (16), $\mathrm{E}[\mu_c(X_i)|W_i=1]-\mathrm{E}[\mu_c(X_i)|W_i=0]$. After simple algebra, it can be shown that if $\mu_c(x)=\mathrm{P}(W_i=0)\cdot\mu_1(x)+\mathrm{P}(W_i=1)\cdot\mu_0(x)$, then

$$\begin{aligned}(17)\qquad \mathrm{E}[\mu_c(X_i)|W_i=1]-\mathrm{E}[\mu_c(X_i)|W_i=0] &= \mathrm{E}[\mu_1(X_i)|W_i=1]\\ &\quad -\mathrm{E}[\mu_1(X_i)]+\mathrm{E}[\mu_0(X_i)]-\mathrm{E}[\mu_0(X_i)|W_i=0].\end{aligned}$$

We can easily interpret both elements of this explained component. The first element, $\mathrm{E}[\mu_1(X_i)|W_i=1]-\mathrm{E}[\mu_1(X_i)]$, is equal to the difference between actual mean outcomes of advantaged individuals and counterfactual mean outcomes that would be observed if the outcomes of the whole population had been determined according to the conditional mean of these individuals. This parameter is also equal to the amount by which actual mean outcomes of advantaged individuals would decrease if their characteristics were the same as those of the whole population. Whenever advantaged individuals have "better" characteristics than disadvantaged individuals, it will be the case that $\mathrm{E}[\mu_1(X_i)|W_i=1]>\mathrm{E}[\mu_1(X_i)]$. Therefore, this element of the explained component will contribute positively to the raw mean difference in outcomes. Similarly, the second element of this explained component, $\mathrm{E}[\mu_0(X_i)]-\mathrm{E}[\mu_0(X_i)|W_i=0]$, can be interpreted as the difference between counterfactual mean outcomes that would be observed if the outcomes of the whole population had been determined according to the conditional mean of disadvantaged individuals, and their actual mean outcomes. This parameter is also equal to the amount by which actual mean outcomes of disadvantaged individuals would increase if their characteristics were the same as those of the whole population. Again, if

advantaged individuals have "better" characteristics than disadvantaged individuals, then $\mathrm{E}[\mu_0(X_i)] > \mathrm{E}[\mu_0(X_i)|W_i = 0]$, and, therefore, this element of the explained component will also contribute positively to the raw mean difference in outcomes. This interpretation is analogous to that of the explained component in other versions of the Oaxaca–Blinder decomposition, but in this case we do not need to interpret the counterfactual conditional mean as "nondiscriminatory" or "competitive."

Of course, the same interpretation holds in the case of the explained component in Proposition 1, $[\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_c$. Namely, if $\beta_c = \mathrm{P}(W_i = 0) \cdot \beta_1 + \mathrm{P}(W_i = 1) \cdot \beta_0$, then this component takes the form

$$[\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_c = [\mathrm{E}(X_i|W_i = 1) - \mathrm{E}(X_i)] \cdot \beta_1 + [\mathrm{E}(X_i) - \mathrm{E}(X_i|W_i = 0)] \cdot \beta_0, \tag{18}$$

which is a linear special case of Equation (17). Fortin et al. (2011) also briefly discussed a similar explained component.

Reinterpreting Reimers (1983), Cotton (1988), and Fortin (2008)

Finally, the logic of Proposition 1 applies also to several versions of the Oaxaca–Blinder decomposition in Reimers (1983), Cotton (1988), and Fortin (2008). We can easily verify that 1) the unexplained component of the Reimers (1983) decomposition is equal to the arithmetic mean of δ_{gain} and δ_{loss}; 2) the unexplained component of the Cotton (1988) decomposition is equal to a weighted mean of δ_{gain} and δ_{loss}, with reversed weights attached to both these parameters (i.e., the proportion of disadvantaged individuals is used to weight δ_{gain} and the proportion of advantaged individuals is used to weight δ_{loss}); and 3) the unexplained component of the Fortin (2008) decomposition is approximately equal to the same parameter. This last interpretation follows from the earlier discussion of regression estimates of between-group differences in economic outcomes. Elder et al. (2010) observed that regression estimates and the unexplained component of the Cotton (1988) decomposition were generally similar. Because of this similarity, they recommended focusing on regression estimates in studies of between-group outcome gaps. In this article, I demonstrate that this similarity is not necessarily an advantage.

To be clear, these interpretations of the Reimers (1983), Cotton (1988), and Fortin (2008) decompositions assume simple counterfactual treatment (Assumption 1), whereas this assumption is not invoked in any of these studies. More precisely, each tries to account for the presence of general equilibrium effects, which are ruled out by Assumption 1, and to derive a counterfactual conditional mean, which would be observed—in the context of wage gaps—if discrimination ceased to exist. It is very difficult, however, to correctly guess the form of this "nondiscriminatory" or "competitive" wage structure—and Reimers (1983), Cotton (1988), and Fortin (2008) did not offer any theoretical basis to rationalize their choices. Here it might be

easier to invoke the assumption of simple counterfactual treatment instead of relying on the general equilibrium approach. In this case, the Reimers (1983), Cotton (1988), and Fortin (2008) decompositions would be problematic.

Black–White Differences in Test Scores and Wages

Clearly, the interpretation of regression estimates of black–white gaps in economic outcomes depends on the relative sizes of black and white subsamples. Still, OLS estimation of the model in Equation (5) constitutes a standard approach in empirical work (Charles and Guryan 2011). While we can always solve this problem using a variety of semi- and nonparametric methods, it might be sufficient to use one of several versions of the Oaxaca–Blinder decomposition. To estimate δ_{gain} or δ_{loss} we need to choose one of the basic decompositions (Oaxaca 1973; Blinder 1973). If instead we focus on δ_{gap}, then we need to choose the new decomposition, as derived in Proposition 1.

These methodological considerations will be illustrated in various empirical applications to black–white differences in test scores and wages. Whenever blacks are a numerical minority, regression estimates will be similar to their average loss. When, however, blacks become a disadvantaged majority, regression estimates will mimic the average gain for whites. Yet, the estimates based on decomposition methods will always have the desired interpretation: $\hat{\delta}_{gap}$, $\hat{\delta}_{gain}$, or $\hat{\delta}_{loss}$.

Black–White Test Score Gaps in ECLS-K

Following Neal and Johnson (1996), labor economists widely agree that a substantial portion of the black–white wage gap is a consequence of differences in premarket factors. Consequently, in an attempt to explain the emergence of this gap, many researchers have focused on education and cognitive development in children. An example is Fryer and Levitt's (2004) influential study of the black–white test score gap in kindergarten and first grade. This research concluded that the gap among incoming kindergartners practically disappeared when controlling for a small number of covariates. It appeared, however, to re-emerge during the first two years of school.

Recent follow-up studies by Bond and Lang (2013) and Penney (2017) have focused on the (lack of) robustness of these conclusions related to the ordinality of test scores. More precisely, Fryer and Levitt (2004) have treated test scores as interval scales, even though this is inappropriate and any monotonic transformation of the test score scale is also a valid scale. On the one hand, using several such transformations, Bond and Lang (2013) have cast doubt on many of the conclusions in Fryer and Levitt (2004). On the other hand, Penney (2017) considered a normalization of test scores that is invariant to monotonic transformations; his preferred estimates were very similar to regression estimates in Fryer and Levitt (2004). In this article, I

ignore the issue of ordinality of test scores and focus on the interpretation of regression estimates of the black–white test score gap as a weighted average of the average gain for whites and the average loss for blacks. In this sense, my analysis should be treated as an illustration of a different methodological issue, rather than a stand-alone contribution to the debate on black–white test score gaps.

All of these previous papers—namely, Fryer and Levitt (2004), Bond and Lang (2013), and Penney (2017)—are based on data from the ECLS-K. The sample included more than 20,000 children who entered kindergarten in 1998. The main outcomes of interest are standardized test scores in math and reading. I borrow the sample and covariate selections from Penney (2017), who followed Fryer and Levitt (2004). However, unlike Penney (2017), I restrict my attention to test scores in the fall and spring of kindergarten and drop individuals whose race is coded as Hispanic, Asian, or other.

Table 1 reports regression estimates of the black–white test score gap and supplements them with estimates of δ_{gap}, δ_{gain}, and δ_{loss}. Blacks are a clear minority in this sample, and they account for 19% of all observations. Hence, in line with Equation (8), $\hat{\delta}_{OLS}$ is relatively similar to the estimated average loss for blacks. These results suggest that in the fall of kindergarten the black–white test score gap is quite small; in fact, blacks enjoy a slight advantage in reading. By the spring of kindergarten, the relative position of blacks worsens: the math gap more than doubles and their advantage in reading shrinks.[8]

At the same time, the minority status of blacks has an additional consequence. Namely, the estimated average gaps and average gains for whites are always very similar. In the case of math test scores, they are also quite different from both $\hat{\delta}_{OLS}$ and $\hat{\delta}_{loss}$. The average gap in math is 42% to 107% larger than suggested by $\hat{\delta}_{OLS}$. The average gap in reading is more similar to $\hat{\delta}_{OLS}$; however, it also suggests a smaller black advantage in the fall of kindergarten.

To be clear, it is not unreasonable to believe that δ_{loss} is the most interesting parameter in this empirical context. It is natural to ask whether the test scores of blacks are significantly different from those of similar whites. However, the fact that $\hat{\delta}_{loss}$ is relatively well approximated by $\hat{\delta}_{OLS}$ is purely a virtue of the small proportion of blacks in the ECLS-K data or, more generally, in the US population. Moreover, if we decided to focus on δ_{gap}, which is also a very useful measure, we would conclude that black disadvantage in kindergarten math scores is substantially larger than suggested by Fryer and Levitt (2004).

Black–White Wage Gaps in CPS

Many studies have documented that the trend toward black–white wage convergence stopped in the mid-1970s or around 1980 (see, e.g., Grogger

[8]Again, given the results in Bond and Lang (2013), such statements need to be treated with caution.

Table 1. Black–White Test Score Gaps in ECLS-K

	Math test scores					
Time of interview	$\hat{\delta}_{OLS}$	$\hat{\delta}_{gap}$	$\hat{\delta}_{gain}$	$\hat{\delta}_{loss}$	$\hat{\mathrm{P}}(W_i = 1)$	N
Fall kindergarten	0.076***	0.157***	0.181***	0.051**	0.811	11,826
	(0.021)	(0.031)	(0.036)	(0.021)		
Spring kindergarten	0.166***	0.235***	0.255***	0.145***	0.813	11,566
	(0.022)	(0.034)	(0.039)	(0.023)		
	Reading test scores					
Time of interview	$\hat{\delta}_{OLS}$	$\hat{\delta}_{gap}$	$\hat{\delta}_{gain}$	$\hat{\delta}_{loss}$	$\hat{\mathrm{P}}(W_i = 1)$	N
Fall kindergarten	–0.080***	–0.064*	–0.060	–0.085***	0.811	11,826
	(0.021)	(0.036)	(0.042)	(0.021)		
Spring kindergarten	–0.027	–0.035	–0.038	–0.023	0.812	11,573
	(0.022)	(0.037)	(0.042)	(0.022)		

Notes: See also Fryer and Levitt (2004), Bond and Lang (2013), and Penney (2017) for more details on these data. All regressions control for gender, age, birth weight, participation in the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), socioeconomic status, the number of books in the home and its square, and two indicators for mother's age at first birth (teenager and age 30 or over). $\hat{\delta}_{OLS}$ is a least squares estimate of δ in Equation (5). $\hat{\delta}_{gap}$, $\hat{\delta}_{gain}$, and $\hat{\delta}_{loss}$ are based on least squares and sample analogue estimation of Equations (12) and (14). $\hat{\mathrm{P}}(W_i = 1)$ is the sample proportion of whites. N is the sample size. Huber–White standard errors are in parentheses. Positive values reflect black disadvantage. ECLS-K, Early Childhood Longitudinal Study-Kindergarten. *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

1996; Chay and Lee 2000; Juhn 2003; Bayer and Charles 2018). While some research has also revealed a sharp decline in the black–white wage gap in the 1990s (Juhn 2003), other studies have not (Elder et al. 2010). Moreover, several recent contributions have concluded that the current magnitude of the racial wage gap in the United States is the largest in several decades (see, e.g., Hirsch and Winters 2014; Bayer and Charles 2018).

In this article, as in Juhn (2003) and Elder et al. (2010), I focus on data from the March CPS, which are distributed by Flood, King, Ruggles, and Warren (2017). I also borrow the sample and covariate selections from Elder et al. (2010) and extend their analysis by 10 years, from 2008 to 2017. Thus, I study a subsample of full-time, full-year working males; this category is defined as those participants who are at least 18 years old, have earned non-zero wage or salary income, and have worked more than 40 weeks a year and 30 hours in a typical week. Following Elder et al. (2010), I also restrict my attention to individuals whose race is coded as either black or white. The outcome variable of interest is the log hourly wage, and the hourly wage is measured as annual earnings divided by annual hours. The set of control variables is relatively sparse and is listed in Table 2.

Table 2 and Figure 1 report the estimates of δ, δ_{gap}, δ_{gain}, and δ_{loss} for each year between 2000 and 2017. It follows immediately that these results corroborate the earlier conclusion that black–white wage convergence in

Table 2. Black–White Wage Gaps in the Current Population Survey (CPS)

	Log hourly wages					
Year	$\hat{\delta}_{OLS}$	$\hat{\delta}_{gap}$	$\hat{\delta}_{gain}$	$\hat{\delta}_{loss}$	$\hat{\mathrm{P}}(W_i = 1)$	N
2000	0.079***	0.086***	0.086***	0.079***	0.918	25,924
	(0.013)	(0.014)	(0.014)	(0.013)		
2001	0.109***	0.126***	0.127***	0.108***	0.902	40,949
	(0.009)	(0.011)	(0.011)	(0.009)		
2002	0.107***	0.126***	0.128***	0.105***	0.902	40,215
	(0.010)	(0.012)	(0.012)	(0.010)		
2003	0.124***	0.146***	0.148***	0.122***	0.907	38,836
	(0.011)	(0.012)	(0.012)	(0.011)		
2004	0.109***	0.136***	0.139***	0.107***	0.908	37,825
	(0.010)	(0.012)	(0.012)	(0.010)		
2005	0.118***	0.110***	0.109***	0.119***	0.906	37,430
	(0.011)	(0.011)	(0.012)	(0.011)		
2006	0.113***	0.133***	0.135***	0.110***	0.910	37,697
	(0.010)	(0.013)	(0.013)	(0.010)		
2007	0.112***	0.124***	0.125***	0.110***	0.905	37,785
	(0.010)	(0.010)	(0.010)	(0.010)		
2008	0.118***	0.130***	0.131***	0.117***	0.902	37,437
	(0.009)	(0.010)	(0.010)	(0.009)		
2009	0.112***	0.119***	0.120***	0.111***	0.902	36,402
	(0.011)	(0.012)	(0.012)	(0.011)		
2010	0.117***	0.116***	0.116***	0.116***	0.899	34,262
	(0.011)	(0.013)	(0.013)	(0.011)		
2011	0.124***	0.138***	0.140***	0.122***	0.902	33,457
	(0.010)	(0.011)	(0.011)	(0.010)		
2012	0.115***	0.127***	0.128***	0.113***	0.904	33,276
	(0.011)	(0.012)	(0.012)	(0.011)		
2013	0.131***	0.138***	0.139***	0.129***	0.903	33,928
	(0.011)	(0.011)	(0.011)	(0.011)		
2014	0.127***	0.131***	0.132***	0.126***	0.901	33,945
	(0.012)	(0.013)	(0.013)	(0.012)		
2015	0.112***	0.116***	0.117***	0.111***	0.894	34,060
	(0.010)	(0.011)	(0.011)	(0.010)		
2016	0.129***	0.136***	0.137***	0.128***	0.891	31,895
	(0.011)	(0.012)	(0.012)	(0.011)		
2017	0.136***	0.131***	0.131***	0.136***	0.892	32,391
	(0.011)	(0.011)	(0.011)	(0.011)		

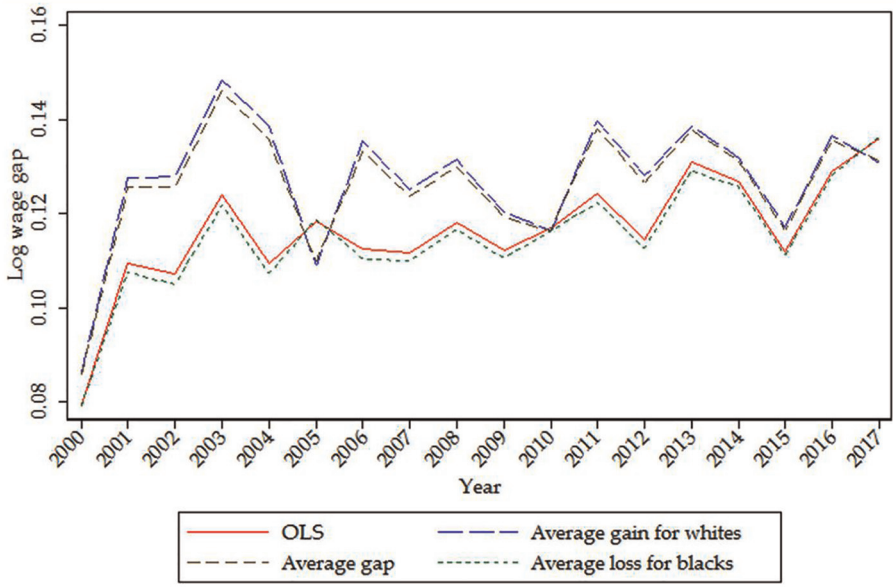
Notes: See also Elder et al. (2010) for more details on these data. All regressions control for a quartic in age, four education categories (no high school diploma, high school diploma either obtained or unclear, three years of college or less, and four years of college or more), and 12 "major occupation" categories listed in the CPS. $\hat{\delta}_{OLS}$ is a least squares estimate of δ in Equation (5). $\hat{\delta}_{gap}$, $\hat{\delta}_{gain}$, and $\hat{\delta}_{loss}$ are based on least squares and sample analogue estimation of Equations (12) and (14). $\hat{\mathrm{P}}(W_i = 1)$ is the sample proportion of whites. N is the sample size. Huber–White standard errors are in parentheses. Positive values reflect black disadvantage.

*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

the United States came to a halt. In fact, all measures of the black–white wage gap were slightly larger in magnitude in 2017 than around 2000.

It should also be noted that, generally speaking, the differences between the average loss for blacks and the average gain for whites are rather small

Figure 1. Black–White Wage Gaps in the Current Population Survey (CPS)



Notes: Numbers are based on point estimates reported in Table 2. Positive values reflect black disadvantage. OLS, ordinary least squares.

in the CPS data, and hence $\hat{\delta}_{OLS}$ is also of the same order of magnitude. Still, the average loss for blacks is typically smaller than the average gain for whites.[9] Because blacks are again a numerical minority, as they account for 8% to 11% of all observations, this translates into a very consistent differential between $\hat{\delta}_{OLS}$ and $\hat{\delta}_{gap}$. Namely, regression estimates understate the average wage gap in most years. As expected, $\hat{\delta}_{OLS}$ is generally indistinguishable from the average loss for blacks; $\hat{\delta}_{gap}$ and $\hat{\delta}_{gain}$ are also practically identical—and larger than $\hat{\delta}_{OLS}$.

Black–White Wage Gaps in NLSY79

A common concern about the CPS data is that it lacks information about some important determinants of wages. In particular, Neal and Johnson (1996) demonstrated that the black–white wage gap nearly disappeared after controlling for age and performance on the Armed Forces Qualifying Test (AFQT). Unsurprisingly, this measure of ability is unavailable in most microeconomic data sets, including CPS. It is recorded, however, as part of the NLSY79, which is a panel study of individuals born between 1957 and

[9]At first, this finding might seem inconsistent with the stylized fact reported in Lang and Lehmann (2012) that black–white wage gaps decrease with education, to the extent that no significant wage differences occur between high-skilled blacks and high-skilled whites. If this is true, then we should expect δ_{gain} to be relatively small, as whites are, on average, more highly educated than blacks. A detailed analysis of this problem, however, is beyond the scope of this article.

Table 3. Black–White Wage Gaps in NLSY79

	Log hourly wages					
Control variables	$\hat{\delta}_{OLS}$	$\hat{\delta}_{gap}$	$\hat{\delta}_{gain}$	$\hat{\delta}_{loss}$	$\hat{P}(W_i = 1)$	N
Age	0.362***	0.363***	0.363***	0.362***	0.857	3,119
	(0.021)	(0.021)	(0.021)	(0.021)		
Age, AFQT	0.088***	0.054*	0.047	0.096***	0.857	3,119
	(0.022)	(0.031)	(0.035)	(0.023)		
Age, AFQT, education	0.149***	0.120***	0.115***	0.151***	0.857	3,119
	(0.022)	(0.029)	(0.031)	(0.023)		
Age, AFQT, other controls	0.052	0.012	0.008	0.056*	0.910	1,586
	(0.033)	(0.047)	(0.050)	(0.034)		
Age, AFQT, education, other controls	0.104***	0.085**	0.083*	0.103***	0.910	1,586
	(0.033)	(0.043)	(0.046)	(0.034)		

Notes: See also Lang and Manove (2011) for more details on these data. "Hourly wages" correspond to mean adjusted wages from the 1996, 1998, and 2000 waves of the survey. "AFQT" includes the AFQT score and its square. "Other controls" include school inputs and family background. School inputs include log of enrollment, log number of teachers, log number of guidance counselors, log number of library books, proportion of teachers with MA/PhD, proportion of teachers who left during the year, and average teacher salary. Family background includes mother's education, father's education, number of siblings, and indicators for whether the respondent was born in the United States, lived in the US at age 14, lived in an urban area at age 14, whether his mother was born in the US, and whether his father was born in the US. $\hat{\delta}_{OLS}$ is a least squares estimate of δ in Equation (5). $\hat{\delta}_{gap}$, $\hat{\delta}_{gain}$, and $\hat{\delta}_{loss}$ are based on least squares and sample analogue estimation of Equations (12) and (14). $\hat{P}(W_i = 1)$ is the sample proportion of whites. N is the sample size. Huber–White standard errors are in parentheses. Positive values reflect black disadvantage. All estimation procedures follow Lang and Manove (2011) in using sampling weights. AFQT, Armed Forces Qualifying Test; NLSY79, National Longitudinal Survey of Youth. *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

1964 that began in 1979 and which was also the source of data in Neal and Johnson (1996).

More recently, Lang and Manove (2011) built a model of educational attainment predicting that, conditional on ability (as proxied by AFQT scores), blacks should receive more education than whites. On the basis of this model, whose predictions are broadly consistent with the NLSY79 data, Lang and Manove (2011) recommended that one should control for both AFQT scores and education when studying black–white differences in wages. Interestingly, when Lang and Manove (2011) augmented the specifications of Neal and Johnson (1996) with education, a substantial black–white wage gap re-emerged.

In this article, I borrow the sample and covariate selections from Lang and Manove (2011: table 5). Because I focus entirely on the black–white gap, I also drop all Hispanics. I study log hourly wages of black and white men from the 1996, 1998, and 2000 waves of the survey. The list of control variables is reported in Table 3, together with regression estimates of the black–white wage gap as well as estimates of δ_{gap}, δ_{gain}, and δ_{loss}. As in previous applications, the proportion of blacks in the NLSY79 data is small; they account for 9% to 14% of all observations. Thus, in line with Equation (8), $\hat{\delta}_{OLS}$ is always very similar to the average loss for blacks. Similarly, $\hat{\delta}_{gap}$ and

$\hat{\delta}_{gain}$ are also hardly distinguishable. Finally, of note is that, unlike in CPS, the average loss for blacks is usually larger than the average gain for whites.

The second and fourth rows of Table 3 correspond to the specifications of Neal and Johnson (1996). It turns out that focusing on the average wage gap, as opposed to regression estimates, would have strengthened their conclusions. Even though $\hat{\delta}_{OLS}$ and $\hat{\delta}_{loss}$ are already quite small in the second and fourth rows, $\hat{\delta}_{gap}$ and $\hat{\delta}_{gain}$ are even smaller; in fact, they are extremely close to zero, and not statistically significant, in the fourth row. In other words, a moderately large set of control variables—including age, AFQT scores, school inputs, and family background—shrinks the average black–white wage gap to (practically) zero.

At the same time, the main conclusion of Lang and Manove (2011) still holds true. When we also control for education, as in the third and fifth rows of Table 3, all measures of the black–white wage gap become substantially larger. Still, $\hat{\delta}_{gap}$ is smaller than the regression estimates, which are similar to those reported in Lang and Manove (2011), but they are both larger than the estimates in the second and fourth rows.

Black–White Wage Gaps in NSW

My results on black–white differences in ECLS-K, CPS, and NLSY79 data share an essential feature: in each case, $\hat{\delta}_{OLS}$ provides a good approximation to $\hat{\delta}_{loss}$. At first, this might seem like a useful property of $\hat{\delta}_{OLS}$, as δ_{loss} is definitely a very interesting parameter. However, as explained earlier, this relationship between $\hat{\delta}_{OLS}$ and $\hat{\delta}_{loss}$ is purely an artifact of the small proportions of blacks represented in ECLS-K, CPS, and NLSY79 data. If instead we focus on an empirical context in which blacks constitute a numerical majority, this supposedly useful property will disappear.

Following LaLonde (1986), Dehejia and Wahba (1999), and Smith and Todd (2005), many studies have used the data on men from the NSW Demonstration, together with non-experimental data sets constructed by LaLonde (1986), to compare the effectiveness of various identification strategies and estimation methods for average treatment effects. In short, NSW was a US work experience program that operated in the mid-1970s and that randomized treatment assignment among eligible participants. This program served a highly disadvantaged population whose members were disproportionately black (Smith and Todd 2005).

As noted previously, these data are typically used to study the effects of the NSW program itself. There is little reason, however, why they should not be used to study black–white wage gaps, although of course the results may not reflect the magnitudes of these gaps in the whole US population. I analyze the original data on the experimental treatment and control groups, as in LaLonde (1986). To be consistent with the previous empirical applications, I focus on log wages and exclude Hispanics; these two restrictions reduce the sample size to 460 individuals, 87% of whom are black.

Table 4. Black–White Wage Gaps in the National Supported Work (NSW) Demonstration

	Log wages in 1978					
Control variables	$\hat{\delta}_{OLS}$	$\hat{\delta}_{gap}$	$\hat{\delta}_{gain}$	$\hat{\delta}_{loss}$	$\hat{\mathrm{P}}(W_i = 1)$	N
Baseline controls	0.121	0.229*	0.098	0.249*	0.130	460
	(0.134)	(0.129)	(0.149)	(0.131)		
+ Nonemployment	0.127	0.258**	0.102	0.282**	0.130	460
	(0.134)	(0.130)	(0.150)	(0.133)		
+ Higher-order terms	0.138	0.188	0.123	0.198	0.130	460
	(0.138)	(0.145)	(0.158)	(0.150)		

Notes: See also LaLonde (1986), Dehejia and Wahba (1999), and Smith and Todd (2005) for more details on these data. "Baseline controls" include age, education, earnings in 1975, and indicators for whether married, whether a high school dropout, and whether treated. "Nonemployment" includes an indicator for whether had zero earnings in 1975. "Higher-order terms" include age squared, age cubed, education squared, and earnings in 1975 squared. $\hat{\delta}_{OLS}$ is a least squares estimate of δ in Equation (5). $\hat{\delta}_{gap}$, $\hat{\delta}_{gain}$, and $\hat{\delta}_{loss}$ are based on least squares and sample analogue estimation of Equations (12) and (14). $\hat{\mathrm{P}}(W_i = 1)$ is the sample proportion of whites. N is the sample size. Huber–White standard errors are in parentheses. Positive values reflect black disadvantage.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 4 reports the estimates of δ, δ_{gap}, δ_{gain}, and δ_{loss}; it also includes the list of control variables. In general, the differences between the average loss for blacks and the average gain for whites are large. This statement is especially true for the first and second rows of Table 4, where we control for a number of baseline covariates (both rows) and employment status in 1975 (second row only).

Unlike previously, the average loss for blacks is not approximated by $\hat{\delta}_{OLS}$ in any useful way. On the contrary, regression estimates, $\hat{\delta}_{OLS}$, are always relatively similar to the average gain for whites. This is, however, a clear implication of Equation (8). When one of two groups is large and the other is small, $\hat{\delta}_{OLS}$ is similar to the "effect" on the smaller group. The difference between $\hat{\delta}_{OLS}$ and $\hat{\delta}_{loss}$ (and also $\hat{\delta}_{gap}$) is particularly striking in the second row of Table 4. While the regression estimate suggests a black–white wage gap of 12.7 log points (which is also not significantly different from zero), the estimated average loss for blacks is 28.2 log points and the estimated average gap is 25.8 log points, more than twice as large as $\hat{\delta}_{OLS}$.[10] These differences are very substantial. When we additionally control for a number of higher-order terms in the third row of Table 4, these differences become smaller, although $\hat{\delta}_{OLS}$ ($\hat{\delta}_{gap}$) remains similar to $\hat{\delta}_{gain}$ ($\hat{\delta}_{loss}$).

[10]In an earlier working paper version of this article, I focused on the subset of the experimental treatment and control groups constructed by Dehejia and Wahba (1999); I also did not exclude Hispanics from the sample. Many of the estimates were quite different than currently reported, although the main message remains unchanged: With a large proportion of blacks, $\hat{\delta}_{OLS}$ is relatively different from $\hat{\delta}_{gap}$ and $\hat{\delta}_{loss}$ but similar to $\hat{\delta}_{gain}$.

Conclusion

In this article, I have borrowed a recent result from the program evaluation literature to demonstrate that the interpretation of regression estimates of between-group differences in economic outcomes necessarily depends on the relative proportions of these groups. If the disadvantaged group is also a numerical minority, as is often the case with blacks, regression estimates will be similar to the average loss for this group. I have demonstrated the empirical relevance of this prediction in applications to black–white test score gaps in ECLS-K data and black–white wage gaps in CPS and NLSY79 data.

Sometimes, however, the disadvantaged group does not constitute a numerical minority, in which case regression estimates will not approximate the average loss for this group. When the majority group is, in fact, disadvantaged—say, blacks in an urban school district, in South Africa, or in NSW data—regression estimates will be similar to the average gain for advantaged individuals. Unfortunately, in most applications, this parameter is also less likely to be of direct interest.

In an intermediate case, where the proportions of both groups are similar—which is to be expected, for example, in a typical study of gender wage gaps—regression estimates will be similar to the average outcome gap. There are reasons to believe that this is an interesting parameter, as it is equal to the difference between mean outcomes in two counterfactual distributions. In the first distribution, outcomes of both groups are determined according to the conditional mean of advantaged individuals. In the second distribution, all outcomes are determined using the conditional mean of disadvantaged individuals.

Of course, instead of relying on regression estimates, researchers may prefer to explicitly choose their parameter of interest. While its estimation would be easy to implement semi- or nonparametrically, following a more traditional approach of using parametric decomposition methods is also possible. If we wish to estimate the average gain for advantaged individuals or the average loss for disadvantaged individuals, we need to use one of the most basic versions of the Oaxaca–Blinder decomposition (Oaxaca 1973; Blinder 1973). If instead we are interested in the average outcome gap, we need to apply the main contribution of this article: a new decomposition whose unexplained component is equal to this parameter. Interestingly, under a particular conditional independence assumption, this object is also equivalent to the average treatment effect.

These decompositions and the framework of this article can be relevant for public policy. For example, Blau and Kahn (2017: 800) explained, without using this particular notation, that $\mu_1(X_i)$ corresponds to the wage a particular woman would receive if her employer was found to have discriminated against women and was now required to treat them identically as it treats men. In this case, the average loss, δ_{loss}, would be useful in determining the total shortfall in female pay at that firm, $N_f \cdot \delta_{loss}$, where N_f is the

number of female employees. Moreover, $N_f \cdot \delta_{loss}$ would also correspond to the potential liability of this firm in a discrimination case, where the plaintiffs are its N_f female workers.

There are also other contexts in which focusing on δ_{gap}, δ_{gain}, or δ_{loss} might be most appropriate. Traditionally, decomposition methods were based on a choice of the "nondiscriminatory" or "competitive" wage structure, which required the researcher to take a stand on what would happen in general equilibrium if discrimination were eradicated. Many researchers seem to forget, however, that only one version of the "generalized" Oaxaca–Blinder decomposition, namely the Neumark (1988) method, offers a functional form of the nondiscriminatory wage structure that results from a theoretical model of the labor market. In all other cases, the comparison wage structure does not seem to have any theoretical underpinnings.

What follows, even if we were potentially interested in such general equilibrium effects, focusing instead on δ_{gap}, δ_{gain}, or δ_{loss} might often be more realistic. Although in this article I avoid referring to the comparison wage structure as "nondiscriminatory" or "competitive," I also believe that nonzero values of δ_{gap}, δ_{gain}, or δ_{loss} might sometimes be interpreted as evidence of discrimination. If we observe all determinants of wages that are also correlated with group membership (Assumption 3), then the existence of systematic differences between observed wages of disadvantaged workers and observed wages of similar advantaged workers directly contradicts the notion of "equal pay for equal work," which is largely synonymous with a lack of wage discrimination.

Finally, important empirical contexts exist, in which the notion of a "nondiscriminatory" or "competitive" conditional mean generally does not apply. Among these are black–white gaps in test scores, infant mortality, as well as comparisons of wage structures across time. In these cases, a partial equilibrium approach is perhaps natural, as is the focus on parameters such as δ_{gap}, δ_{gain}, or δ_{loss}.

Future work might add to our understanding of formal conditions under which causal effects of race, gender, and other immutable characteristics can be identified and estimated (see Kunze 2008, Greiner and Rubin 2011, and Huber 2015 for recent discussions). As already suggested by Fortin et al. (2011), the economic structure behind decomposition methods should also be improved. Finally, understanding the links between the decomposition methods and the program evaluation literature is essential. Following an important review in Fortin et al. (2011), this article has attempted to take this ongoing discussion one step further by providing an interpretation of regression estimates of between-group differences in economic outcomes and developing a new decomposition compatible with the treatment effects framework.

References

Allanson, Paul, and Jonathan P. Atkins. 2005. The evolution of the racial wage hierarchy in post-apartheid South Africa. *Journal of Development Studies* 41(6): 1023–50.

Angrist, Joshua D. 1998. Estimating the labor market impact of voluntary military service using Social Security data on military applicants. *Econometrica* 66(2): 249–88.

Barsky, Robert, John Bound, Kerwin Kofi Charles, and Joseph P. Lupton. 2002. Accounting for the black–white wealth gap: A nonparametric approach. *Journal of the American Statistical Association* 97(459): 663–73.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. Divergent paths: A new perspective on earnings differences between black and white men since 1940. *Quarterly Journal of Economics* 133(3): 1459–1501.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2006. Why do minority men earn less? A study of wage differentials among the highly educated. *Review of Economics and Statistics* 88(1): 300–313.

———. 2008. Gender wage disparities among the highly educated. *Journal of Human Resources* 43(3): 630–59.

Blau, Francine D., and Andrea H. Beller. 1988. Trends in earnings differentials by gender, 1971–1981. *Industrial and Labor Relations Review* 41(4): 513–29.

Blau, Francine D., and John W. Graham. 1990. Black-white differences in wealth and asset composition. *Quarterly Journal of Economics* 105(2): 321–39.

Blau, Francine D., and Lawrence M. Kahn. 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55(3): 789–865.

Blinder, Alan S. 1973. Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* 8(4): 436–55.

Bond, Timothy N., and Kevin Lang. 2013. The evolution of the black–white test score gap in grades K–3: The fragility of results. *Review of Economics and Statistics* 95(5): 1468–79.

Boustan, Leah Platt, and William J. Collins. 2014. The origin and persistence of black–white differences in women's labor force participation. In Leah Platt Boustan, Carola Frydman, and Robert A. Margo (Eds.), *Human Capital in History: The American Record*, pp. 205–40. Chicago: University of Chicago Press.

Card, David, and Alan B. Krueger. 1992. School quality and black–white relative earnings: A direct assessment. *Quarterly Journal of Economics* 107(1): 151–200.

Charles, Kerwin Kofi, and Jonathan Guryan. 2011. Studying discrimination: Fundamental challenges and recent progress. *Annual Review of Economics* 3: 479–511.

Charles, Kerwin Kofi, and Erik Hurst. 2002. The transition to home ownership and the black–white wealth gap. *Review of Economics and Statistics* 84(2): 281–97.

Chay, Kenneth Y., and David S. Lee. 2000. Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black–white wage differentials. *Journal of Econometrics* 99(1): 1–38.

Collins, William J., and Robert A. Margo. 2001. Race and home ownership: A century-long view. *Explorations in Economic History* 38(1): 68–92.

Cotton, Jeremiah. 1988. On the decomposition of wage differentials. *Review of Economics and Statistics* 70(2): 236–43.

Dehejia, Rajeev H., and Sadek Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94(448): 1053–62.

Duncan, Gregory M., and Duane E. Leigh. 1985. The endogeneity of union status: An empirical test. *Journal of Labor Economics* 3(3): 385–402.

Elder, Todd E., and Yuqing Zhou. 2017. The black–white gap in non-cognitive skills among elementary school children. Unpublished working paper. Accessed at https://msu.edu/~telder/Noncog_Skills_Current.pdf.

Elder, Todd E., John H. Goddeeris, and Steven J. Haider. 2010. Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics* 17(1): 284–90.

———. 2016. Racial and ethnic infant mortality gaps and the role of socio-economic status. *Labour Economics* 43: 42–54.

Firpo, Sergio. 2017. Identifying and measuring economic discrimination. *IZA World of Labor* 347: 1–10.

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2017. Integrated Public Use Microdata Series, Current Population Survey: Version 5.0 [dataset]. Minneapolis: University of Minnesota.

Fortin, Nicole M. 2008. The gender wage gap among young adults in the United States: The importance of money versus people. *Journal of Human Resources* 43(4): 884–918.

Fortin, Nicole M., Thomas Lemieux, and Sergio Firpo. 2011. Decomposition methods in economics. In Orley Ashenfelter and David Card (Eds.), *Handbook of Labor Economics*, Vol. 4A, pp. 1–102. Amsterdam and New York: Elsevier.

Frölich, Markus. 2007. Propensity score matching without conditional independence assumption—with an application to the gender wage gap in the United Kingdom. *Econometrics Journal* 10(2): 359–407.

Fryer, Roland G. 2011. Racial inequality in the 21st century: The declining significance of discrimination. In Orley Ashenfelter and David Card (Eds.), *Handbook of Labor Economics*, Vol. 4B, pp. 855–971. Amsterdam and New York: Elsevier.

Fryer, Roland G., and Steven D. Levitt. 2004. Understanding the black–white test score gap in the first two years of school. *Review of Economics and Statistics* 86(2): 447–64.

———. 2006. The black–white test score gap through third grade. *American Law and Economics Review* 8(2): 249–81.

———. 2013. Testing for racial differences in the mental ability of young children. *American Economic Review* 103(2): 981–1005.

Fryer, Roland G., Devah Pager, and Jörg L. Spenkuch. 2013. Racial disparities in job finding and offered wages. *Journal of Law and Economics* 56(3): 633–89.

Greiner, D. James, and Donald B. Rubin. 2011. Causal effects of perceived immutable characteristics. *Review of Economics and Statistics* 93(3): 775–85.

Grogger, Jeff. 1996. Does school quality explain the recent black/white wage trend? *Journal of Labor Economics* 14(2): 231–53.

Hirsch, Barry T., and John V. Winters. 2014. An anatomy of racial and ethnic trends in male earnings in the U.S. *Review of Income and Wealth* 60(4): 930–47.

Holland, Paul W. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81(396): 945–60.

Huber, Martin. 2015. Causal pitfalls in the decomposition of wage gaps. *Journal of Business & Economic Statistics* 33(2): 179–91.

Humphreys, Macartan. 2009. Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities. Unpublished working paper. Accessed at http://www.columbia.edu/~mh2245/papers1/monotonicity7.pdf.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47(1): 5–86.

Juhn, Chinhui. 2003. Labor market dropouts and trends in the wages of black and white men. *Industrial and Labor Relations Review* 56(4): 643–62.

Kline, Patrick. 2011. Oaxaca–Blinder as a reweighting estimator. *American Economic Review: Papers & Proceedings* 101(3): 532–37.

Kunze, Astrid. 2008. Gender wage gap studies: Consistency and decomposition. *Empirical Economics* 35(1): 63–76.

LaLonde, Robert J. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76(4): 604–20.

Lang, Kevin, and Jee-Yeon K. Lehmann. 2012. Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature* 50(4): 959–1006.

Lang, Kevin, and Michael Manove. 2011. Education and labor market discrimination. *American Economic Review* 101(4): 1467–96.

Li, Fan, Kari Lock Morgan, and Alan M. Zaslavsky. 2018. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* 113(521): 390–400.

Melly, Blaise. 2006. Applied quantile regression. Unpublished doctoral dissertation. St. Gallen, Switzerland: University of St. Gallen.

Mora, Ricardo. 2008. A nonparametric decomposition of the Mexican American average wage gap. *Journal of Applied Econometrics* 23(4): 463–85.

Neal, Derek A., and William R. Johnson. 1996. The role of premarket factors in black–white wage differences. *Journal of Political Economy* 104(5): 869–95.

Neumark, David. 1988. Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources* 23(3): 279–95.

Ñopo, Hugo. 2008. Matching as a tool to decompose wage gaps. *Review of Economics and Statistics* 90(2): 290–99.

Oaxaca, Ronald L. 1973. Male–female wage differentials in urban labor markets. *International Economic Review* 14(3): 693–709.

Oaxaca, Ronald L., and Michael R. Ransom. 1988. Searching for the effect of unionism on the wages of union and nonunion workers. *Journal of Labor Research* 9(2): 139–48.

———. 1994. On discrimination and the decomposition of wage differentials. *Journal of Econometrics* 61(1): 5–21.

Penney, Jeffrey. 2017. Test score measurement and the black–white test score gap. *Review of Economics and Statistics* 99(4): 652–56.

Reimers, Cordelia W. 1983. Labor market discrimination against Hispanic and black men. *Review of Economics and Statistics* 65(4): 570–79.

Ritter, Joseph A., and Lowell J. Taylor. 2011. Racial disparity in unemployment. *Review of Economics and Statistics* 93(1): 30–42.

Rothstein, Jesse, and Nathan Wozny. 2013. Permanent income and the black–white test score gap. *Journal of Human Resources* 48(3): 509–544.

Sherer, George. 2000. Intergroup economic inequality in South Africa: The post-apartheid era. *American Economic Review: Papers & Proceedings* 90(2): 317–21.

Słoczyński, Tymon. 2018. A general weighted average representation of the ordinary and two-stage least squares estimands. IZA Discussion Paper No. 11866. Bonn, Germany: Institute of Labor Economics.

Smith, Jeffrey A., and Petra E. Todd. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125(1–2): 305–53.

Stiefel, Leanna, Amy Ellen Schwartz, and Ingrid Gould Ellen. 2006. Disentangling the racial test score gap: Probing the evidence in a large urban school district. *Journal of Policy Analysis and Management* 26(1): 7–30.

Weichselbaumer, Doris, and Rudolf Winter-Ebmer. 2005. A meta-analysis of the international gender wage gap. *Journal of Economic Surveys* 19(3): 479–511.

Weinberger, Catherine J., and Peter J. Kuhn. 2010. Changing levels or changing slopes? The narrowing of the gender earnings gap 1959–1999. *Industrial and Labor Relations Review* 63(3): 384–406.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.