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

TYMON SŁOCZYŃSKI†

Abstract

In this paper I demonstrate, both theoretically and empirically, that the interpretation of regression estimates of between-group differences in economic outcomes depends on the relative sizes of subpopulations under study. When the disadvantaged group is small, regression estimates are similar to its average loss. When this group is instead a numerical majority, regression estimates are similar to the average gain for advantaged individuals. I analyze black–white test score gaps using ECLS-K data and black–white wage gaps using CPS, NLSY79, and NSW data, documenting that the interpretation of regression estimates varies dramatically across applications. Methodologically, I also develop a new version of the Oaxaca–Blinder decomposition whose unexplained component recovers a parameter referred to as *the average outcome gap*. Under a particular conditional independence assumption, this estimand is equivalent to the average treatment effect (ATE). Finally, I provide treatment-effects reinterpretations of the Reimers, Cotton, and Fortin decompositions.

JEL Classification: C21, I24, J15, J31, J71

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

---

*This version: January 26, 2018. I am grateful to Arun Advani, Joshua Angrist, Anna Baranowska-Rataj, Elizabeth Brainerd, Brantly Callaway, Thomas Crossley, Steven Haider, Krzysztof Karbownik, Patrick Kline, Michał Myck, Mateusz Myśliwski, Ronald Oaxaca, Jörg Schwiebert, Gary Solon, Adam Szulc, Joanna Tyrowicz, Glen Waddell, Rudolf Winter-Ebmer, Jeffrey Wooldridge, and seminar and conference participants at many institutions for useful 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. Previous versions of this paper circulated under the titles "Population Average Gender Effects" and "Average Wage Gaps and Oaxaca–Blinder Decompositions."

†Brandeis University & IZA. Correspondence: Department of Economics & International Business School, Brandeis University, MS 021, 415 South Street, Waltham, MA 02453. E-mail: tslocz@brandeis.edu.

1 Introduction

Despite five decades of progress since the civil rights movement, black–white gaps in economic outcomes are very persistent in the United States. A large number of papers study 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 (Barsky *et al.* 2002), cognitive skills (Fryer and Levitt 2004, 2006, 2013; Bond and Lang 2013), non-cognitive skills (Elder and Zhou 2017), infant mortality (Elder *et al.* 2016), and other outcomes.[1] 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 were studied using decomposition methods (Elder *et al.* 2010; Fortin *et al.* 2011; Firpo 2017).[2] However, as noted by Charles and Guryan (2011), in recent empirical work researchers have typically focused on a simpler approach of estimating the following model using ordinary least squares:

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

where Y_i is the outcome under study, B_i is the binary variable which indicates race (1 if black, 0 if white), and X_i is the vector of observed characteristics. Indeed, this simple method is used in a large number of important papers on black–white gaps, including Collins and Margo (2001), Charles and Hurst (2002), Fryer and Levitt (2004, 2006), Clotfelter *et al.* (2009), Fryer (2011), Lang and Manove (2011), Bond and Lang (2013), Fryer and Levitt (2013), Fryer *et al.* (2013), Rothstein and Wozny (2013), Boustan and Collins (2014), Carruthers and Wanamaker (2017), Elder and Zhou (2017), and many others.

In this paper I borrow from the recent program evaluation literature to call this practice into question. As discussed by, among others, Angrist (1998), Humphreys (2009), and Słoczyński (2017), ordinary least squares estimation of a model analogous to (1) does not recover, in general, the average treatment effect (ATE), unless there is no heterogeneity in the effects of the treatment. These results immediately extend to studies of between-

[1] Recent surveys of this topic—and the related problem of racial discrimination—include Charles and Guryan (2011), Fryer (2011), and Lang and Lehmann (2012).

[2] Recent contributions to the decomposition literature have concentrated on semi- and nonparametric analogues of standard decomposition techniques (Barsky *et al.* 2002; Black *et al.* 2006, 2008; Frölich 2007; Mora 2008; Ñopo 2008), extensions to other statistics besides the mean (Juhn *et al.* 1993; DiNardo *et al.* 1996; Machado and Mata 2005; Melly 2005; Firpo *et al.* 2007; Rothe 2010, 2012; Chernozhukov *et al.* 2013), and causal interpretations of decomposition methods (Huber 2015).

group differences in economic outcomes. In particular, Słoczyński (2017) shows that

$$\hat{\tau}_{\text{OLS}} \simeq \hat{\text{P}}(D_i = 0) \cdot \tilde{\tau}_{\text{ATT}} + \hat{\text{P}}(D_i = 1) \cdot \tilde{\tau}_{\text{ATC}}, \tag{2}$$

where D_i is the binary variable which indicates treatment status (1 if treated, 0 if control), $\hat{\tau}_{\text{OLS}}$ is the ordinary least squares estimate of the coefficient on D_i, and $\tilde{\tau}_{\text{ATT}}$ and $\tilde{\tau}_{\text{ATC}}$ are estimates of the average treatment effect on the treated (ATT) and the average treatment effect on the controls (ATC), as discussed in Słoczyński (2017).

This result has important implications for the interpretation of regression estimates of between-group differences in economic outcomes. 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 these individuals also constitute a numerical minority. When instead they 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 a number of applications. First, I study black–white differences in kindergarten test scores using ECLS-K data, similar to Fryer and Levitt (2004, 2006), Bond and Lang (2013), Penney (2017), and many others. My analysis weakens some of the conclusions in Fryer and Levitt (2004) but the general picture remains unchanged. Next, I study black–white differences in wages using CPS (as in Juhn 2003 and Elder *et al.* 2010), NLSY79 (as in Neal and Johnson 1996 and Lang and Manove 2011), and NSW data. In the first two cases, regression estimates are very similar to the average wage loss for blacks. In the last application, however, regression estimates appear to mimic the average wage gain for whites; they also dramatically overstate both the average wage gap and the average wage loss for blacks. The source of this discrepancy is very simple. Participation in the NSW program was intended to provide assistance to a highly disadvantaged population whose members were disproportionately black (Smith and Todd 2005). In this case, however, the interpretation of regression estimates of black–white gaps is substantially different than in the "standard" case where blacks are also a numerical minority.

To address this important issue, I derive a new version of the Oaxaca–Blinder decomposition (Oaxaca 1973; Blinder 1973) whose "unexplained component" can be interpreted as the average treatment effect, which is likely to be the primary object of interest in various empirical contexts.[3] Because the potential outcome model (see, *e.g.*, Holland 1986;

[3]Recently, several researchers (Barsky *et al.* 2002; Black *et al.* 2006, 2008; Melly 2006; Fortin *et al.* 2011;

Imbens and Wooldridge 2009) is rarely invoked in the decomposition literature, I usually refer to this object as *the average outcome gap*—an equivalent parameter which lacks a causal interpretation. It is important to note that this new procedure is distinct from previous versions of the "generalized" Oaxaca–Blinder decomposition (Reimers 1983; Cotton 1988; Neumark 1988; Oaxaca and Ransom 1994; Fortin 2008), although it easily fits into this class of methods. Different members of this class are defined by the choice of *the comparison coefficients* which in turn determine the counterfactual conditional mean with which all actual outcomes are compared. In this paper I study, among other things, whether the average outcome gap can be recovered with some version of the generalized Oaxaca–Blinder decomposition. I derive such a new procedure which uses a linear combination of the regression coefficients for both subpopulations to construct the counterfactual conditional mean. However, these coefficients are weighted in a nonstandard way, namely the proportion of advantaged individuals (*e.g.*, whites) is used to weight the coefficients for disadvantaged individuals (*e.g.*, blacks), and vice versa. Clearly, such a weighting scheme may at first look counterintuitive.[4] Nevertheless, within the framework of this paper the role of each group's coefficients is to serve as the counterfactual for the other group, and therefore we should indeed put more weight on the coefficients for the smaller group in order to recover the average outcome gap. Note that a similar intuition can be applied to provide a reinterpretation of the Reimers (1983), Cotton (1988), and Fortin (2008) decompositions. Each of these procedures is easily shown to recover some generally uninteresting weighted average of conditional outcome gaps.

2 Theory

Consider a population which is 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 unit i, we also observe an outcome, Y_i, and a row vector of observed characteristics, X_i. In this case, $\mu_1(x) = \mathrm{E}(Y_i \mid X_i = x, W_i = 1)$ is the expected outcome of an advantaged individual with $X_i = x$ and $\mu_0(x) = \mathrm{E}(Y_i \mid X_i = x, W_i = 0)$ is the expected outcome of a disadvantaged individual with these characteristics. Moreover, define *the conditional 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

Kline 2011) have noted that the unexplained component in the most basic version of the Oaxaca–Blinder decomposition can sometimes be interpreted as the average treatment effect on the treated.

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

object is also referred to as the conditional average controlled difference by Li *et al.* (2016). 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

$$\delta_{\text{gap}} = \mathrm{E}\left[\delta(X_i)\right]. \tag{3}$$

Within the framework of a potential outcome model, and under additional assumptions, this parameter is equivalent to the average treatment effect. Moreover, define *the average gain for advantaged individuals* and *the average loss for disadvantaged individuals* as

$$\delta_{\text{gain}} = \mathrm{E}\left[\delta(X_i) \mid W_i = 1\right] \qquad \text{and} \qquad \delta_{\text{loss}} = \mathrm{E}\left[\delta(X_i) \mid W_i = 0\right], \tag{4}$$

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

$$\delta_{\text{gap}} = \mathrm{P}(W_i = 1) \cdot \delta_{\text{gain}} + \mathrm{P}(W_i = 0) \cdot \delta_{\text{loss}}. \tag{5}$$

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 is important to note 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 discussed below for completeness.

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 which 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 is important to note that this assumption rules out the presence of general equilibrium effects, and this might be a substantial restriction in some empirical contexts.

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

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, $W_i \perp \varepsilon_i \mid X_i$, 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.[5] Indeed, Card and Krueger (1992) argue that omission of measures of school quality might affect estimates of black–white wage gaps; on the other hand, Grogger (1996) presents a different view.

It is important to note that Assumptions 1 to 3 guarantee identification of the aggregate decomposition (Fortin *et al.* 2011). If we maintain these assumptions, it becomes possible to construct a counterfactual distribution which would be observed if 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, outcomes of all individuals are determined according to the conditional mean of advantaged individuals; in the second distribution, 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 *(i)* actual outcomes of these individuals and *(ii)* 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

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

average gap between *(i)* their counterfactual outcomes which would be observed if these outcomes were determined according to the conditional mean of advantaged individuals and *(ii)* 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. On the other hand, 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. Especially, the average outcome gap—a noncausal equivalent of the average treatment effect—is likely to be the primary object of interest in many empirical studies. It is intuitively appealing to compare mean outcomes of *all individuals* in two counterfactual distributions, which differ only in the choice of the conditional mean that is used to generate these counterfactual outcomes.

Regression Estimates

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

$$Y_i = \delta W_i + X_i \beta + \varepsilon_i. \tag{6}$$

Now, unlike in equation (1), the disadvantaged group is the omitted category. This ensures that the sign of δ is consistent with the signs of δ_{gap}, δ_{gain}, and δ_{loss}. Moreover, it is a straightforward extension of a result in Słoczyński (2017) that

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

where $\hat{\pi}$ is decreasing in $\hat{\mathrm{P}}(W_i = 0)$.[6] 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, $\hat{\pi}$ is equal to $\hat{\mathrm{P}}(W_i = 1)$. What follows,

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

This result has important implications for the interpretation of $\hat{\delta}$. Consider, for example, the problem of analyzing gender wage gaps. Intuitively, in a typical study, proportions of male and female workers are roughly similar (see, *e.g.*, Blau and Beller 1988; Weinberger

[6] The exact expressions for $\tilde{\delta}_{\text{gap}}$, $\tilde{\delta}_{\text{gain}}$, and $\tilde{\delta}_{\text{loss}}$ also follow from Słoczyński (2017).

and Kuhn 2010; Blau and Kahn 2017). In this case, $\hat{\delta} \simeq \tilde{\delta}_{\text{gap}}$. If instead we are interested in the average wage loss for women, δ_{loss}, we need to use a different method.

On the other hand, 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 (see, *e.g.*, Elder *et al.* 2010). In this case, $\hat{\delta} \simeq \tilde{\delta}_{\text{loss}}$, 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 which is also disproportionately black. For example, Stiefel *et al.* (2006) analyze 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 the average gain for whites, $\hat{\delta} \simeq \tilde{\delta}_{\text{gain}}$, while this parameter is less likely to be of direct interest.

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:

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

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

$$\begin{aligned} \delta_{\text{raw}} &= \mathrm{E}(X_i \mid W_i = 1) \cdot (\beta_1 - \beta_0) \\ &+ [\mathrm{E}(X_i \mid W_i = 1) - \mathrm{E}(X_i \mid W_i = 0)] \cdot \beta_0, \end{aligned} \tag{10}$$

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

$$\begin{aligned} \delta_{\text{raw}} &= \mathrm{E}(X_i \mid W_i = 0) \cdot (\beta_1 - \beta_0) \\ &+ [\mathrm{E}(X_i \mid W_i = 1) - \mathrm{E}(X_i \mid W_i = 0)] \cdot \beta_1. \end{aligned} \tag{11}$$

The difference between equations (10) and (11) rests upon 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:

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

Traditionally, the decomposition literature regarded 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 papers suggest 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:

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

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 $\beta_{\text{c}} = \beta_1 = \beta_0$, then there is no unexplained component, because $\beta_1 = \beta_0$ implies that both groups have the same conditional mean.

As noted previously, a number of papers suggest alternative comparison coefficients for equation (13). These coefficients are often of the form $\beta_{\text{c}} = \lambda \cdot \beta_1 + (1 - \lambda) \cdot \beta_0$, where $\lambda \in [0, 1]$ is a weighting factor. If $\lambda = 0$, then the disadvantaged group is used as reference, $\beta_{\text{c}} = \beta_0$, and equation (13) simplifies to equation (10). Similarly, if $\lambda = 1$, then the advantaged group is used as reference, $\beta_{\text{c}} = \beta_1$, and equation (13) simplifies to equation (11). Alternatively, Reimers (1983) suggests $\lambda = \frac{1}{2}$ and Cotton (1988) suggests $\lambda = \mathrm{P}(W_i = 1)$, the proportion of advantaged individuals. Moreover, in the context of wage gaps, Neumark (1988) develops a simple model of Beckerian discrimination and shows that identification of the nondiscriminatory wage structure is ensured, for example, if the utility function of the representative producer is homogeneous of degree zero with respect to labor inputs of advantaged and disadvantaged workers. Such a wage structure can be approximated by regression coefficients in a pooled model which 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 is criticized by both Fortin (2008) and Elder *et al.* (2010), as exclusion of the indicator for group membership can bias coefficients on other covariates which also affects the unexplained component. Therefore, Fortin (2008) proposes to use a pooled model including this variable as the comparison wage structure. By construction, the unexplained component in such a decomposition is equal to the coefficient on the indicator for group membership in a pooled regression.

Recovering the Average Outcome Gap

A number of papers (Barsky *et al.* 2002; Black *et al.* 2006, 2008; Melly 2006; Fortin *et al.* 2011; Kline 2011) note 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

$$\begin{aligned} \delta_{\mathrm{gap}} &= \mathrm{E}(X_i) \cdot (\beta_1 - \beta_0) \\ &= \mathrm{E}(X_i \mid W_i = 1) \cdot (\beta_1 - \beta_{\mathrm{c}}) \;+\; \mathrm{E}(X_i \mid W_i = 0) \cdot (\beta_{\mathrm{c}} - \beta_0)\,. \end{aligned} \tag{14}$$

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* $\beta_{\mathrm{c}} = \mathrm{P}(W_i = 0) \cdot \beta_1 + \mathrm{P}(W_i = 1) \cdot \beta_0$*. Then, equation (13) takes the form*

$$\delta_{\mathrm{raw}} \;=\; \delta_{\mathrm{gap}} \;+\; [\mathrm{E}(X_i \mid W_i = 1) - \mathrm{E}(X_i \mid W_i = 0)] \cdot \beta_{\mathrm{c}}.$$

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 the proportion of disadvantaged individuals is used to weight the coefficients for advantaged individuals. Although this weighting scheme may at first look counterintuitive, both sets of coefficients play a clearly defined role in this decomposition—as the counterfactual for the other group (Assumption 1). This is exactly the reason why more weight should be put on the coefficients of the smaller group which are used to provide the counterfactual for the larger one.

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

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

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 $W_i \cdot [X_i - \mathrm{E}(X_i \mid W_i = 1)]$ 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 $W_i \cdot [X_i - \mathrm{E}(X_i \mid W_i = 0)]$.

Several recent papers criticize 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 $\mu_{\text{c}}(x) = \lambda \cdot \mu_1(x) + (1 - \lambda) \cdot \mu_0(x)$, we can always decompose δ_{raw} as

$$\begin{aligned} \delta_{\text{raw}} &= (1 - \lambda) \cdot \mathrm{E}\left[\delta(X_i) \mid W_i = 1\right] \; + \; \lambda \cdot \mathrm{E}\left[\delta(X_i) \mid W_i = 0\right] \\ &+ \; \{\mathrm{E}\left[\mu_{\text{c}}(X_i) \mid W_i = 1\right] - \mathrm{E}\left[\mu_{\text{c}}(X_i) \mid W_i = 0\right]\} \,. \end{aligned} \tag{16}$$

As before, the choice of $\lambda = \mathrm{P}(W_i = 0)$ and, equivalently, $\mu_{\text{c}}(x) = \mathrm{P}(W_i = 0) \cdot \mu_1(x) + \mathrm{P}(W_i = 1) \cdot \mu_0(x)$ ensures that $\delta_{\text{gap}} = (1 - \lambda) \cdot \mathrm{E}\left[\delta(X_i) \mid W_i = 1\right] + \lambda \cdot \mathrm{E}\left[\delta(X_i) \mid W_i = 0\right] = \mathrm{P}(W_i = 1) \cdot \delta_{\text{gain}} + \mathrm{P}(W_i = 0) \cdot \delta_{\text{loss}}$. Clearly, if one group is "small" and the other is "large," we need to put a "large" weight 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 are present in 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 $\delta_{\text{gain}}/\delta_{\text{loss}}$, respectively. We can probably assume that the better an estimator is for various average treatment effects, the better it is also for various averages of conditional outcome gaps (see, *e.g.*, Fortin *et al.* 2011). Indeed, several recent applications use matching on covariates (Black *et al.* 2006, 2008; Ñopo 2008), methods based on the propensity score (Frölich 2007), reweighting (Barsky *et al.* 2002), and regression trees (Mora 2008) to study between-group differences in various outcomes.

Interpreting the Explained Component

Traditionally, decomposition methods were 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, $[\mathrm{E}(X_i \mid W_i = 1) - \mathrm{E}(X_i \mid W_i = 0)] \cdot \beta_{\mathrm{c}}$, and equation (16), $\mathrm{E}\left[\mu_{\mathrm{c}}(X_i) \mid W_i = 1\right] - \mathrm{E}\left[\mu_{\mathrm{c}}(X_i) \mid W_i = 0\right]$. Interestingly, after simple algebra, it can be shown that if $\mu_{\mathrm{c}}(x) = \mathrm{P}(W_i = 0) \cdot \mu_1(x) + \mathrm{P}(W_i = 1) \cdot \mu_0(x)$, then

$$\begin{aligned} \mathrm{E}\left[\mu_{\mathrm{c}}(X_i) \mid W_i = 1\right] - \mathrm{E}\left[\mu_{\mathrm{c}}(X_i) \mid W_i = 0\right] &= \mathrm{E}\left[\mu_1(X_i) \mid W_i = 1\right] - \mathrm{E}\left[\mu_1(X_i)\right] \\ &\quad + \mathrm{E}\left[\mu_0(X_i)\right] - \mathrm{E}\left[\mu_0(X_i) \mid W_i = 0\right]. \end{aligned} \tag{17}$$

We can easily interpret both elements of this explained component. The first element, $\mathrm{E}\left[\mu_1(X_i) \mid W_i = 1\right] - \mathrm{E}\left[\mu_1(X_i)\right]$, is equal to the difference between actual mean outcomes of advantaged individuals and counterfactual mean outcomes which would be observed if the whole population had their outcomes determined according to the conditional mean of these individuals. This 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 $\mathrm{E}\left[\mu_1(X_i) \mid W_i = 1\right] > \mathrm{E}\left[\mu_1(X_i)\right]$. 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, $\mathrm{E}\left[\mu_0(X_i)\right] - \mathrm{E}\left[\mu_0(X_i) \mid W_i = 0\right]$, can be interpreted as the difference between counterfactual mean outcomes which would be observed if the whole population had their outcomes determined according to the conditional mean of disadvantaged individuals and actual mean outcomes of these individuals—and this is 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 $\mathrm{E}\left[\mu_0(X_i)\right] > \mathrm{E}\left[\mu_0(X_i) \mid W_i = 0\right]$, and therefore this element of the explained component will also contribute positively to the raw mean difference in outcomes. This is analogous to the interpretation 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, $[\mathrm{E}(X_i \mid W_i = 1) - \mathrm{E}(X_i \mid W_i = 0)] \cdot \beta_{\mathrm{c}}$. Namely, if $\beta_{\mathrm{c}} = \mathrm{P}(W_i = 0) \cdot \beta_1 +$

$P(W_i = 1) \cdot \beta_0$, then this component takes the form

$$[E(X_i \mid W_i = 1) - E(X_i \mid W_i = 0)] \cdot \beta_c = [E(X_i \mid W_i = 1) - E(X_i)] \cdot \beta_1 + [E(X_i) - E(X_i \mid W_i = 0)] \cdot \beta_0, \quad (18)$$

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

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). It can be easily verified that *(i)* the unexplained component of the Reimers (1983) decomposition is equal to the arithmetic mean of δ_{gain} and δ_{loss}; *(ii)* 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 *(iii)* 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. A related point is made in Elder *et al.* (2010) who recommend, however, focusing on regression estimates, as they are similar to the unexplained component of the Cotton (1988) decomposition. In this paper I demonstrate that this is not necessarily an advantage.

To be clear, these interpretations of the Reimers (1983), Cotton (1988), and Fortin (2008) decompositions are based on the assumption of simple counterfactual treatment (Assumption 1), while this assumption is not invoked in any of these papers. More precisely, each of these papers 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) do not offer any theoretical basis to rationalize their choices. In this case it might be easier to invoke the assumption of simple counterfactual treatment instead of relying on the general-equilibrium approach—in which case the Reimers (1983), Cotton (1988), and Fortin (2008) decompositions would be problematic.

3 Black–White Differences in Test Scores and Wages

It is now clear that regression estimates of black–white gaps in economic outcomes have an interpretation that is dependent on the relative sizes of black and white subsamples. Still, ordinary least squares estimation of the model in (6) 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 a number of 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. On the other hand, the estimates based on decomposition methods will always have the desired interpretation: $\hat{\delta}_{\text{gap}}$, $\hat{\delta}_{\text{gain}}$, or $\hat{\delta}_{\text{loss}}$.

Black–White Test Score Gaps in ECLS-K

Following Neal and Johnson (1996), it has been widely agreed among labor economists that a substantial portion of the black–white wage gap is a consequence of differences in premarket factors. Consequently, in a search for an explanation of the emergence of this gap, many papers have focused on education and cognitive development in children. For example, in an influential paper, Fryer and Levitt (2004) study the black–white test score gap in kindergarten and first grade; strikingly, they conclude that the gap among incoming kindergartners practically disappears when we control for a small number of covariates. This gap, however, appears to reemerge during the first two years of school.

Recent follow-up studies by Bond and Lang (2013) and Penney (2017) focus on the (lack of) robustness of these conclusions that is related to the ordinality of test scores. More precisely, Fryer and Levitt (2004) treat test scores as interval scales, even though this is inappropriate and any monotonic transformation of the test score scale is also a valid scale. Considering a number of such transformations, Bond and Lang (2013) present a very pessimistic view of the main conclusions in Fryer and Levitt (2004). On the other hand, Penney (2017) corroborates this earlier study; his preferred estimates are very similar to regression estimates in Fryer and Levitt (2004).

All of these papers are based on data from the Early Childhood Longitudinal Study kindergarten cohort (ECLS-K). The sample includes more than 20,000 children who en-

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

	Math test scores				
	$\hat{\delta}_{\text{OLS}}$	$\hat{\delta}_{\text{gap}}$	$\hat{\delta}_{\text{gain}}$	$\hat{\delta}_{\text{loss}}$	
Fall kindergarten	0.068***	0.113***	0.123***	0.051***	$\hat{\text{P}}(W_i = 1) = 0.861$
	(0.020)	(0.028)	(0.030)	(0.019)	$N = 16{,}097$
Spring kindergarten	0.152***	0.194***	0.203***	0.137***	$\hat{\text{P}}(W_i = 1) = 0.863$
	(0.021)	(0.030)	(0.032)	(0.021)	$N = 15{,}823$

	Reading test scores				
	$\hat{\delta}_{\text{OLS}}$	$\hat{\delta}_{\text{gap}}$	$\hat{\delta}_{\text{gain}}$	$\hat{\delta}_{\text{loss}}$	
Fall kindergarten	–0.107***	–0.064*	–0.055	–0.115***	$\hat{\text{P}}(W_i = 1) = 0.854$
	(0.021)	(0.034)	(0.037)	(0.020)	$N = 15{,}310$
Spring kindergarten	–0.048**	–0.019	–0.014	–0.050***	$\hat{\text{P}}(W_i = 1) = 0.858$
	(0.022)	(0.033)	(0.037)	(0.021)	$N = 15{,}315$

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, WIC participation, socioeconomic status, two indicators for mother's age at first birth (teenager and age 30 or over), the number of books in the home and its square, and three additional race categories (Hispanic, Asian, and other). $\hat{\delta}_{\text{OLS}}$ is a least squares estimate of δ in equation (6). $\hat{\delta}_{\text{gap}}$, $\hat{\delta}_{\text{gain}}$, and $\hat{\delta}_{\text{loss}}$ are based on least squares and sample analogue estimation of equations (12) and (14). 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.

tered kindergarten in 1998. The main outcomes of interest are standardized test scores in math and reading. In this paper I borrow the sample and covariate selections from Penney (2017) who follows Fryer and Levitt (2004). I also restrict my attention to test scores in the fall and spring of kindergarten.

Table 1 reproduces the estimates of δ from Penney (2017) 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 14–15% of all observations. Hence, in line with equation (8), $\hat{\delta}_{\text{OLS}}$ is very 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 widens and their advantage in reading shrinks.

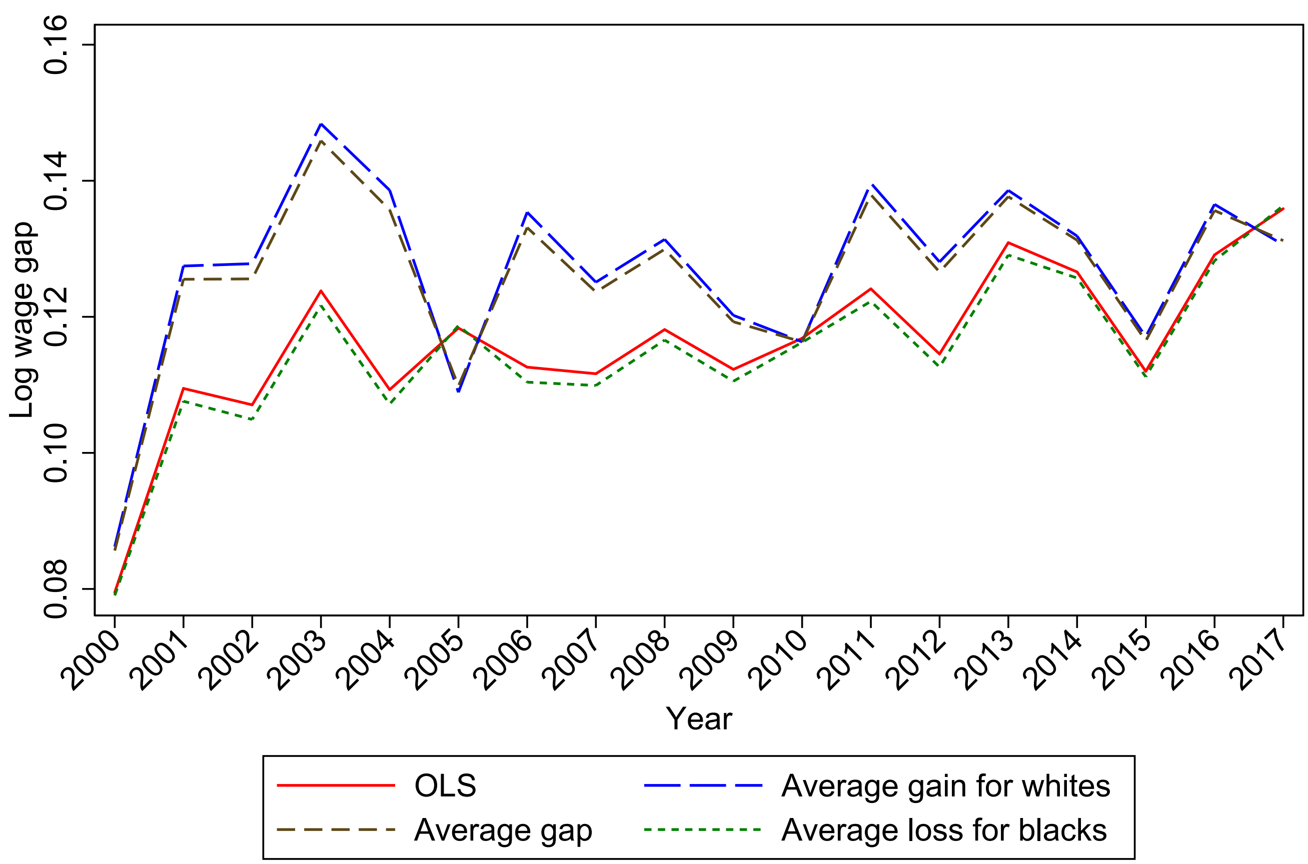
At the same time, the minority status of blacks has an additional consequence. Namely, the estimated average gaps and average gains for whites are always very similar. In fact, they are also quite different from both $\hat{\delta}_{\text{OLS}}$ and $\hat{\delta}_{\text{loss}}$. The average gap in math is 28–66% larger than suggested by $\hat{\delta}_{\text{OLS}}$. The average gap in reading is much closer to zero than $\hat{\delta}_{\text{OLS}}$; it is also not significantly different from zero in the spring 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 $\hat{\delta}_{\text{loss}}$ is well approximated by $\hat{\delta}_{\text{OLS}}$ is purely a virtue of the small proportion of blacks in the ECLS-K data or, more generally, in the U.S. 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 is substantially larger than suggested by Fryer and Levitt (2004).

Black–White Wage Gaps in CPS

A large number of papers document that the trend towards black–white wage convergence stopped in mid-1970s or around 1980 (see, *e.g.*, Grogger 1996; Chay and Lee 2000; Juhn 2003; Bayer and Charles 2016). While some studies also reveal a sharp decline of the black–white wage gap in the 1990s (Juhn 2003), other papers do not (Elder *et al.* 2010). Moreover, several recent contributions conclude that the current magnitude of the racial wage gap in the United States is the largest in several decades (see, *e.g.*, Hirsch and Win-

Figure 1: Black–White Wage Gaps in CPS



Notes: Numbers are based on point estimates reported in Table 2. Positive values reflect black disadvantage.

Table 2: Black–White Wage Gaps in CPS

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

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, 3 years of college or less, and 4 years of college or more), and twelve "major occupation" categories listed in the CPS. $\hat{\delta}_{\text{OLS}}$ is a least squares estimate of δ in equation (6). $\hat{\delta}_{\text{gap}}$, $\hat{\delta}_{\text{gain}}$, and $\hat{\delta}_{\text{loss}}$ are based on least squares and sample analogue estimation of equations (12) and (14). 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.

ters 2014; Bayer and Charles 2016).

In this paper, as in Juhn (2003) and Elder *et al.* (2010), I focus on data from the March Current Population Surveys (CPS), which are distributed by Flood *et al.* (2017). I also borrow the sample and covariate selections from Elder *et al.* (2010), also extending 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 observations who are at least 18 years old, have earned nonzero wage or salary income, and have worked strictly more than 40 weeks a year and 30 hours in a typical week. 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.

Figure 1 and Table 2 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 the U.S. 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 in the CPS data, and hence $\hat{\delta}_{\text{OLS}}$ is also of the same order of magnitude. Still, the average loss for blacks is typically smaller than the average gain for whites.[7] Because blacks are again a numerical minority, as they account for 8–11% of all observations, this translates into a very consistent differential between $\hat{\delta}_{\text{OLS}}$ and $\hat{\delta}_{\text{gap}}$. Namely, regression estimates understate the average wage gap in most years. As expected, $\hat{\delta}_{\text{OLS}}$ is generally indistinguishable from the average loss for blacks; $\hat{\delta}_{\text{gap}}$ and $\hat{\delta}_{\text{gain}}$ are also practically identical—and larger than $\hat{\delta}_{\text{OLS}}$.

Black–White Wage Gaps in NLSY79

A common concern about the CPS data is that it does not contain information about some important determinants of wages. In particular, Neal and Johnson (1996) demonstrate that the black–white wage gap nearly disappears 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 National Longitudinal Survey of Youth (NLSY79), which is a panel study of individuals born between 1957 and 1964 that began in 1979 and is also

[7] At first, this might seem inconsistent with a stylized fact reported in Lang and Lehmann (2012) that black–white wage gaps decrease with education—to the extent that there are no significant wage differences between high-skilled blacks and high-skilled whites. If this is true, then we should expect δ_{gain} to be relatively small, and not large, as whites are, on average, more highly educated than blacks. A detailed analysis of this problem, however, is beyond the scope of this paper.

Table 3: Black–White Wage Gaps in NLSY79

	Log hourly wages				
	$\hat{\delta}_{\text{OLS}}$	$\hat{\delta}_{\text{gap}}$	$\hat{\delta}_{\text{gain}}$	$\hat{\delta}_{\text{loss}}$	
Age, Hispanic	0.362***	0.350***	0.348***	0.361***	$\hat{P}(W_i = 1) = 0.866$
	(0.021)	(0.020)	(0.020)	(0.021)	$N = 3{,}841$
Age, Hispanic, AFQT	0.089***	0.054*	0.048	0.096***	$\hat{P}(W_i = 1) = 0.866$
	(0.021)	(0.030)	(0.033)	(0.022)	$N = 3{,}841$
Age, Hispanic, AFQT, education	0.151***	0.121***	0.116***	0.153***	$\hat{P}(W_i = 1) = 0.866$
	(0.021)	(0.028)	(0.030)	(0.022)	$N = 3{,}841$
Age, Hispanic, AFQT, other controls	0.056*	0.017	0.013	0.061*	$\hat{P}(W_i = 1) = 0.914$
	(0.033)	(0.046)	(0.049)	(0.033)	$N = 1{,}876$
Age, Hispanic, AFQT, education, other controls	0.109***	0.085**	0.083*	0.109***	$\hat{P}(W_i = 1) = 0.914$
	(0.032)	(0.042)	(0.044)	(0.033)	$N = 1{,}876$

Notes: See also Lang and Manove (2011) for more details on these data. "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 U.S., lived in the U.S. at age 14, lived in an urban area at age 14, whether his mother was born in the U.S., and whether his father was born in the U.S. $\hat{\delta}_{\text{OLS}}$ is a least squares estimate of δ in equation (6). $\hat{\delta}_{\text{gap}}$, $\hat{\delta}_{\text{gain}}$, and $\hat{\delta}_{\text{loss}}$ are based on least squares and sample analogue estimation of equations (12) and (14). 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 source of data in Neal and Johnson (1996).

More recently, Lang and Manove (2011) build a model of educational attainment which predicts that, conditional on ability (as proxied by AFQT scores), blacks should get more education than whites. On the basis of this model—whose predictions are broadly consistent with the NLSY79 data—Lang and Manove (2011) recommend that one should control for both AFQT scores and education when studying black–white differences in wages. Interestingly, when Lang and Manove (2011) augment the specifications of Neal and Johnson (1996) with education, a substantial black–white wage gap reemerges.

In this paper I borrow the sample and covariate selections from Table 5 in Lang and Manove (2011). What follows, I study log hourly wages of black men from the 1996, 1998, and 2000 waves of the survey. The list of control variables is reported in Table 3, together with a replication of regression estimates from Lang and Manove (2011) as well as a number of new 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 9–13% of all observations. Thus, in line with equation (8), $\hat{\delta}_{\text{OLS}}$ is always very similar to the average loss for blacks. Similarly, $\hat{\delta}_{\text{gap}}$ and $\hat{\delta}_{\text{gain}}$ are also hardly distinguishable. Finally, it is useful to note that, unlike in CPS, the average loss for blacks is generally 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 $\hat{\delta}_{\text{OLS}}$ and $\hat{\delta}_{\text{loss}}$ are already quite small in the second and fourth rows, $\hat{\delta}_{\text{gap}}$ and $\hat{\delta}_{\text{gain}}$ are even smaller; in fact, they are very 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, family background, and an indicator for whether the respondent is Hispanic—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 additionally 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, $\hat{\delta}_{\text{gap}}$ is visibly smaller than the regression estimates, which are also 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, $\hat{\delta}_{\text{OLS}}$ provides a good approximation to $\hat{\delta}_{\text{loss}}$. At first, this might seem like a useful property of $\hat{\delta}_{\text{OLS}}$, as δ_{loss} is definitely a very interesting parameter. However, as explained before, this relationship between $\hat{\delta}_{\text{OLS}}$ and $\hat{\delta}_{\text{loss}}$ is purely an artifact of the small proportions of blacks 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 papers use the data from the National Supported Work (NSW) Demonstration, together with nonexperimental 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 U.S. work experience program that operated in the mid-1970s and randomized treatment assignment among eligible participants. Also, 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 will not be informative about the magnitudes of these gaps in the whole U.S. population. In this paper I study a subset of the experimental treatment and control groups which was constructed by Dehejia and

Table 4: Black–White Wage Gaps in NSW

	Log wages in 1978				
	$\hat{\delta}_{\text{OLS}}$	$\hat{\delta}_{\text{gap}}$	$\hat{\delta}_{\text{gain}}$	$\hat{\delta}_{\text{loss}}$	
Baseline controls	0.284***	0.230*	0.303***	0.211	$\hat{P}(W_i = 1) = 0.208$
	(0.119)	(0.139)	(0.122)	(0.149)	$N = 308$
+ Nonemployment	0.285***	0.231*	0.304***	0.212	$\hat{P}(W_i = 1) = 0.208$
	(0.120)	(0.138)	(0.124)	(0.147)	$N = 308$
+ Higher-order terms	0.280***	0.117	0.302***	0.069	$\hat{P}(W_i = 1) = 0.208$
	(0.119)	(0.149)	(0.126)	(0.169)	$N = 308$

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 months 13–24 prior to randomization, earnings in 1975, and indicators for whether married, whether a high school dropout, and whether treated. "Nonemployment" includes indicators for whether had zero earnings in months 13–24 prior to randomization and whether had zero earnings in 1975. "Higher-order terms" include age squared, age cubed, education squared, and squares of both earnings variables. $\hat{\delta}_{\text{OLS}}$ is a least squares estimate of δ in equation (6). $\hat{\delta}_{\text{gap}}$, $\hat{\delta}_{\text{gain}}$, and $\hat{\delta}_{\text{loss}}$ are based on least squares and sample analogue estimation of equations (12) and (14). 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.

Wahba (1999). To be consistent with the previous empirical applications, I focus on log wages, which reduces the sample size to 308 individuals, 79% of whom are black.[8]

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 third row of Table 4, where we control for the standard set of covariates and a number of higher-order terms.

Unlike previously, the average loss for blacks is not approximated by $\hat{\delta}_{\text{OLS}}$ in any useful way. On the contrary, regression estimates, $\hat{\delta}_{\text{OLS}}$, are always very 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, $\hat{\delta}_{\text{OLS}}$ is similar to the "effect" on the smaller group. The difference between $\hat{\delta}_{\text{OLS}}$ and $\hat{\delta}_{\text{loss}}$ (and also $\hat{\delta}_{\text{gap}}$) is particularly striking in the third row of Table 4. While the regression estimate suggests a black–white wage gap of 28 log points, the estimated average loss for blacks is only 6.9 log points and the estimated average gap is 11.7 log points. These differences are very substantial; in the latter two cases, the estimates are also not significantly different from zero.

[8] As explained by Smith and Todd (2005), it is generally preferable to use the "early random assignment" sample which they also construct. When I consider the same specifications using their data, I observe the same relationship between $\hat{\delta}_{\text{OLS}}$, $\hat{\delta}_{\text{gap}}$, $\hat{\delta}_{\text{gain}}$, and $\hat{\delta}_{\text{loss}}$. However, the sample size is much smaller, and hence the estimates are also quite noisy. Thus, I focus on the data set constructed by Dehejia and Wahba (1999).

4 Summary

In this paper 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. Importantly, 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 in a way that actual outcomes of advantaged individuals currently are. In the second distribution, this is true for outcomes 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, it is also possible to follow a more traditional approach of using parametric decomposition methods. 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 a further contribution of this paper—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.

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), it is also important to improve the economic structure behind decomposition methods. Finally, it is essential to understand the links between the decomposition methods and the program evaluation literature. Following an important review in Fortin *et al.* (2011), this paper 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 which is compatible with the treatment effects framework.

References

ALLANSON, P. & ATKINS, J. P. (2005). The evolution of the racial wage hierarchy in post-apartheid South Africa. *Journal of Development Studies* 41, 1023–1050.

ANGRIST, J. D. (1998). Estimating the labor market impact of voluntary military service using Social Security data on military applicants. *Econometrica* 66, 249–288.

BARSKY, R., BOUND, J., CHARLES, K. K. & LUPTON, J. P. (2002). Accounting for the black–white wealth gap: A nonparametric approach. *Journal of the American Statistical Association* 97, 663–673.

BAYER, P. & CHARLES, K. K. (2016). Divergent paths: Structural change, economic rank, and the evolution of black–white earnings differences, 1940–2014. Unpublished.

BLACK, D. A., HAVILAND, A. M., SANDERS, S. G. & TAYLOR, L. J. (2006). Why do minority men earn less? A study of wage differentials among the highly educated. *Review of Economics and Statistics* 88, 300–313.

BLACK, D. A., HAVILAND, A. M., SANDERS, S. G. & TAYLOR, L. J. (2008). Gender wage disparities among the highly educated. *Journal of Human Resources* 43, 630–659.

BLAU, F. D. & BELLER, A. H. (1988). Trends in earnings differentials by gender, 1971–1981. *Industrial and Labor Relations Review* 41, 513–529.

BLAU, F. D. & KAHN, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55, 789–865.

BLINDER, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* 8, 436–455.

BOND, T. N. & LANG, K. (2013). The evolution of the black–white test score gap in grades K–3: The fragility of results. *Review of Economics and Statistics* 95, 1468–1479.

BOUSTAN, L. P. & COLLINS, W. J. (2014). The origin and persistence of black–white differences in women's labor force participation. In: *Human Capital in History: The American Record* (BOUSTAN, L. P., FRYDMAN, C. & MARGO, R. A., eds.). University of Chicago Press.

CARD, D. & KRUEGER, A. B. (1992). School quality and black–white relative earnings: A direct assessment. *Quarterly Journal of Economics* 107, 151–200.

CARRUTHERS, C. K. & WANAMAKER, M. H. (2017). Separate and unequal in the labor market: Human capital and the Jim Crow wage gap. *Journal of Labor Economics* 35, 655–696.

CHARLES, K. K. & GURYAN, J. (2011). Studying discrimination: Fundamental challenges and recent progress. *Annual Review of Economics* 3, 479–511.

CHARLES, K. K. & HURST, E. (2002). The transition to home ownership and the black–white wealth gap. *Review of Economics and Statistics* 84, 281–297.

CHAY, K. Y. & LEE, D. S. (2000). Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black–white wage differentials. *Journal of Econometrics* 99, 1–38.

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. & MELLY, B. (2013). Inference on counterfactual distributions. *Econometrica* 81, 2205–2268.

CLOTFELTER, C. T., LADD, H. F. & VIGDOR, J. L. (2009). The academic achievement gap in grades 3 to 8. *Review of Economics and Statistics* 91, 398–419.

COLLINS, W. J. & MARGO, R. A. (2001). Race and home ownership: A century-long view. *Explorations in Economic History* 38, 68–92.

COTTON, J. (1988). On the decomposition of wage differentials. *Review of Economics and Statistics* 70, 236–243.

DEHEJIA, R. H. & WAHBA, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94, 1053–1062.

DINARDO, J., FORTIN, N. M. & LEMIEUX, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica* 64, 1001–1044.

DUNCAN, G. M. & LEIGH, D. E. (1985). The endogeneity of union status: An empirical test. *Journal of Labor Economics* 3, 385–402.

ELDER, T. E., GODDEERIS, J. H. & HAIDER, S. J. (2010). Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics* 17, 284–290.

ELDER, T. E., GODDEERIS, J. H. & HAIDER, S. J. (2016). Racial and ethnic infant mortality gaps and the role of socio-economic status. *Labour Economics* 43, 42–54.

ELDER, T. E. & ZHOU, Y. (2017). The black–white gap in non-cognitive skills among elementary school children. Unpublished.

FIRPO, S. (2017). Identifying and measuring economic discrimination. *IZA World of Labor* 347, 1–10.

FIRPO, S., FORTIN, N. M. & LEMIEUX, T. (2007). Decomposing wage distributions using recentered influence function regressions. Unpublished.

FLOOD, S., KING, M., RUGGLES, S. & WARREN, J. R. (2017). *Integrated Public Use Microdata Series, Current Population Survey: Version 5.0 [dataset]*. University of Minnesota.

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

FORTIN, N. M., LEMIEUX, T. & FIRPO, S. (2011). Decomposition methods in economics. In: *Handbook of Labor Economics* (ASHENFELTER, O. & CARD, D., eds.), vol. 4A. Elsevier.

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

FRYER, R. G. (2011). Racial inequality in the 21st century: The declining significance of discrimination. In: *Handbook of Labor Economics* (ASHENFELTER, O. & CARD, D., eds.), vol. 4B. Elsevier.

FRYER, R. G. & LEVITT, S. D. (2004). Understanding the black–white test score gap in the first two years of school. *Review of Economics and Statistics* 86, 447–464.

FRYER, R. G. & LEVITT, S. D. (2006). The black–white test score gap through third grade. *American Law and Economics Review* 8, 249–281.

FRYER, R. G. & LEVITT, S. D. (2013). Testing for racial differences in the mental ability of young children. *American Economic Review* 103, 981–1005.

FRYER, R. G., PAGER, D. & SPENKUCH, J. L. (2013). Racial disparities in job finding and offered wages. *Journal of Law and Economics* 56, 633–689.

GREINER, D. J. & RUBIN, D. B. (2011). Causal effects of perceived immutable characteristics. *Review of Economics and Statistics* 93, 775–785.

GROGGER, J. (1996). Does school quality explain the recent black/white wage trend? *Journal of Labor Economics* 14, 231–253.

HIRSCH, B. T. & WINTERS, J. V. (2014). An anatomy of racial and ethnic trends in male earnings in the U.S. *Review of Income and Wealth* 60, 930–947.

HOLLAND, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* 81, 945–960.

HUBER, M. (2015). Causal pitfalls in the decomposition of wage gaps. *Journal of Business & Economic Statistics* 33, 179–191.

HUMPHREYS, M. (2009). Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities. Unpublished.

IMBENS, G. W. & WOOLDRIDGE, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 5–86.

JUHN, C. (2003). Labor market dropouts and trends in the wages of black and white men. *Industrial and Labor Relations Review* 56, 643–662.

JUHN, C., MURPHY, K. M. & PIERCE, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101, 410–442.

KLINE, P. (2011). Oaxaca–Blinder as a reweighting estimator. *American Economic Review: Papers & Proceedings* 101, 532–537.

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

LALONDE, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76, 604–620.

LANG, K. & LEHMANN, J.-Y. K. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature* 50, 959–1006.

LANG, K. & MANOVE, M. (2011). Education and labor market discrimination. *American Economic Review* 101, 1467–1496.

LI, F., MORGAN, K. L. & ZASLAVSKY, A. M. (2016). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* (forthcoming).

MACHADO, J. A. F. & MATA, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* 20, 445–465.

MELLY, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics* 12, 577–590.

MELLY, B. (2006). Applied quantile regression. PhD dissertation.

MORA, R. (2008). A nonparametric decomposition of the Mexican American average wage gap. *Journal of Applied Econometrics* 23, 463–485.

NEAL, D. A. & JOHNSON, W. R. (1996). The role of premarket factors in black–white wage differences. *Journal of Political Economy* 104, 869–895.

NEUMARK, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources* 23, 279–295.

OAXACA, R. L. (1973). Male–female wage differentials in urban labor markets. *International Economic Review* 14, 693–709.

OAXACA, R. L. & RANSOM, M. R. (1988). Searching for the effect of unionism on the wages of union and nonunion workers. *Journal of Labor Research* 9, 139–148.

OAXACA, R. L. & RANSOM, M. R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics* 61, 5–21.

ÑOPO, H. (2008). Matching as a tool to decompose wage gaps. *Review of Economics and Statistics* 90, 290–299.

PENNEY, J. (2017). Test score measurement and the black–white test score gap. *Review of Economics and Statistics* 99, 652–656.

REIMERS, C. W. (1983). Labor market discrimination against Hispanic and black men. *Review of Economics and Statistics* 65, 570–579.

RITTER, J. A. & TAYLOR, L. J. (2011). Racial disparity in unemployment. *Review of Economics and Statistics* 93, 30–42.

ROTHE, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics* 155, 56–70.

ROTHE, C. (2012). Partial distributional policy effects. *Econometrica* 80, 2269–2301.

ROTHSTEIN, J. & WOZNY, N. (2013). Permanent income and the black–white test score gap. *Journal of Human Resources* 48, 509–544.

SHERER, G. (2000). Intergroup economic inequality in South Africa: The post-apartheid era. *American Economic Review: Papers & Proceedings* 90, 317–321.

SMITH, J. A. & TODD, P. E. (2005). Does matching overcome LaLonde's critique of non-experimental estimators? *Journal of Econometrics* 125, 305–353.

SŁOCZYŃSKI, T. (2017). A general weighted average representation of the ordinary and two-stage least squares estimands. Unpublished.

STIEFEL, L., SCHWARTZ, A. E. & ELLEN, I. G. (2006). Disentangling the racial test score gap: Probing the evidence in a large urban school district. *Journal of Policy Analysis and Management* 26, 7–30.

WEICHSELBAUMER, D. & WINTER-EBMER, R. (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys* 19, 479–511.

WEINBERGER, C. J. & KUHN, P. J. (2010). Changing levels or changing slopes? The narrowing of the gender earnings gap 1959–1999. *Industrial and Labor Relations Review* 63, 384–406.

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2nd ed.