Labour Economics 17 (2010) 284–290

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Unexplained gaps and Oaxaca–Blinder decompositions

Todd E. Elder, John H. Goddeeris, Steven J. Haider *

Department of Economics, Michigan State University, 110 Marshall-Adams Hall, East Lansing, MI 48824, United States

ARTICLE INFO

Article history:
Received 7 July 2009
Accepted 2 November 2009
Available online 12 November 2009

JEL classification:
J31
J24
J15
J16

Keywords:
Decompositions
Discrimination

ABSTRACT

We analyze four methods to measure unexplained gaps in mean outcomes: three decompositions based on the seminal work of Oaxaca (1973) and Blinder (1973) and an approach involving a seemingly naïve regression that includes a group indicator variable. Our analysis yields two principal findings. We show that the coefficient on a group indicator variable from an OLS regression is an attractive approach for obtaining a single measure of the unexplained gap. We also show that a commonly-used pooling decomposition systematically overstates the contribution of observable characteristics to mean outcome differences when compared to OLS regression, therefore understating unexplained differences. We then provide three empirical examples that explore the practical importance of our analytic results.

1. Introduction

When faced with a gap in mean outcomes between two groups, researchers frequently examine how much of the gap can be explained by differences in observable characteristics. A common approach to distinguish between explained and unexplained components follows the seminal papers of Oaxaca (1973) and Blinder (1973), with the original "Oaxaca–Blinder" (O–B) decomposition based on separate linear regressions for the two groups. Letting d be an indicator variable for group membership, y^d be the scalar outcome of interest for a member of group d, X^d be a row vector of observable characteristics (including a constant), $\hat{\beta}^d$ be the column vector of coefficients from a linear regression of y^d on X^d, and overbars denote means, it is straightforward to show that

$$\bar{y}^1 - \bar{y}^0 = (\bar{X}^1 - \bar{X}^0)\hat{\beta}^1 + \bar{X}^0(\hat{\beta}^1 - \hat{\beta}^0). \tag{1}$$

In this expression, the first and second terms on the right hand side represent the explained and unexplained components of the difference in mean outcomes, respectively.

Both seminal articles pointed out that the decomposition in Eq. (1) is not unique because an equally compelling alternative decomposition exists:

$$\bar{y}^1 - \bar{y}^0 = (\bar{X}^1 - \bar{X}^0)\hat{\beta}^0 + \bar{X}^1(\hat{\beta}^1 - \hat{\beta}^0). \tag{2}$$

While the first term on the right hand side of Eq. (2) is still interpreted as the explained component, this alternative calculation generally will yield different values from Eq. (1), and there is often little reason to prefer one to the other. Many papers acknowledge this ambiguity by simply reporting both decompositions.

Several papers have proposed alternative O–B decompositions, with perhaps the most widely adopted alternative proposed by Neumark (1988).[1] That paper develops a decomposition based on a pooled regression without group-specific intercepts. It is important to emphasize that Neumark (1988) does not analyze the measurement issue of whether his pooled decomposition or those based on Eqs. (1) and (2) distinguish between explained and unexplained gaps. Rather, he analyzes what fraction of an unexplained wage gap, already purged of productivity differences, represents discrimination, demonstrating that different assumptions regarding employer behavior can lead to each of the three decompositions.[2] Despite this difference in motivation, the pooled decomposition he proposed has been adopted

* Corresponding author. Tel.: +1 517 355 1860.
E-mail addresses: telder@msu.edu (T.E. Elder), goddeeri@msu.edu (J.H. Goddeeris), haider@msu.edu (S.J. Haider).

0927-5371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.labeco.2009.11.002

[1] Other alternatives in the spirit of Oaxaca (1973) and Blinder (1973) have been put forward by Reimers (1983) and Cotton (1988), who both propose decompositions which are convex linear combinations of those given in Eqs. (1) and (2). Oaxaca and Ransom (1994) provide an integrative treatment of the various methods.

[2] Neumark (1988) shows how different assumptions regarding employer preferences lead to different estimates of the wage structure that would prevail in the absence of discrimination, and therefore different estimates of discrimination. His analysis starts from the assumption that the set of observable characteristics is sufficiently rich to remove all productivity differences between the groups of interest, so that any unexplained differences represent discrimination or favoritism. We suspect that few researchers interested in decomposing group differences into explained and unexplained components intend to make such an assumption.

as the primary approach to measure explained and unexplained gaps in a number of empirical studies.[3]

Researchers also routinely use an even simpler approach to measure unexplained gaps. They estimate the pooled regression including an indicator variable for group membership as well as the other observable characteristics, interpreting the coefficient on the group indicator as the unexplained component. For example, this method has been applied to the measurement of union wage premiums (e.g., Lewis, 1986), racial test score gaps (e.g., Fryer and Levitt, 2004), and racial wage gaps (e.g., Neal and Johnson, 1996).

In this paper, we compare these various methods for assessing the unexplained gap in mean outcomes between two groups. Our analysis yields two principal findings. First, we show that the coefficient on the group indicator from a pooled OLS regression is an attractive approach for obtaining a single measure of the unexplained gap. Second, we show that the pooled O–B strategy systematically overstates the role of observables in explaining mean outcome as compared to OLS with a group indicator, thereby understating unexplained differences.[4] The intuition for this result is straightforward: the pooled regression coefficients on observable covariates are biased due to the omission of group-specific intercepts, which in turn causes the role of observables to be overstated. We then provide three empirical examples that explore the practical importance of our analytic results, two based on wage gaps and one based on test score gaps.

2. The relationship among four measures of the unexplained gap

As in the Introduction, let y be the scalar outcome of interest, d be an indicator variable equal to 1 for an individual in group 1 and 0 otherwise, X be the vector of observable characteristics (including a constant but not d), and overbars denote means. We study four different measures of the unexplained gap in $\bar{y}$ between groups 0 and 1. The first two measures come from the standard O–B decompositions listed in Eqs. (1) and (2): define Gap^1 to be $\overline{X}^0(\hat{\beta}^1-\hat{\beta}^0)$, the final term in Eq. (1), and similarly define Gap^0 to be the final term in Eq. (2). The third measure, Gap^{p}, is the unexplained component from Neumark's (1988) proposed decomposition,

$$\bar{y}^1-\bar{y}^0=(\overline{X}^1-\overline{X}^0)\hat{\beta}^p+\overline{X}^1(\hat{\beta}^1-\hat{\beta}^p)+\overline{X}^0(\hat{\beta}^p-\hat{\beta}^0), \tag{3}$$

where $\hat{\beta}^p$ is defined to be the coefficient vector from the pooled regression of y on X. The first term on the right hand side of Eq. (3) is again interpreted as the explained component, and the sum of the final two terms is the unexplained component, Gap^{p}. If y denotes a wage, for example, then these two terms correspond to each group's advantage or disadvantage relative to the pooled wage structure. Finally, the fourth unexplained gap measure, Gap^{OLS}, is the coefficient on d from the pooled OLS regression of y on d and X. We compare these gaps by specifying a population data generating process and then deriving what each of the gaps measure.

2.1. The case in which coefficients are equal across groups

We begin by assuming that the mean outcomes between groups 0 and 1 differ only by a constant and that the outcome is influenced by only one observable characteristic x. These assumptions simplify the exposition substantially, but as we describe below, all of the results in this section extend to the case in which the outcome depends on a vector of characteristics X. We relax the assumption of equal coefficients across groups in the next subsection.

Specifically, suppose the population relationship between y, d, and x is

$$y=\delta_0+\delta_d d+\delta_x x+\varepsilon, \tag{4}$$

with ε orthogonal to d and to x conditional on d.[5] Under these strong assumptions, a sensible definition of the population unexplained gap is δ_d. Moreover, under these assumptions, the probability limit of Gap^{OLS} is δ_d.

To derive probability limits of the other estimates of the unexplained gap, we introduce some additional notation. An O–B unexplained gap can always be written as the difference in overall mean outcomes minus the difference in predicted mean outcomes, and both of these differences can be denoted by linear projections. Letting $b(z|w)$ denote the slope from a linear projection of z on w and a constant, a general expression for an O–B unexplained gap is

$$\begin{aligned}\text{Gap}&=[\bar{y}_1-\bar{y}_0]-[\hat{\theta}(\bar{x}_1-\bar{x}_0)]\\&=b(y|d)-b(\hat{\theta}x|d),\end{aligned} \tag{5}$$

where $\hat{\theta}$ is a coefficient computed from sample data. The choice of $\hat{\theta}$ is what distinguishes different O–B decompositions from each other. For example, Gap^1 is obtained when $\hat{\theta}$ is the OLS slope coefficient from a regression of y on x and a constant using data from group 1, while Gap^0 is obtained when $\hat{\theta}$ is the OLS slope coefficient using data from group 0.

Consider the probability limit of an O–B gap under the data generating process described by Eq. (4):

$$\begin{aligned}\text{plim Gap}&=\text{plim } b(y|d)-b(\hat{\theta}x|d)\\&=\frac{\text{cov}(d,y)}{\text{var}(d)}-\frac{\text{cov}(d,x)}{\text{var}(d)}\text{plim}\hat{\theta}\\&=\frac{\text{cov}(d,\delta_0+\delta_d d+\delta_x x+\varepsilon)}{\text{var}(d)}-\frac{\text{cov}(d,x)}{\text{var}(d)}\text{plim}\hat{\theta}\\&=\delta_d+\frac{\text{cov}(d,x)}{\text{var}(d)}\delta_x-\frac{\text{cov}(d,x)}{\text{var}(d)}\text{plim}\hat{\theta}.\end{aligned} \tag{6}$$

Thus, the estimated gap converges to δ_d whenever plim $\hat{\theta}=\delta_x$. Because plim $\hat{\theta}=\delta_x$ in both group-specific regressions, the probability limits of Gap^0 and Gap^1 are δ_d, implying that Gap^0, Gap^1, and Gap^{OLS} are asymptotically equivalent.

In contrast, Gap^{P} generally will not converge to δ_d. The difference arises because, in a pooled regression that does not include the

[3] For examples of articles that adopt this pooling approach, see Oaxaca and Ransom (1994), Mavromaras and Rudolph (1997), DeLeire (2001), Hersch and Stratton (2002), Jacob (2002), Boden and Galizzi (2003), Gittleman and Wolff (2004), and Yount (2008).

[4] To our knowledge, the only references to this issue in the literature are in Fortin (2006) and Jann (2008). Both papers discuss the idea that omitting a group indicator can transfer some of the unexplained gap to the explained portion. However, neither paper develops a general expression for the bias. In addition, both discuss the issue in a context where the group with better outcomes also has larger values of characteristics positively associated with outcome (e.g., males having more education in a model of male–female wage gaps). It has not previously been shown that the bias relative to pooled OLS with a group indicator exists whenever the groups have different mean characteristics (i.e., whenever O–B decompositions are used).

[5] In regressions with the scalar x, the constant will be denoted separately, while in the more general case X will denote a vector of characteristics including a constant.

group-specific intercept d, plim $\hat{\theta}$ typically does not equal δ_x due to omitted variables bias. To see this formally, consider the probability limit of Gap^{P},

$$\begin{aligned}\text{plim Gap}^{\text{P}} &= \text{plim } b(y|d) - b(xb(y|x)|d)\\ &= \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,[x\text{cov}(x,y)/\text{var}(x)])}{\text{var}(d)}\\ &= \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)} \times \frac{\text{cov}(x,y)}{\text{var}(x)}\\ &= \frac{1}{\text{var}(d)}\left(\text{cov}(d,y) - \frac{\text{cov}(d,x)\text{cov}(x,y)}{\text{var}(y)}\right).\end{aligned} \tag{7}$$

It is useful to compare this expression to an alternative representation of the probability limit of Gap^{OLS}. Defining $\tilde{z}(w)$ to be the component of z that is orthogonal to w in the population (so that $\tilde{z}(w) = z - wb(z|w)$), then

$$\begin{aligned}\text{plim Gap}^{\text{OLS}} &= \text{plim } \hat{\delta}_d\\ &= \frac{\text{cov}(\tilde{d}(x), \tilde{y}(x))}{\text{var}(\tilde{d}(x))}\\ &= \frac{\text{cov}(d,\tilde{y}(x))}{\text{var}(\tilde{d}(x))} - \frac{\text{cov}(x,\tilde{y}(x))}{\text{var}(\tilde{d}(x))} \times \frac{\text{cov}(d,x)}{\text{var}(x)}\\ &= \frac{\text{cov}(d,\tilde{y}(x))}{\text{var}(\tilde{d}(x))}\\ &= \frac{1}{\text{var}(\tilde{d}(x))}\left(\text{cov}(d,y) - \frac{\text{cov}(d,x)\text{cov}(x,y)}{\text{var}(x)}\right),\end{aligned} \tag{8}$$

where the fourth equality follows because $\text{cov}(x, \tilde{y}(x)) = 0$ by the definition of $\tilde{y}(x)$. Comparing Eqs. (7) and (8),

$$\text{plim Gap}^{\text{P}} = \frac{\text{var}(\tilde{d}(x))}{\text{var}(d)} \text{plim Gap}^{\text{OLS}}. \tag{9}$$

The ratio of the two gaps, $\text{var}(\tilde{d}(x))/\text{var}(d)$, equals the probability limit of $(1-R^2)$ from the auxiliary regression of d on x, so the gaps are equivalent only when d is orthogonal to x (in which case observed characteristics explain none of the between-group differences in outcomes). In all other cases, the probability limit of Gap^{P} is smaller than the probability limit of Gap^{OLS}, which we have shown to be equivalent to δ_d and the probability limits of Gap^0 and Gap^1.

The intuition for this result is straightforward. The omission of d from a pooled regression leads to omitted variables bias in the estimated coefficient on x. Because the coefficient on x captures both the direct effect of x on y and the effect of d on y indirectly through the correlation between d and x, it tends to explain “too much” of the gap in outcomes, leading the unexplained gap to be too small. We illustrate this effect in Fig. 1 for the case in which $\bar{x}^1 > \bar{x}^0$, $\bar{y}^1 > \bar{y}^0$, and $\delta_x > 0$. The total gap in mean outcomes is $\bar{y}^1 - \bar{y}^0$, and based on the group 1 regression line (the top line in the figure), the explained gap is $\bar{y}^1 - A$ and the unexplained gap is $A - \bar{y}^0$. Note that the steepness of the line determines the magnitudes of the explained and unexplained gaps, so Gap^1 and Gap^0 are identical because the group 1 and group 0 lines are parallel. In contrast, the regression line for the pooled regression (denoted as the dashed line in the figure) must be steeper than either group line due to omitted variables bias. As a result, Gap^{P} must be less than the other three unexplained gap measures.[6]

[6] Neumark (1988), p. 293, makes a similar point about the case illustrated in Fig. 1. We note, however, that our finding that Gap^{P} is smaller in absolute value than Gap^{OLS} does not require that $\bar{x}_1 > \bar{x}_0$, or that either measure is bounded between zero and the overall difference in mean outcomes.

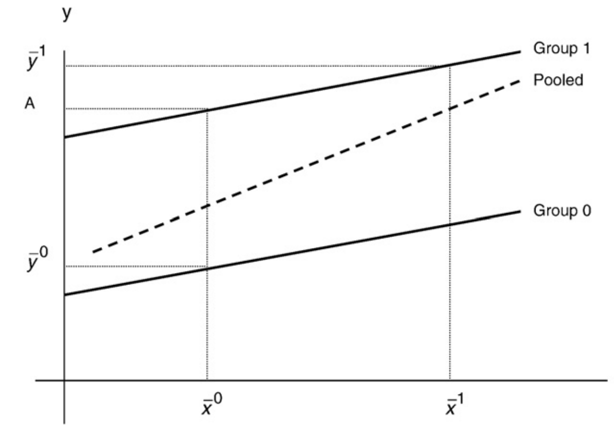

Fig. 1. The relationship between Gap^{P}, Gap^1, and Gap^0.

Finally, in Appendix A1 we show that the asymptotic relationship given in Eq. (9) is also an exact result that holds in finite samples. Further, although we have assumed x is a scalar for notational convenience, the relationship between Gap^{P} and Gap^{OLS} holds when x is vector-valued and regardless of whether model (4) is correct: in all circumstances, Gap^{P} is exactly equal to Gap^{OLS} multiplied by $(1-R^2)$ from the auxiliary regression of d on all observable covariates.[7]

2.2. The case in which coefficients vary across groups

The relationship between Gap^{P} and Gap^{OLS} presented above is exact and general (see Appendix A1). Thus, in the varying coefficients case, Gap^{P} is still systematically less than Gap^{OLS} whenever the averages of observable characteristics differ between the two groups. Turning to the relationship between Gap^{OLS}, Gap^1 and Gap^0, we once again begin by assuming that the outcome is influenced by only one observable characteristic x. We assume again that ε is orthogonal to d and to x conditional on d, but now we allow the coefficient on x to vary between the two groups,

$$y = \lambda_0 + \lambda_d d + \lambda_x x + \lambda_{dx} dx + \varepsilon. \tag{4a}$$

Eq. (6) showed that the probability limit for an O–B unexplained gap based on $\hat{\theta}$ can be written as

$$\text{plim Gap} = \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)} \text{plim } \hat{\theta}. \tag{10}$$

Based on this expression, it is straightforward to see that

$$\text{plim Gap}^1 = \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)} \frac{\text{cov}(x,y|d=1)}{\text{var}(x|d=1)} \tag{11}$$

and

$$\text{plim Gap}^0 = \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)} \frac{\text{cov}(x,y|d=0)}{\text{var}(x|d=0)}. \tag{12}$$

[7] An implication of these results is that, while Gap^{OLS} and Gap^{P} will always have the same sign, the sign of the explained component can differ depending on which approach is used. If $\bar{y}_1 > \bar{y}_0$, $\bar{x}_1 < \bar{x}_0$, and $\delta_x > 0$, then $\bar{y}_1 > \bar{y}_0$ will be smaller than Gap^{OLS} and the associated explained component will be negative. In this situation, $\bar{y}_1 > \bar{y}_0$ may be larger than Gap^{OLS} multiplied by $(1-R^2)$ from the auxiliary regression of d on x. If so, $\bar{y}_1 > \bar{y}_0$ will be larger than Gap^{P}, so that the explained component will be positive. The use of Gap^{P} would therefore imply that observable characteristics explain a positive fraction of an outcome gap, despite the fact that the group with “better” outcomes has “worse” observable characteristics.

As we show in the Appendix A, Gap^{OLS} is a weighted average of Gap^1 and Gap^0,

$$\text{Gap}^{OLS} = \hat{w}^1\text{Gap}^1 + \hat{w}^0\text{Gap}^0, \tag{13}$$

with the weights given by sample analogs of the following:

$$w^1 \equiv \text{plim}\,\hat{w}^1 = \frac{\Pr(d=1)\text{var}(x|d=1)}{\Pr(d=1)\text{var}(x|d=1) + \Pr(d=0)\text{var}(x|d=0)} \tag{14a}$$

$$w^0 \equiv \text{plim}\,\hat{w}^0 = \frac{\Pr(d=0)\text{var}(x|d=0)}{\Pr(d=1)\text{var}(x|d=1) + \Pr(d=0)\text{var}(x|d=0)}. \tag{14b}$$

It is straightforward to show that these weights are bounded by 0 and 1, implying that Gap^{OLS} is always bounded by Gap^1 and Gap^0. In addition, the structure of these weights is intuitively appealing, with Gap^{OLS} approaching Gap^1 for large values of $\text{var}(x|d=1)/\text{var}(x|d=0)$ and for values of $\Pr(d=1)$ close to 1. When $\text{var}(x)$ does not vary across groups, the weights are the sample analogs of $\Pr(d=1)$ and $\Pr(d=0)$ so that Gap^{OLS} is simply the group-size weighted average of Gap^0 and Gap^1, identical to the O–B decomposition proposed by Cotton (1988).

Although Eqs. (14a) and (14b) hold exactly when x is a scalar, they are only suggestive of the relationships among the gaps when X is a vector of characteristics. In the vector case, Gap^{OLS} is not necessarily bounded between Gap^0 and Gap^1.[8] However, the relationship among the three measures depends on similar quantities as in the scalar case, making it unlikely that Gap^{OLS} will deviate far from the bounds; the extent it deviates from the bounds is an empirical issue that we return to below. In the vector case, Gap^{OLS} will still tend to Gap^0 or Gap^1 depending on the relative group sample size and the variance of observables. To see why, recall that Eq. (5) implies that all three gaps can be written as the total gap minus a gap-specific explained portion, $(\overline{X}_1-\overline{X}_0)'\hat{\theta}$, with the choice of $\hat{\theta}$ uniquely determining the gap. Gap^0 is based on $\hat{\beta}_0$, Gap^1 is based on $\hat{\beta}^1$, and Gap^{OLS} is based on $\hat{\beta}^{OLS}$, the slope coefficient on X from a pooled regression of y on X and d. The coefficient vector $\hat{\beta}^{OLS}$ will be closer to $\hat{\beta}^1$ (and therefore Gap^{OLS} will be closer to Gap^1) when group 1 is a larger share of the sample and when the variances of the X's are larger in group 1 than in group 0. Our empirical results in the next section will demonstrate the extent to which Gap^{OLS} deviates from Gap^1 and Gap^0 in three different contexts, as well as the extent to which Gap^P deviates from Gap^{OLS}.

3. Empirical examples

We demonstrate the practical importance of the analytic results shown above by presenting three empirical examples: the male–female wage gap among full-time, full-year workers using Current Population Survey (CPS) data; the white–black wage gap among full-time, full-year working males using CPS data; and the white–black test score gap in kindergarten using the fall 1998 assessment of the Early Childhood Longitudinal Study — Kindergarten Cohort (ECLS-K). For the first two outcomes we first show results for 1985 and 2001, and for the last we present separate results for reading and math test scores.

Table 1
Empirical results.

	White–black log wage gap		Male–female log wage gap		White–black test score gap	
	1985	2001	1985	2001	Math	Reading
N	28,163	40,949	48,499	76,747	13,040	12,374
Total gap	0.254	0.216	0.372	0.285	14.660	11.352
Share in group 1	0.927	0.902	0.598	0.570	0.871	0.865
Gap^1	0.130	0.105	0.346	0.280	0.380	−0.272
Gap^0	0.126	0.129	0.388	0.297	4.103	2.800
Gap^P	0.127	0.105	0.276	0.233	0.680	0.109
Gap^{OLS}	0.131	0.108	0.361	0.294	0.782	0.124
Auxiliary R^2	0.034	0.028	0.238	0.208	0.131	0.122

For both sets of wage gap results, we use a relatively sparse set of regressors, controlling for age, education, and occupation.[9] We define full-time, full-year workers as those who are at least 18 years old and are working more than 30 h a week and 40 weeks a year. The hourly wage is measured as annual earnings divided by annual hours, and all models examine the gap in the log hourly wage. For the male–female results, we include all men (group 1) and women (group 0) and control for whether an individual is black. For the white–black results, we only include males who report being black (group 0) or white (group 1). In analyzing white–black test score differentials, we follow the specifications of Fryer and Levitt (2004), who show that seven covariates are sufficient to explain the entire gap in kindergarten test scores between whites and blacks based on Gap^{OLS}.[10]

We provide the results in Table 1. For each example, we list the sample size, the total gap between the two groups, the four measures of unexplained gaps discussed in the previous section (Gap^1, Gap^0, Gap^P, and Gap^{OLS}), and the R^2 from the auxiliary regression of group status on the other regressors.

These examples illustrate several of the analytic results discussed in the previous section. First, the two standard O–B decompositions can yield dissimilar estimates. Although the results are reasonably similar for the male–female and white–black wage gaps, they lead to noticeably different conclusions for the white–black test score gaps. In particular, Gap^1 (using regression coefficients from the white sample) suggests that only 2.5% (0.380/14.660) of the racial gap in math scores remains unexplained after controlling for this small set of covariates, but nearly 28% (4.103/14.660) of the math gap is unexplained based on Gap^0.

Second, Gap^{OLS} usually lies between Gap^1 and Gap^0, but not always; Gap^{OLS} is outside of the bounds for the white–black wage gap in 1985. In addition, Gap^{OLS} tends to be closer to the bound corresponding to the group that represents a larger fraction of the data. For all four white–black gaps, Gap^{OLS} is very close to the estimate evaluated at the white coefficients (Gap^1), but it is approximately in the middle of Gap^1 and Gap^0 for the male–female wage differential, consistent with the roughly equal shares of males and females in the population.

Third, the deviation between Gap^P and Gap^{OLS} is exactly related to the R^2 from the auxiliary regression of the group indicator on the other explanatory variables ($\text{Gap}^P = (1-R^2)\times\text{Gap}^{OLS}$). Gap^P still falls between Gap^1 and Gap^0 in three cases (in the white–black wage

[8] Consider the generalization of Eq. (4a) in the case in which X may be vector-valued, $y=\lambda_0+\lambda_d d+\lambda_x' X+\lambda_{dx}'(d\times X)+\varepsilon$, where here X does not include a constant (this abuse of notation is useful for readability). As Eq. (5) above implies, $\text{Gap}^1=\hat{\lambda}_d+\overline{X}_0'\hat{\lambda}_{dx}$ and $\text{Gap}^0=\hat{\lambda}_d+\overline{X}_1'\hat{\lambda}_{dx}$. Gap^{OLS}, which is the coefficient on d from a regression omitting the $(d\times X)$ terms, equals $\hat{\lambda}_d+r'\hat{\lambda}_{dx}$, where r is a vector of coefficients on d from auxiliary regressions of the $d\times X$ interactions on X, d, and a constant. In the case of a single x, it can be shown that $r=\hat{w}^1\bar{x}_0+\hat{w}^0\bar{x}_1$, where $\hat{w}^1$ and $\hat{w}^0$ are defined as in Eqs. (14a) and (14b); because r is bounded by $\bar{x}_0$ and $\bar{x}_1$, Gap^{OLS} is bounded by Gap^1 and Gap^0. However, when $\dim(X)>1$, even if each element of r is bounded between the corresponding elements of $\overline{X}_0$ and $\overline{X}_1$, Gap^{OLS} will not necessarily lie between Gap^0 and Gap^1.

[9] We include a quartic in age, 4 education categories (less than high school, high school, some college, and completed college), and 14 occupation categories (the complete "Major Occupation" codes listed in the CPS for these years).

[10] Specifically, we include indicators for whether the mother's age at first birth was over 30 or less than 20, an indicator for whether the mother received WIC payments, a quadratic in the number of books in the home, the child's birthweight in ounces, and an NCES-created summary measure of the family's SES. See Fryer and Levitt (2004) for more details on these measures, and see Appendix Tables 1 and 2 for summary statistics for the estimation samples we use.

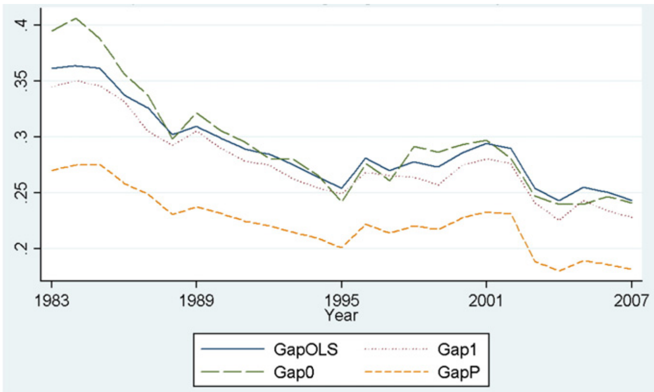

Fig. 2. Unexplained male–female log wage differentials by year, CPS.

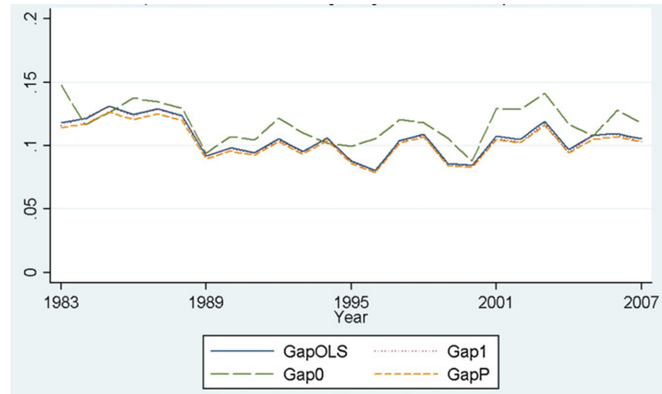

Fig. 3. Unexplained black–white log wage differentials by year, CPS.

differential in 1985 and both test score differences), but in the other three it does not. Gap^{P} is substantially outside the Gap^1 and Gap^0 estimates for both male–female wage gaps because of the high R^2 of the auxiliary regression and the associated attenuation of Gap^{P} relative to Gap^{OLS}.[11]

As further illustration of the relationships among the four Gap measures, Figs. 2 and 3 show the white–black and male–female wage gaps for each year between 1985 and 2007. In the male–female case shown in Fig. 2, the plots of Gap^{OLS}, Gap^1, and Gap^0 are quite similar. Gap^{P} is substantially lower in all years, due to the relatively high power of the covariates in explaining group membership, i.e., men and women are substantially different on observable dimensions. In the white–black case shown in Fig. 3, Gap^0 is consistently larger than Gap^1, but the plots of Gap^{OLS} and Gap^{P} are essentially identical to Gap^1 because the white group represents a large fraction of the population and because the explanatory variables do not predict group membership. In both graphs, Gap^{OLS} lies in between or nearly in between Gap^1 and Gap^0 for every year.

4. Discussion and conclusion

We analyze four methods to measure unexplained gaps in mean outcomes, three based on the decomposition methods of Oaxaca (1973) and Blinder (1973) and one based on a pooled regression with a group indicator variable. Our analysis yields two principal findings. We show that, in the case of a single observable characteristic, the coefficient on the group indicator from a pooled OLS regression is a weighted average of the unexplained gaps from the two standard O–B approaches, with intuitively sensible weights that are bounded between 0 and 1 and sum to 1. The strict bounding result on these weights, however, does not extend to the case when there is more than one regressor. Thus, although the unexplained gap from a pooled OLS regression reflects the overall relationship between the observable characteristics and the outcome variable, this unexplained gap is no longer strictly bounded by the two standard O–B gaps. In contrast, we show that the O–B pooling strategy without a group indicator systematically overstates the contribution of observables to mean outcome differences, therefore understating unexplained differences. Thus, in circumstances where the decompositions are used to separate between explained and unexplained gaps, the O–B pooling strategy systematically fails to do so.

To explore the practical significance of our results, we provide empirical examples involving white–black and male–female wage differentials and the white–black kindergarten test score gap. These examples demonstrate that Gap^{OLS} is typically close to the standard Oaxaca–Blinder unexplained gaps but systematically larger than Gap^{P}. Gap^{P} will deviate from Gap^{OLS} to the extent that there are differences in the means of observable characteristics between the two groups. This deviation can be large enough to drive Gap^{P} substantially below both standard O–B measures, as is the case with the male–female wage differentials.

Taken together, our analytical and empirical results suggest that the pooling O–B decomposition without a group-specific indicator should not be used to distinguish between explained and unexplained gaps, although this method may be useful to assess how much of an unexplained gap represents discrimination if specific assumptions are met. In contrast, Gap^{OLS} provides an attractive summary approach to separate between-group mean differences into explained and unexplained components.

Acknowledgment

We thank Jeff Biddle, Marianne Bitler, Jonah Gelbach, Kevin Hallock, David Neumark, Mathias Sinning, Gary Solon, Mel Stephens, and Steve Woodbury for very useful comments on an initial draft. All errors remain our own, of course. Haider gratefully acknowledges the financial support of the Australian National University as a Gruen Fellow.

Appendix A

A1. The exact relationship between Gap^{OLS} and Gap^{P}

Consider a sample of observations on y, a scalar outcome of interest, d, an indicator variable for group membership, and X, a vector of observed characteristics. For this appendix section define each to be the vector or matrix of deviations from its respective sample mean. Further, define $P = X(X'X)^{-1}X'$ to be the projection matrix onto X and $M = I - P$ to be its complement. Note that we need make no assumptions about relationships in the population.

Gap^{OLS} is then given by

$$\text{Gap}^{\text{OLS}} = (d'Md)^{-1}(d'My). \tag{A1}$$

Similar to Eq. (7) in the text, Gap^{P} can be expressed as the difference of two regression coefficients, one that equals the total gap between the two groups and one that equals the predicted gap, which is constructed using fitted values from a pooled regression of y on X.

[11] Oaxaca and Ransom (1994) found a similar result in their male–female wage example (see their Table 3, column 2), but they did not comment that this result was to be expected.

Therefore,

$$
\begin{aligned}
\text{Gap}^{\text{P}} &= (d'd)^{-1}d'y - (d'd)^{-1}d'Py \\
&= (d'd)^{-1}d'My \\
&= (d'd)^{-1}(d'Md) \times \text{Gap}^{\text{OLS}} \\
&= (1 - R^2_{d,X}) \times \text{Gap}^{\text{OLS}},
\end{aligned} \tag{A2}
$$

where $R^2_{d,X}$ is the R^2 from the auxiliary regression of d on X. As a result, Gap^{P} will always be smaller than Gap^{OLS} except when d is orthogonal to X, which corresponds to the case in which covariates can explain *none* of the difference across groups in average outcomes.

A2. The relationship between Gap^0, Gap^1, and Gap^{OLS} in the scalar x case

We first derive two expressions that will be useful in the final result. Defining $\pi = \Pr(d=1)$, note that

$$
\begin{aligned}
\text{var}(x) &= (1-\pi) \times [\text{var}(x|d=0) + [E(x|d=0) - E(x)]^2] \\
&\quad + \pi \times [\text{var}(x|d=1) + [E(x|d=1) - E(x)]^2] \\
&= (1-\pi) \times [\text{var}(x|d=0) + [E(x|d=1) - E(x|d=0)]^2 \pi^2] \\
&\quad + \pi \times [\text{var}(x|d=1) + [E(x|d=1) - E(x|d=0)]^2 (1-\pi)^2] \\
&= (1-\pi) \times \text{var}(x|d=0) + \pi \times \text{var}(x|d=1) + \frac{\text{cov}(x,d)^2}{\text{var}(d)^2} \\
&\quad \times [(1-\pi)\pi^2 + \pi(1-\pi)^2] \\
&= (1-\pi) \times \text{var}(x|d=0) + \pi \times (x|d=1) + \frac{\text{cov}(x,d)^2}{\text{var}(d)^2} \\
&\quad \times [\pi \text{var}(d) + (1-\pi)\text{var}(d)] \\
&= (1-\pi) \times \text{var}(x|d=0) + \pi \times \text{var}(x|d=1) + \frac{\text{cov}(x,d)^2}{\text{var}(d)}.
\end{aligned} \tag{A3}
$$

The first equality is the decomposition of the variance of x into "within group" and "between group" components, the second equality follows from applying the law of iterated expectations to $E(x)$, the third follows because $E(x|d=1) - E(x|d=0) = \text{cov}(x,d)/\text{var}(d)$ for any binary variable d, and the fourth because $\text{var}(d) = \pi(1-\pi)$ for any binary variable d. Similar logic implies that

$$
\begin{aligned}
\text{cov}(x,y) &= (1-\pi) \times \text{cov}(x,y|d=0) + \pi \times \text{cov}(x,y|d=1) \\
&\quad + \frac{\text{cov}(d,x)\text{cov}(d,y)}{\text{var}(d)}.
\end{aligned} \tag{A4}
$$

Beginning with the result in Eq. (13), we combine Eqs. (11), (12), (14a), and (14b):

$$
\begin{aligned}
\text{plim}\left[\hat{w}^1\text{Gap}^1 + \hat{w}^0\text{Gap}^0\right] &= w^1\left[\frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)}\frac{\text{cov}(x,y|d=1)}{\text{var}(x|d=1)}\right] \\
&\quad + w^0\left[\frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)}\frac{\text{cov}(x,y|d=0)}{\text{var}(x|d=0)}\right] \\
&= \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)}\Pi,
\end{aligned} \tag{A5}
$$

where

$$
\begin{aligned}
\Pi &= \frac{\pi \times \text{var}(x|d=1)}{\pi \times \text{var}(x|d=1) + (1-\pi) \times \text{var}(x|d=0)} \frac{\text{cov}(x,y|d=1)}{\text{var}(x|d=1)} \\
&\quad + \frac{(1-\pi)\text{var}(x|d=0)}{\pi \times \text{var}(x|d=1) + (1-\pi) \times \text{var}(x|d=0)} \frac{\text{cov}(x,y|d=0)}{\text{var}(x|d=0)}.
\end{aligned}
$$

Simplifying Π,

$$
\begin{aligned}
\Pi &= \frac{\pi \times \text{var}(x|d=1)}{\pi \times \text{var}(x|d=1) + (1-\pi) \times \text{var}(x|d=0)} \frac{\text{cov}(x,y|d=1)}{\text{var}(x|d=1)} \\
&\quad + \frac{(1-\pi) \times \text{var}(x|d=0)}{\pi \times \text{var}(x|d=1) + (1-\pi) \times \text{var}(x|d=0)} \frac{\text{cov}(x,y|d=0)}{\text{var}(x|d=0)} \\
&= \frac{\pi \times \text{cov}(x,y|d=1) + (1-\pi) \times \text{cov}(x,y|d=0)}{\pi \times \text{var}(x|d=1) + (1-\pi)\text{var}(x|d=0)} \\
&= \frac{\text{cov}(x,y) - \dfrac{\text{cov}(d,x)\text{cov}(d,y)}{\text{var}(d)}}{\text{var}(x) - \dfrac{\text{cov}(d,x)^2}{\text{var}(d)}}.
\end{aligned} \tag{A6}
$$

The equality on the last line follows from using Eqs. (A3) and (A4) to simplify the preceding line. As a result,

$$
\begin{aligned}
\text{plim}[\hat{w}^1\text{Gap}^1 + \hat{w}^0\text{Gap}^0] &= \frac{\text{cov}(d,y)}{\text{var}(d)} - \frac{\text{cov}(d,x)}{\text{var}(d)}\left[\frac{\text{cov}(x,y) - \dfrac{\text{cov}(d,x)\text{cov}(d,y)}{\text{var}(d)}}{\text{var}(x) - \dfrac{\text{cov}(d,x)^2}{\text{var}(d)}}\right] \\
&= \frac{\text{cov}(d,y)\text{var}(x) - \text{cov}(d,x)\text{cov}(x,y)}{\text{var}(x)\text{var}(d) - \text{cov}(d,x)^2}.
\end{aligned} \tag{A7}
$$

Recall from the text that

$$
\text{plim Gap}^{\text{OLS}} = \frac{1}{\text{var}(\tilde{d}(x))} \times \left(\text{cov}(d,y) - \frac{\text{cov}(d,x)\text{cov}(x,y)}{\text{var}(x)}\right). \tag{A8}
$$

Since $\tilde{d}(x)$ represents the residuals from a population regression of d on x,

$$
\begin{aligned}
\text{var}(\tilde{d}(x)) &= \text{var}\left[d - x\frac{\text{cov}(d,x)}{\text{var}(x)}\right] \\
&= \text{var}(d) - \frac{\text{cov}(d,x)^2}{\text{var}(x)}.
\end{aligned} \tag{A9}
$$

This implies that Eq. (A8) can be rewritten as follows:

$$
\begin{aligned}
\text{plim Gap}^{\text{OLS}} &= \frac{1}{\text{var}(\tilde{d}(x))} \times \left(\text{cov}(d,y) - \frac{\text{cov}(d,x)\text{cov}(x,y)}{\text{var}(x)}\right) \\
&= \frac{\text{var}(x)}{\text{var}(d)(x) - \text{cov}(d,x)^2}\left(\frac{\text{cov}(d,y)\text{var}(x)}{\text{var}(x)} - \frac{\text{cov}(d,x)\text{cov}(x,y)}{\text{var}(x)}\right) \\
&= \frac{\text{cov}(d,y)\text{var}(x) - \text{cov}(d,x)\text{cov}(x,y)}{\text{var}(x)\text{var}(d) - \text{cov}(d,x)^2}.
\end{aligned} \tag{A10}
$$

Comparing Eqs. (A7) and (A10) gives the desired result.

Appendix Table 1
Descriptive statistics for CPS.

	N	Wage	Black	Female	Age	HS+
1983	45,637	8.79	0.08	0.40	38.72	0.84
1984	46,196	9.10	0.08	0.40	38.61	0.85
1985	48,499	9.62	0.09	0.40	38.52	0.85
1986	48,365	10.09	0.09	0.40	38.44	0.86
1987	48,402	10.44	0.09	0.41	38.47	0.86
1988	49,495	10.80	0.09	0.41	38.58	0.87

(continued on next page)

Appendix Table 1 (*continued*)

	N	Wage	Black	Female	Age	HS+
1989	46,741	11.12	0.09	0.41	38.68	0.87
1990	52,015	11.71	0.09	0.41	38.68	0.87
1991	51,402	11.99	0.09	0.41	38.89	0.88
1992	50,018	12.40	0.09	0.43	39.08	0.89
1993	49,405	12.91	0.09	0.43	39.35	0.89
1994	47,948	13.19	0.09	0.43	39.54	0.90
1995	48,839	13.67	0.09	0.42	39.66	0.90
1996	43,719	14.55	0.09	0.42	39.86	0.89
1997	44,727	15.14	0.09	0.42	40.06	0.89
1998	44,941	15.82	0.09	0.43	40.15	0.90
1999	46,314	16.41	0.09	0.43	40.27	0.89
2000	47,551	16.61	0.09	0.43	40.45	0.89
2001	76,647	18.12	0.11	0.43	40.25	0.90
2002	75,429	19.02	0.11	0.43	40.63	0.90
2003	73,809	19.48	0.11	0.43	40.98	0.90
2004	72,531	19.88	0.11	0.43	41.34	.091
2005	71,711	20.39	0.11	0.43	41.40	0.91
2006	72,170	20.97	0.10	0.43	41.48	0.90
2007	72,500	21.93	0.11	0.43	41.69	0.91

Note: Entries are unweighted means of the variables listed in the column headings, listed by survey year. Everyone who worked full-time and full-year is included (at least 30 h per week and 40 weeks per year). Wages are in nominal dollars.

Appendix Table 2
Descriptive statistics for ECLS-K.

	Full sample	Blacks	Whites
N	13,040	1708	11,332
SES composite	0.10	−0.38	0.17
	(0.78)	(0.70)	(0.77)
# books in home	80.78	40.44	87.09
	(59.70)	(39.97)	(59.82)
Mother's age at first birth	24.20	20.93	24.71
	(5.45)	(4.76)	(5.37)
Child's birthweight in pounds	7.42	6.97	7.49
	(1.30)	(1.37)	(1.28)
WIC participation	0.35	0.71	0.29
	(0.48)	(0.45)	(0.46)
Fall K math	53.64	39.21	55.90
	(28.17)	(25.39)	(27.92)
Fall K reading	51.78	41.96	53.31
	(28.72)	(26.95)	(28.69)

Note: Cell entries are unweighted means of the variable listed in the row headings, with standard deviations reported in parentheses. The "Full sample" column includes both black and white kindergarten students in ECLS-K. All covariates are measured as described in Fryer and Levitt (2004).

References

Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. Journal of Human Resources 8 (4), 436–455.
Boden, L.I., Galizzi, M., 2003. Income losses of women and men injured at work. Journal of Human Resources 38 (3), 722–757.
Cotton, J., 1988. On the decomposition of wage differentials. Review of Economics and Statistics 70 (2), 236–243.
DeLeire, T., 2001. Changes in wage discrimination against people with disabilities: 1984–93. Journal of Human Resources 36 (1), 144–158.
Fortin, N.M., 2006. Greed, Altruism, and the Gender Wage Gap. Unpublished manuscript, University of British Columbia.
Fryer, R., Levitt, S., 2004. Understanding the black–white test score gap in the first two years of school. Review of Economics and Statistics 86 (2), 447–464.
Gittleman, M., Wolff, E.N., 2004. Racial differences in patterns of wealth accumulation. Journal of Human Resources 39 (1), 193–227.
Hersch, J., Stratton, L.S., 2002. Housework and wages. Journal of Human Resources 37 (1), 217–229.
Jacob, B.A., 2002. Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. Economics of Education Review 21 (6), 589–598.
Jann, B., 2008. The Blinder–Oaxaca decomposition for linear regression models. Stata Journal 8 (4), 453–479.
Lewis, H.G., 1986. Union Relative Wage Effects: A Survey. The University of Chicago Press, Chicago.
Mavromaras, K.G., Rudolph, H., 1997. Wage discrimination in the reemployment process. Journal of Human Resources 32 (4), 812–860.
Neal, D., Johnson, W., 1996. The role of premarket factors in black–white wage differences. Journal of Political Economy 104 (5), 869–895.
Neumark, D., 1988. Employers' discriminatory behavior and the estimation of wage discrimination. Journal of Human Resources 23 (3), 279–295.
Oaxaca, R., 1973. Male–female wage differentials in urban labor markets. International Economic Review 14 (3), 693–709.
Oaxaca, R.L., Ransom, M.R., 1994. On discrimination and the decomposition of wage differentials. Journal of Econometrics 61 (1), 5–21.
Reimers, C.W., 1983. Labor market discrimination against Hispanic and Black men. Review of Economics and Statistics 65 (4), 570–579.
Yount, K., 2008. Gender, resources across the life course, and cognitive functioning in Egypt. Demography 45 (4), 907–926.