When Do Covariates Matter? And Which Ones, and How Much?

Jonah B. Gelbach*
Department of Economics
Eller College of Management
University of Arizona

June 25, 2009

Abstract

Many authors add variables sequentially to their covariate sets when using linear estimators to investigate the effect of a variable of interest X_1, on some outcome y. One justification for this practice involves robustness: if estimates of the coefficient on X_1 are stable across specifications, then researchers conclude that their findings are robust. A second justification involves accounting: by measuring the difference in X_1's estimated coefficient as they add sets of covariates to the specification, researchers sometimes claim to have measured the effects of covariate variation on this coefficient. In this paper, I show that sequential covariate addition can be very misleading. The relationship between X_1 and a given covariate set may be sensitive to the order in which other covariates have been added. This sensitivity is especially problematic for accounting exercises, as I show using the canonical example of the black-white wage gap. The paper's main contribution is to show how to use the population and sample omitted variables bias formulas to define an economically and econometrically meaningful conditional decomposition that explains how much various covariates account for sensitivity in the estimated coefficient on X_1. I illustrate the conditional decomposition using NLSY data on the black-white wage gap, with interesting empirical results. I also briefly discuss several extensions, including: instrumental variables estimators; the fact that my decomposition nests the Oaxaca-Blinder decomposition; and using the properties of the omitted variables bias formula to construct a Hausman test for cross-specification differences in coefficient estimates under the null that X_1 and X_2 are uncorrelated. Finally, I provide asymptotic variance formulas in an appendix, as well as a link to Stata code that implements my estimators.

*I would like to thank David Card, Mark Duggan, Bill Evans, Steve Haider, Judy Hellerstein, Kei Hirano, Jon Klick, Justin McCrary, Ingmar Prucha, Jim Smith, Alex Whalley, Justin Wolfers, and seminar participants at Berkeley, the BLS, LSU, UC-Irvine, University of Maryland, and RAND for helpful comments and discussions.

1 Introduction

It is a staple of contemporary empirical research, across many fields of economics, to estimate multiple versions of a linear regression model. One starts from a base specification and then adds progressively more covariates. One basic objective of sequential addition is to evaluate the robustness of the coefficient on some variable of interest, X_1, as various covariates in X_2 are partialled out. In this paper, I evaluate what can be learned about robustness from sequential covariate addition exercises. The answer generally is "not much". For example, consider a classic empirical topic, the black-white wage gap. The conditional gap given both education and occupation is not the same population parameter as the conditional black-white wage gap given only education. Different covariate sets map to different probability limits for the coefficient on X_1. Thus, changing the order in which covariates are added changes the difference in the estimated coefficient on X_1 that will be attributed to each covariate set.

A second purpose of this paper to offer a computationally simple and econometrically meaningful alternative to the following accounting question: how much of the change in X_1 coefficients can be attributed to different variables in X_2 as we move from the base specification that has no X_2 covariates to the full specification that has all X_2 covariates? Many cases of interest involve attempts to determine how much of a cross-group difference in some outcome variable's mean is due to group-level heterogeneity in observables, as with the case of the black-white wage gap. In this example, y represents log wages; X_1 is a dummy variable indicating race, and possibly also some other baseline controls; and X_2 contains a variety of observed covariates that are potentially related to wages and may vary systematically across race. The change in the estimated race-dummy coefficient as each new set of covariates is added is then interpreted as being due to variation in the most recently added set of variables.

I show below that using sequential covariate addition in the case of the black-white wage gap gives empirically non-robust answers to the accounting question. For example, depending on the sequence in which education and test score variables are added to the wage equation, education variation either increases the black-white wage gap a little or reduces it substantially. My theoretical discussion shows that this non-robustness occurs because of an empirical fact previously noted by Lang and Manove (2006): the black-white <u>education</u> gap depends critically on whether one conditions on test scores.

Because a robustness measure that is itself not robust to the order of computation is problematic, I develop an alternative, conditional decomposition. This decomposition is based on a simple least-squares identity that links the estimates of the base- and full-specification coefficients on X_1: $\widehat{\beta}_1^{\text{base}} = \widehat{\beta}_1^{\text{full}} + (X_1'X_1)^{-1} X_1'X_2\widehat{\beta}_2$.[1] An advantage of my decomposition is that it yields consistent estimates of economically and econometrically meaningful population parameters. In the context of the black-white wage gap, this derivation gives a clear meaning to the "effect of adding covariates". Whether variation in a covariate increases or reduces the gap depends on (a) whether the covariate has a positive effect on wage, and (b) whether the covariate has a higher mean among blacks or whites (after partialling out other X_1 variables present in the base specification). These basic ideas have a long history in econometrics and statistics. But to my knowledge, the decomposition usefulness of the sample omitted variables bias formula has not previously been recognized.

Since my decomposition is based on parameter estimates computed from the full specification, it is order-invariant. Moreover, it can be implemented by estimating a manageable set of simple auxiliary models and thus requires only minor effort. Also, asymptotic variance formulas are straightforward to derive. I also provide Stata code that implements my decomposition and provides consistent coefficient and variance estimates.[2]

My results show that nothing is gained, and much can be lost, via the practice of sequentially adding covariates. Researchers interested in a formal decomposition of the difference in X_1 coefficient estimates would be better served by determining which covariate sets are of interest when it comes to robustness and then using my conditional decomposition to estimate how much these covariate sets move the coefficient of interest. The same goes for researchers interested in the robustness of the estimated coefficient on X_1 as a general matter. In sum, everything of interest can be learned by properly using results from the full specification, and anything else that would be learned via sequential addition isn't true.

The rest of this paper proceeds as follows. In section 2, I discuss a large number of papers in which authors have added covariates sequentially for one or both of the reasons listed above. In section 3, I revisit Neal and Johnson's (1996) well-known study of the black-white wage gap, as

[1] This identity appears in Goldberger (1991), though it may have appeared elsewhere first.

[2] Code for my `b1x2` command is available for download from `http://gelbach.eller.arizona.edu/papers/b1x2/`.

well as Lang and Manove's (2006) interesting reconsideration. In section 4, I develop a conditional decomposition, using the sample omitted variables bias formula. Conditional on education, industry dummies, and occupation dummies, I find that heterogeneity in test scores explains roughly a third of the overall black-white wage gap in a sample similar to Neal and Johnson's (1996). Education explains less than 10 percent of the overall gap, variation in occupation of employment explains about 20 percent of the wage gap, and variation in industry explains virtually none.

In section 5, I briefly discuss three extensions. First, my decomposition nests the Oaxaca-Blinder decomposition (OBD). Second, my decomposition can be extended to instrumental variables estimators. Third, under the null hypothesis that X_2 covariates are orthogonal to X_1, one can use Hausman's (1978) well-known variance-of-differences result to test for statistically significant changes in the coefficient on X_1. The practical usefulness of this result is that when a single coefficient is of interest, one can assess the null with very little information: only the estimated coefficients and standard errors from the two specifications are needed. While my discussion of these issues is very brief, details can be found in Gelbach (2009).

Finally, I conclude in section 6.

2 Examples From the Literature

Consider the case of cross-group gaps in wages. In their outstanding review of work on race, gender and labor markets, Altonji and Blank (1999) report results from estimating a variety of models relating wages to characteristics of workers and their jobs.[3] Among their Table 4 results are three sets of estimates, Models 4 through 6, of wage regressions using data from the 1995 March CPS. My Table 1 reports their estimates of the coefficient on the black indicator and provides details about included covariates. Altonji and Blank write that

> As control variables are added to the model the negative effect of race ... on hourly wages becomes less significant. In 1995, black males received 21% lower hourly wages than white males if no control variables were included; they received 12% less once education, experience and region were controlled for, and they received 9% less when a full set of control variables were included. [Page 3156.]

[3]Unless noted otherwise, all references to wages in this paper are to the natural log of wages.

A common inference to draw from this discussion is that education, experience and region account $21\% - 12\% = 9\%$ of the black-white wage gap, while additional controls account for $12\% - 9\% = 3\%$. Indeed, the word "account" is sometimes used explicitly to characterize the change in a coefficient of interest when covariates are added. Consider Krueger's (1993) famous study relating the wage structure to computer use. Krueger writes that

> [b]etween October 1984 and 1989 the conventional OLS estimate of the return to education in the private sector increased by 0.96 points [from 7.92 to 8.88]. However, if computer use is held constant, the return to education is estimated to have increased by 0.56 points [from 7.06 to 7.62]. Thus, it appears that increased computer use can "account" for 41.6 percent ($= 100 \bullet (0.96 - 0.56/0.96)$) of the increase in the return to education in the private sector. [Pages 51–52.][4]

For another accounting example, consider Levitt and Syverson (2008), who use a sample of Chicago-area home sales to test whether agents exploit informational advantages over their principals. They estimate regressions of the log price of home sales on a dummy variable indicating whether the home sold is owned by a real estate agent. When the agent owns the home, the agent is also the principal. A principal-agent problem would cause sale prices of non-agent owned homes to be lower than those of agents' own homes. Thus, presence of a principal-agent problem implies a positive coefficient on the agent dummy.

My Table 2 reports some of Levitt and Syverson's results, from their Tables 1 and 2. When only city and year dummies are included, the coefficient on the agent dummy is 12.8 log points; I will consider this the base specification. The agent-dummy coefficient estimate falls to 4.8 when the authors include city-by-year dummies and basic characteristics related to the home's scale (e.g., number of bedrooms). Levitt and Syverson write that "the sales price difference between agent-owned and non-agent-owned homes is . . . almost two-thirds less than the within-city difference obtained before conditioning on scale." The authors then add more detailed house qualilty indicators, the listing agent's total number of sales, indicators of various keywords in the home's listing, and, finally, block dummies. As my Table 2 shows, most of the change in the coefficient

[4]DiNardo and Pischke (1997) have suggested that Krueger's estimates of the coefficient on the computer-use dummy itself are actually picking up occupational characteristics. Their argument is largely based on the fact that the computer-use dummy's coefficient moves substantially when they add occupation dummies to a base specification.

on the agent dummy between the first (base) and last (full) specification occurs when basic house characteristics are added. Levitt and Syverson conclude that

> the primary dimension along which agent-owned houses differ from other homes is in terms of scale and readily identifiable amenities.. . . Including a wide range of controls in addition to these basic ones . . . has a relatively small impact on the measured impact of agent-ownership. (Page 604)

However, it is possible that the scale and basic-amenity variables are correlated with detailed indicators of house quality, description keywords, and block dummies. Thus, it is possible that the coefficient on the agent dummy would move just as much when any of these other sets of covariates is added before adding the basic scale and amenity characteristics. Without a conditional accounting method, it is not possible to assess Levitt and Syverson's conclusion concerning the cause of the difference in unconditional and conditional agent-dummy effects.

In the examples above, authors draw inferences concerning the importance of different sets of covariates in affecting the magnitude of the coefficient of interest. Other cases involve only the general question of whether the coefficient on the variable of interest moves much at all. Consider the following passage from Lee and McCrary's (2005) excellent study, which uses regression discontinuity methods to evaluate discounting behavior by young potential criminal offenders:

> Moving from left to right in Table II, we control for an increasing number of factors. Column (1) gives our most parsimonious model, controlling only for a juvenile/adult dummy and a cubic polynomial in age at current arrest; column (8) gives our most complex model, adding controls for race, size of county in which the baseline arrest24 occurred, offense type of baseline arrest, and a quintic polynomial in age at baseline arrest. In each column, the added controls are good predictors of the probability of arrest, but in no case does including additional controls affect the estimated discontinuity importantly. (Page 17.)

The practice of sequentially adding covariates, whether for purposes of formal accounting or as a robustness check on a coefficient of interest, is now common throughout myriad subfields of economics. Further examples that I will not discuss in detail include the following:

- **Economics of crime:** Lee and McCrary (2005), Donohue and Levitt (2001).
- **Population economics and economics of the family:** Goldin and Katz (2002), Stevenson and Wolfers (2006), Caceres-Delpiano (2006).
- **Welfare Reform:** Haider, Schoeni, Bao and Danielson (2004).
- **Economics of Education:** Cullen, Jacob and Levitt (2005).
- **Economics of desegregation:** Kane, Riegg and Staiger (2006), Guryan (2004).
- **Wage gaps:** Hellerstein and Neumark (2008), Charles and Guryan (2008), Bound and Freeman (1992).
- **Health economics and tort reform:** Sharkey (2005), Klick and Stratmann (2007), Kessler and McClellan (1996).
- **Economic development:** Duflo (2001), Deaton (2008), Hanushek and Woessmann (2008).
- **Banking and financial economics:** Khwaja and Mian (2008).
- **Bankruptcy:** Lefgren and McIntyre (2009).
- **Economics of the minimum wage:** Card and Krueger (1995, e.g, Table 4.10).
- **Positive political economy:** Edlund and Pande (2002).
- **Economics of environmental regulation:** Greenstone (2002).
- **Human capital and seniority:** Moore, Newman and Turnbull (1998).

I emphasize that all of the work listed here is thoughtful, interesting, and well executed. My criticism is not of any particular paper, but rather the clearly accepted practice of sequential covariate addition itself.

3 Sequential Addition and the Black-White Wage Gap

I now turn to an empirical example involving the black-white wage gap, based on Lang and Manove's (2006) reconsideration of influential work by Neal and Johnson (1996, henceforth, NJ). NJ argue that educational attainment should be excluded from wage equations, since schooling decisions may be endogenous to labor market conditions. Neal and Johnson use data from the NLSY79 to evaluate the effects of both excluding schooling and controlling for a person's score on the Armed Forces Qualification Test (AFQT), a commonly used measure of ability. Because

NLSY subjects generally take the AFQT before they would have finished school, Neal and Johnson argue that this variable is an exogenous measure of ability based on pre-labor market information.

Neal and Johnson find a conventionally large unconditional black-white wage gap, which persists in large part when one includes years of schooling as a covariate. In Table 3, I report results from wage regressions using my own NLSY sample, which I constructed to roughly match NJ's.[5] In the first row of Panel A, I report the estimated coefficient on a race dummy from a different specification. All specifications include a dummy variable indicating whether a person is Hispanic, as well as a linear age term. The specification in the second column also includes years of schooling; the specification in the third column drops schooling but includes a measure of AFQT score; and the specification in the fourth column includes both of these variables.

The column (1) coefficient shows that, partialling out age and a Hispanic indicator, the average wage among NLSY young blacks with 1990 earnings is 22.1 log points below the average among young whites. Thus, $\widehat{\beta}_1^{\text{base}} = -22.1$. Because this estimate is essentially the same as the simple black-white difference in mean log wages when no age control is included and Hispanics are excluded from the sample (21.9 log points), I will also refer to $\widehat{\beta}_1^{\text{base}}$ as the simple black-white wage gap. Adding schooling, in column (2), reduces the black-white wage gap to 19.6 log points; using obvious notation, we thus have $\widehat{\beta}_{1,\text{race}}(X_2^{\text{educ}}) = -19.6$.

NJ's central result is that when schooling is excluded from the model, adding the AFQT test score as a covariate drives the coefficient on the race dummy much of the way toward zero. My column (3) coefficient estimate replicates this result. Adding AFQT to the base model in column (1) reduces the black-white wage gap from 22.1 log points to just 7.2 log points: $\widehat{\beta}_{1,\text{race}}(X_2^{\text{AFQT}}) = -7.2$. Based on this result, Neal and Johnson argue that most of the observed black-white wage gap is due to pre-market differences across race in ability, whatever the cause of these differences, rather than features of the labor market.

Lang and Manove (2006) construct a signalling model that implies that both educational attainment and test scores must appear as covariates in wage equations. In column (4) of Table 3, I report the estimated black-white wage gap from a specification that includes both AFQT and

[5] I discuss the sample in Appendix Agiven. While my specifications have 156 more observations than do NJ's, my results in specifications with given covariate sets are very similar to theirs. I have thus chosen to focus on the sample used here, rather than seeking to match the samples more closely.

educational attainment, like the one Lang and Manove estimate.[6] This specification, which I treat as the full specification for now, yields a black-white wage gap of 11.7 log points. This gap is only about half the size of the column (1) estimate, but it is 4.5 log points greater than NJ's preferred specification. For purposes of discussion in the present paper, it will be easiest to accept Lang and Manove's argument that both education and AFQT belong in the wage equation.[7] Thus, I will regard the column (4) model that includes both AFQT and education as the full specification, so $\widehat{\beta}_1^{\text{full}} = -11.7$.

The total difference between the base and full specifications is $\widehat{\delta} = \widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}} = -10.4$ log points. In other words, partialling out both education and AFQT explains away 10.4 log points of the simple column-(1) gap of 22.1 log points. In Panel B of Table 3, I turn to the issue of accounting for this 10.4-point explained. As discussed above, a common approach to such accounting is to add covariates sequentially and tally up the incremental changes in the black-white wage gap. The entry in the first row and column of Panel B shows that adding education to the base model reduces the black-white wage gap by 2.5 log points, which is the difference between the column (1) and column (2) coefficients. The entry in the second row shows that when we add AFQT to the column (2) specification, yielding column (4), the black-white wage gap falls by an additional 7.9 log points. A typical interpretation of these results would be that variation in education explains $2.5 \div 10.4 = 24\%$ of the explained part of the black-white wage gap, while variation in AFQT scores explains $7.9 \div 10.4 = 76\%$.

However, the education-AFQT sequence is not the only possible one for adding covariates. The second column in Panel B of Table 3 reports sequential-accounting results when we first add AFQT to the base specification and then add education. Comparing the estimates in columns (1) and (3) of Panel A now suggests that AFQT variation explains 14.9 log points of the black-white wage gap, for a total of $14.9 \div 10.4 = 143\%$ of the explained part of the black-white wage gap. Meanwhile, when we add education to the column (3) specification, the black-white wage gap increases in magnitude by 4.5 log points. In other words, variation in education now appears to

[6]LM's sample and use of weights differs somewhat from NJ's, but the pattern of estimates they report is similar to the one exhibited in Table 3.

[7]NJ's position is that AFQT is a better measure of skill than educational attainment. They argue that any correlation between wages and schooling that holds conditional on AFQT must as a matter of logic be due either to measurement error or endogeneity of post-AFQT administration schooling. Lang and Manove's signalling model provides an alternative, theory-based explanation for the need to include schooling in wage equations. Because they provide a variety of evidence in favor of their preferred approach, I will accept their position for purposes of this paper.

reduce the black-white wage gap, so holding education constant increases the gap.

How can this counter-intuitive result occur? As Lang and Manove comment and demonstrate in their Table 2, blacks in the sample have fewer years of schooling than whites, but the sign of this correlation is reversed after we partial out AFQT scores. In Table 4, I report results from auxiliary regressions of education on a black indicator. Both specifications include a Hispanic indicator and a linear age control. The specification in column (1) includes no other covariates, and it shows that by 1990, blacks had completed an average of 0.37 fewer years of schooling than whites. In column (2), I add AFQT to the covariate set in this auxiliary model. The results show that, conditional on AFQT, blacks get over a year more education than whites. As I discuss below, this reversal in the sign of the relationship between educational attainment and race is the key to understanding the sensitivity of the accounting results in Table 3 to the sequence in which education and AFQT are added to the base wage specification.

The results in this section should serve as a wake-up call. In a classic application, they point to the possibility that sequential addition of covariates can lead to very different conclusions about how much different covariates contribute to explaining the black-white wage gap. Results in Table 3 could be read to imply that the black-white differences in education either increase the black-white wage gap by 2.5 log points or reduce it by nearly twice as much. Meanwhile, black-white differences in test scores either reduce the wage gap by three-fourths of the total explained amount, or by more than 140%. Given that one believes both schooling and test scores belong in the wage equation, there is no clear basis on which to choose from these results, because the sequence of addition is arbitrary. Yet all of these results have potentially important policy and research consequences.

The results in this section throw into stark relief the sequence-sensitivity of the sequential addition-based approach to accounting for the change in a coefficient between the base and full specifications. I now turn to my own decomposition, which is conditional and therefore does not involve such problems.

4 Accounting for Coefficient Changes

I assume the two sets of right-hand side variables, X_1 and X_2, have k_1 and k_2 variables in each. The population linear relationship between Y and $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ is

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{1}$$

With no assumptions on the conditional expectation of ε given the regressors, we can think of (1) as simply specifiying a linear projection relationship: assuming all relevant expectations exist, we can define $\beta = (\beta_1', \beta_2')'$ as $\beta = E[X'X]^{-1}E[X'Y]$. Under the assumption that $E[\varepsilon_i|X_i] = 0$ for all observations i, β has a causal interpretation as the partial effect of X on Y, since then $E[Y_i|X_i = x] = x\beta$ and the partial derivative of the conditional expectation is β. In general I will assume that $E[\varepsilon|X] = 0$ does hold, though none of the results below require this assumption.[8]

As usual, the OLS estimator for the full vector of partial effects β is given by $\widehat{\beta} \equiv (X'X)^{-1}X'Y$. I label $\widehat{\beta}_1^{\text{full}}$ and $\widehat{\beta}_2$ the components of $\widehat{\beta}$ that correspond to the variables in X_1 and X_2, respectively. Standard results imply that $\widehat{\beta}_1^{\text{full}}$ is consistent for β_1 without assuming anything about either β_2 or the correlation between X_1 and X_2, since all X_2 variation is partialled out in the full specification. Now consider the coefficient on X_1 from the base specification that ignores X_2. The estimator for this coefficient is $\widehat{\beta}_1^{\text{base}} \equiv (X_1'X_1)^{-1}X_1'y$, whose probability limit is

$$\text{plim } \widehat{\beta}_1^{\text{base}} = \beta_1 + \text{ plim } \left(X_1'X_1\right)^{-1} X_1'X_2\beta_2, \tag{2}$$

$$\text{or } \beta_1^{\text{base}} = \beta_1 + \Gamma\beta_2 = \beta_1 + \delta. \tag{3}$$

Here, the parameter Γ is the matrix of coefficients from projecting the columns of X_2 on the columns of X_1:

$$X_2 = X_1\Gamma + W, \tag{4}$$

[8] See Wooldridge (2002) for a useful discussion of projections and partial effects.

where W is a matrix of conformable projection residuals. Equation (3) has two interpretations. In the first, δ is the population omitted variables bias from excluding X_2 when estimating β_1 via least squares: β_1^{base} is the parameter identified by this approach, and β_1^{base} differs from the β_1 by δ. As noted above, the second interpretation is that β_1^{base} is an interesting parameter in its own right. Under this interpretation, it is natural to consider decomposing δ into components related to different X_2 variables. Empirical decomposition exercises then involve breaking δ into meaningful components and estimating them.

In the example discussed in section 3, X_2 has two columns: one for education and one for AFQT. Thus, Γ also must have two columns. In this example, the projection relationship in (4) can also be written, in more detailed form, as

$$X_2^{\text{AFQT}} = \Gamma_0^{\text{AFQT}} + X_{1,\text{race}}\Gamma_{\text{race}}^{\text{AFQT}} + X_{1,\text{Hispanic}}\Gamma_{\text{Hispanic}}^{\text{AFQT}} + X_{1,\text{Age}}\Gamma_{\text{Age}}^{\text{AFQT}} + W^{\text{AFQT}}, \tag{5}$$

$$X_2^{\text{educ}} = \Gamma_0^{\text{AFQT}} + X_{1,\text{race}}\Gamma_{\text{race}}^{\text{educ}} + X_{1,\text{Hispanic}}\Gamma_{\text{Hispanic}}^{\text{educ}} + X_{1,\text{Age}}\Gamma_{\text{Age}}^{\text{educ}} + W^{\text{educ}}. \tag{6}$$

In this example, X_1 contains a constant, a race indicator, an Hispanic indicator, and age, as in Table 3 above. The row of Γ^{AFQT} corresponding to the race dummy tells us the black-white mean difference in residual test scores after partialling out the other elements of X_1; the race dummy's row in Γ^{educ} tells us the same thing for years of educational attainment.

To see how the elements of (5) affect the difference between β_1^{base} and β_1^{full}, I rewrite (3) as follows:

$$\beta_1^{\text{base}} - \beta_1 = \Gamma^{\text{AFQT}}\beta_2^{\text{AFQT}} + \Gamma^{\text{educ}}\beta_2^{\text{educ}}, \tag{7}$$

$$\text{so } \delta_{\text{race}} \equiv \beta_{1,\text{race}}^{\text{base}} - \beta_{1,\text{race}} = \Gamma_{\text{race}}^{\text{AFQT}}\beta_2^{\text{AFQT}} + \Gamma_{\text{race}}^{\text{educ}}\beta_2^{\text{educ}}. \tag{8}$$

We can see from (7) how to decompose the explained part of δ_{race}, the simple black-white wage gap:

$$\text{AFQT component}: \qquad \delta_{\text{race}}^{\text{AFQT}} = \Gamma_{\text{race}}^{\text{AFQT}} \beta_2^{\text{AFQT}}. \tag{9}$$

$$\text{Education component}: \qquad \delta_{\text{race}}^{\text{educ}} = \Gamma_{\text{race}}^{\text{educ}} \beta_2^{\text{educ}}. \tag{10}$$

While I will generalize this claim below, one of my primary contributions here is to point out that, given the model (1), the parameters in (9) provide the best answer to the question: "How much of the black-white wage gap is explained by variation in education and AFQT?" These parameters are clearly interpretable as the mean black-white gap in AFQT or education, scaled by each covariate's wage-equation impact. These covariate mean differences and wage-equation effects are population parameters that do not depend on the order in which covariates are partialled out. Thus, consistent estimates of the δ_{race} parameters do not suffer from the sequence sensitivity observed in section 3. Moreover, the δ_{race} parameters are intuitively exactly what researchers must mean when they attempt to decompose the causes of the black-white wage gap. If there were no mean difference in a covariate across race, then its Γ_{race} coefficient would be zero. In this case, variation in the covariate would explain none of the black-white gap in mean wages. The same would hold if the covariate does not affect the wage, so that $\beta_2 = 0$.

In the first row of Panel A of Table 5, I report estimates of Γ_{race} from my NLSY sample. The estimate in the column headed education shows that after we partial out variation due to Hispanic background and age, blacks have 0.37 fewer years of education (this result also appears in Table 4 above). Thus, $\widehat{\Gamma}_{\text{race}}^{\text{educ}} = -0.37$. The next row of Table 5 reports the estimated coefficients on the education variable from OLS estimation of the full specification given in (1). An additional year of education is associated with an increase in average wages of 4.1 log points, conditional on race, Hispanic background, age, and AFQT. Thus, $\widehat{\beta}_2^{\text{educ}} = 0.041$. Equation (9) tells us that the conditional contribution of education is the product of these two estimates: $\widehat{\delta}_{\text{race}}^{\text{educ}} = -0.37 \times 0.041$. The next row of the table shows that this product equals -0.0155: conditional on all the other variables in the model, the black-white wage gap would be roughly 1.6 log points smaller if mean years of schooling were the same across blacks and whites.

I provide analogous results for AFQT in the second column of Table 5. Because I have standardized AFQT scores within my sample (see Appendix A), the coefficient in the first row shows that the average AFQT score among blacks is 0.87 standard deviations lower than among whites, conditional on Hispanic background and age. Thus, $\widehat{\Gamma}_{\text{race}}^{\text{AFQT}} = -0.87$. The next row shows that a one standard deviation increase in AFQT is associated with a wage increase of roughly 10 log points: $\widehat{\beta}_2^{\text{AFQT}} = 0.1013$. The third row reports the product of these two estimates, $\widehat{\delta}_{\text{race}}^{\text{AFQT}} = -0.088$. This estimate tells us that if mean AFQT were the same in this sample and all other relationships were held constant, the black-white wage gap would be 8.8 log points smaller.

Adding together our estimates of these decomposition components shows that variation in education and AFQT account for a total of 10.4 log points of the black-white wage gap. This figure exactly equals the difference between the column (1) and column (4) specifications of Table 3. As I discuss below, this equality is an identity: the sample omitted variables bias formula and the difference between the base- and full-specification coefficient estimates are necessarily the same.

First, though, it will be helpful to re-visit the issue of sequential covariate addition. For comparison's sake, Panel B of Table 5 shows the two sets of accounting estimates reported above in Table 3. The first set of accounting estimates comes from first adding education to the column (1) base specification, to arrive at column (2) Table 3, and then adding AFQT to that specification to arrive at column (4). Since the columns (1) and (2) estimates of the black-white wage gap are 22.07 and 19.58 log points, this sequence suggests that the black-white wage gap would be 2.49 log points lower if education were held constant, as reported in the first row and column of Panel B of Table 5. Because the full-specification estimate of the black-white wage gap is 11.71 log points, the difference between columns (2) and (4) is 7.86 log points, as reported in Table 3 and repeated in the first row and second column of Panel B of Table 5.

The $\widehat{\delta}_{\text{race}}$ parameters are consistent estimates of the conditional contribution of each covariates. Thus, the bias from using the education-AFQT sequential addition approach equals the difference between the estimates in the first row of Panel B and the corresponding $\widehat{\delta}_{\text{race}}$ estimates. The second row in Panel B shows the resulting bias estimates. These bias estimates sum exactly to zero, because my conditional decomposition and the sequential addition decomposition necessarily yield the same total explained gap (again, I discuss this equality below).

The bias from using the education-AFQT sequence is relatively small in each case–a bit less than one log point. However, the next two rows in Panel B show that the reverse sequence causes substantial bias in accounting for the explained part of the gap. While my conditional decomposition shows that variation in educational attainment contributes about 1.5 log points to the black-white wage gap, the AFQT-education sequential-addition approach suggests that variation in years of schooling actually reduces the black-white wage gap by 4.5 log points. Moreover, the same bias magnitude applies to AFQT with opposite sign. The AFQT-education sequence of covariate addition would therefore suggest that nearly 15 log points of the black-white wage gap can be attributed to AFQT variation, rather than the consistent estimate of about 9 log points computed when we hold all other variables constant.

What drives the large bias from using the AFQT-education sequential addition approach? Consider the difference between the estimate of -0.0718 in column (3) of Table 3, and column (4)'s full-specification estimate of -0.1171. The difference between these estimates, +0.0453, is the explained-gap component attributed to education using the AFQT-education sequential addition approach. To understand the relationship between this difference and the education component of my conditional decomposition, it will help to consider the following projection relationship:

$$\begin{aligned} X_2^{\text{educ}} = {} & X_{1,\text{race}}\Gamma_{\text{race}}^{\text{educ(AFQT)}} + X_{1,\text{Hispanic}}\Gamma_{\text{Hispanic}}^{\text{educ(AFQT)}} + X_{1,\text{Age}}\Gamma_{\text{Age}}^{\text{educ(AFQT)}} \\ & + X_2^{\text{AFQT}}\Gamma_{\text{AFQT}}^{\text{educ(AFQT)}} + W^{\text{educ(AFQT)}}. \end{aligned} \tag{11}$$

The "educ(AFQT)" superscript notation indicates that the Γ parameters come from an equation relating education to all the X_1 variables, as well as AFQT. Inclusion of AFQT on the right hand side of (11) is the only difference between this projection relationship and the one in (5).

Let $\widehat{\beta}_{1,\text{race}}(AFQT)$ be the race-dummy estimate from column (3) of Table 3, the one that includes AFQT. The equivalence result below can be used to show that

$$\widehat{\delta}_{\text{race}}(AFQT) \equiv \widehat{\beta}_{1,\text{race}}(AFQT) - \widehat{\beta}_{1,\text{race}}^{\text{full}} = \widehat{\Gamma}_{\text{race}}^{\text{educ(AFQT)}}\widehat{\beta}_2^{\text{educ}}.$$

The estimate $\widehat{\delta}_{\text{race}}(AFQT)$ is the education component of the explained gap computed using the education-AFQT sequential addition approach, which we have seen equals +0.0453. We can thus find $\widehat{\Gamma}_{\text{race}}^{\text{educ(AFQT)}}$ by dividing $\widehat{\delta}_{\text{race}}(AFQT)$ by the Table 5 education coefficient in the full-specification wage equation, $\widehat{\beta}_2^{\text{educ}} = 0.041$. The result is $\widehat{\Gamma}_{\text{race}}^{\text{educ(AFQT)}} = 1.10$, which I earlier reported in the second column of Table 4. Relative to my conditional decomposition, the bias in $\widehat{\delta}_{\text{race}}(AFQT)$ is thus

$$\begin{aligned}\text{Bias}^{\text{educ(AFQT)}} &= \widehat{\delta}_{\text{race}}(AFQT) - \widehat{\delta}_{\text{race}} \\ &= \left[\widehat{\Gamma}^{\text{educ(AFQT)}}_{\text{race}} - \widehat{\Gamma}^{\text{educ}}_{\text{race}}\right] \times \widehat{\beta}^{\text{educ}}_{2}. \end{aligned} \tag{12}$$

Equation (12) shows that the estimated bias in the part of the explained gap attributed to a covariate depends on two factors. One factor is the difference made by conditioning on AFQT when estimating the black-white gap in education, i.e., the difference between $\widehat{\Gamma}^{\text{educ(AFQT)}}_{\text{race}}$ and $\widehat{\Gamma}^{\text{educ}}_{\text{race}}$. In the present example, these projection coefficients are very dis-similar, having different signs and very different magnitudes. Their difference equals $1.10-(-0.37)=1.47$. The other factor that affects the bias from using sequential covariate addition to estimate the education share of the explained gap is $\widehat{\beta}^{\text{educ}}_{2}$, the estimated return to education. While this coefficient is relatively small, its product with 1.47 is substantial, at +6 log points.

Since the biases in the components attributed to AFQT and education must sum to zero, it immediately follows that the sequential-addition bias in the component attributed to AFQT when it is added first to the base specification satisfies $\text{Bias}^{\text{AFQT(AFQT)}} = -\text{Bias}^{\text{educ(AFQT)}}$. By symmetry of all definitions above, it can then be shown that the bias in the component attributed to education when it is added first to the base specification satisfies

$$\begin{aligned}\text{Bias}^{\text{educ(educ)}} &= \widehat{\delta}_{\text{race}}(X^{\text{educ}}_{2}) - \widehat{\delta}_{\text{race}} \\ &= \left[\widehat{\Gamma}^{\text{AFQT(educ)}}_{\text{race}} - \widehat{\Gamma}^{\text{AFQT}}_{\text{race}}\right] \times \widehat{\beta}^{\text{AFQT}}_{2}, \end{aligned} \tag{13}$$

where $\widehat{\Gamma}^{\text{AFQT(educ)}}_{\text{race}}$ is defined analogously to $\widehat{\Gamma}^{\text{educ(AFQT)}}_{\text{race}}$: it is the race-dummy coefficient from least-squares estimation of the projection relationship between the left hand side variable AFQT and all X_1 variables plus education. This coefficient's value is -0.78, quite close to the estimate of -0.87 for $\widehat{\Gamma}^{\text{AFQT}}_{\text{race}}$. That is, the black-white AFQT gap changes by less than a tenth of a standard deviation when we partial out education in computing this gap. Meanwhie, the return to AFQT from the full-specification wage equation, $\widehat{\beta}_2^{\text{AFQT}}$, is roughly 10 log points per standard deviation. That is why the bias of the education component's part of the sequential-addition explained gap is less than one log point when we add education first.

We can summarize this discussion by supposing we have re-scaled all X_2 variables so that all wage-equation coefficients equal one. Whether sequential covariate addition accounting exercises are sensitive to the sequence of addition then depends entirely on the pattern of differences between $\widehat{\Gamma}$ projection coefficients estimated with and without conditioning on other X_2 covariates. When the relationship between X_1 and one of the covariates is very sensitive to conditioning on additional covariates, the sequence in which covariates are added will make a large difference in accounting exercises, as in the black-white wage gap.

This example clearly illustrates two key propositions. First, as simple as it is, my conditional decomposition has an intuitive justification as the correct way to decompose cross-group mean differences. Second, using sequential covariate addition can lead to seriously biased estimates of the contributions of different variables. I now discuss general formulas for the decomposition with arbitrary numbers of covariates.

4.1 The General Decomposition

In this section, I provide general notation for my conditional decomposition. This discussion is largely a formalization of the basic idea in the previous section, that the sample analogue of the omitted variables bias formula can serve as the basis for explaining sensitivity of the estimated coefficient on a vector of base-specification regressors.

Recall that the parameter δ from (3) represents the difference between the population coefficients on X_1 in the base and full specifications. Since plim $\widehat{\beta}_1^{\text{base}} = \beta_1 + \delta$, and since $\widehat{\beta}_1^{\text{full}}$ is consistent for β_1, a natural approach to estimating δ is to just use $\widehat{\delta} \equiv \widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}}$, which is what authors do when they compare base and full specifications. The key algebraic identity driving my decomposition is that

$$\widehat{\delta} = \left(X_1'X_1\right)^{-1} X_1'X_2\widehat{\beta}_2, \tag{14}$$

so that the base-full difference in the X_1 coefficient vector identically equals the sample analogue of the omitted variables bias formula. For a very simple proof,[9] observe that the usual least-squares algebra implies

$$y = X_1\widehat{\beta}_1^{\text{full}} + X_2\widehat{\beta}_2 + \widehat{\varepsilon},$$

where $\widehat{\varepsilon} = y - X\widehat{\beta}$ is the usual vector of fitted residuals. Now pre-multiply both sides by $\left(X_1'X_1\right)^{-1}X_1'$, delivering

$$\widehat{\beta}_1^{\text{base}} = \widehat{\beta}_1^{\text{full}} + \left(X_1'X_1\right)^{-1}X_1'X_2\widehat{\beta}_2,$$

since the fitted residuals are exactly orthogonal to the columns of X_1 by construction of the least-squares estimator. Thus

$$\begin{aligned}\widehat{\delta} &\equiv \widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}} \\ &= \left(X_1'X_1\right)^{-1}X_1'X_2\widehat{\beta}_2.\end{aligned}$$

This identity is the key to conditionally decomposing the difference in base and full regression coefficients. Since $\widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}}$ identically equals the sample omitted variables bias formula, any decomposition of $\left(X_1'X_1\right)^{-1}X_1'X_2\widehat{\beta}_2$ is also a decomposition of $\widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}}$. Let

- X_{2k} be the column of observations on the k^{th} covariate in X_2;
- $\widehat{\Gamma}_k = \left(X_1'X_1\right)^{-1}X_1'X_{2k}$ be the estimated OLS coefficient on X_1 from an auxiliary model with X_{2k} as the dependent variable;

[9] I am grateful to Ingmar Prucha for suggesting this proof. A similar proof using different notation is available in Goldberger (1991, p. 184). A slightly more detailed alternative proof that makes use of results on orthogonal projections is also available. First, express $\widehat{\beta}_2$ as $(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}Y$, where $X_2^* \equiv M_1X_2$ and $M_1 = I - P_1$, with $P_1 = X_1(X_1'X_1)^{-1}X_1$ being the projection matrix for X_1. Thus X_2^* is the part of X_2 that is orthogonal to X_1. By the Frisch-Waugh-Lovell theorem, $P = P_1 + P_2^*$, where P is the projection matrix using all variables in $X \equiv \begin{bmatrix} X_1 & X_2 \end{bmatrix}$, P_1 is the projection matrix for X_1 only, and P_2^* is the projection matrix for X_2^*. Post-multiplying both sides of the projection identity by Y yields $X_1\widehat{\beta}_1^{\text{full}} + X_2\widehat{\beta}_2 = X_1\widehat{\beta}_1^{\text{base}} + X_2^*\widehat{\beta}_2$. Rearranging implies $X_1\widehat{\delta} = (X_2 - X_2^*)\widehat{\beta}_2 = P_1X_2\widehat{\beta}_2$. Observing that $X_1'P_1 = X_1'$ and pre-multiplying both sides by $(X_1'X_1)^{-1}X_1'$ then yields $\widehat{\delta} = (X_1'X_1)^{-1}X_1'X_2\widehat{\beta}_2 = \widehat{\Gamma}\widehat{\beta}_2$, as in the theorem.

- $\widehat{\beta}_{2k}$ be the estimated coefficient on X_{2k} in the full specification of interest.

Then the part of the sample omitted variables bias attributable to covariate k can be written as $(X_1'X_1)^{-1}X_1'X_{2k}\widehat{\beta}_{2k} = \widehat{\Gamma}_k\widehat{\beta}_{2k}$. Because the sample omitted variables bias formula is linear in its k_2 components, it follows that

$$\begin{aligned}\widehat{\delta} &= \sum_{k=1}^{k_2} \widehat{\Gamma}_k\widehat{\beta}_{2k} \\ &= \sum_{k=1}^{k_2} \widehat{\delta}_k, \text{ where } \widehat{\delta}_k \equiv \widehat{\Gamma}_k\widehat{\beta}_{2k}.\end{aligned} \tag{15}$$

Equation (15) suggests a straightforward way of carrying out a decomposition:

1. Estimate the full model of interest, which yields the estimated vector $\widehat{\beta}_2$.
2. Use OLS to estimate the vector of coefficients on X_1 in a set of auxiliary models with each of the k_2 covariates X_{2k} acting as the dependent variable; this estimate is $\widehat{\Gamma}_k$.
3. Multiply $\widehat{\Gamma}_k$ by $\widehat{\beta}_{2k}$, which yields $\widehat{\delta}_k$, the component of the omitted variables bias estimated to be due to each variable k.

Because $\widehat{\delta} = \sum_k \widehat{\delta}_k$, the estimates of $\widehat{\delta}_k$ for groups of covariates can be summed together, e.g., into industry, occupation, and so on. Each sum then yields group-specific omitted variables bias components as desired. For group g, we have $\widehat{\delta}^g = \sum_{k \text{ in group } g} \widehat{\delta}_k$. The overall difference in coefficients, $\widehat{\delta}$, can thus also be written as the sum of group-wise components: $\widehat{\delta} = \sum_g \widehat{\delta}^g$. A practical drawback to this approach is that there will often be many covariates, so that k_2 is quite large. In such cases, a simple alternative approach is available: rather than estimating k_2 regressions and then creating G group-specific sums, one can sum first and regress later. Define the heterogeneity variable $\widehat{H}_{k(i)} = X_{2k(i)}\widehat{\beta}_{2k}$. For any observation i, $\widehat{H}_{k(i)}$ tells us how much of i's wage is estimated to be due to i's value of the k^{th} covariate, X_{2k}. The practical decomposition I suggest here is based on the obvious equality

$$\left(X_1'X_1\right)^{-1}X_1'\widehat{H}_k = \left(X_1'X_1\right)^{-1}X_1'X_{2k}\widehat{\beta}_{2k} = \widehat{\Gamma}_k\widehat{\beta}_{2k} = \widehat{\delta}_k.$$

Summing over all covariates in group g and using the fact that $(X_1'X_1)^{-1}X_1'$ does not vary with the covariate index k, we have

$$\begin{aligned}\widehat{\delta}^g &= \sum_{k \in \text{group } g} \widehat{\delta}_k \\ &= (X_1'X_1)^{-1} X_1' \sum_{k \in \text{group } g} \widehat{H}_k.\end{aligned}$$

This small manipulation shows that we can calculate the total group-g contribution to $\widehat{\delta}$ as follows:

1. Estimate the full model of interest.
2. For each observation i, Create $\widehat{H}^g_{(i)}$ by summing up $X_{2k(i)}\widehat{\beta}_{2k}$ over all covariates k in group g; call the resulting group-level heterogeneity term $\widehat{H}^g_{(i)} = \sum \widehat{H}_{k(i)}$.
3. Use OLS to estimate the coefficients on X_1 in an auxiliary model in which the dependent variable for observation i is the group-level heterogeneity term $\widehat{H}^g_{(i)}$.

The coefficient estimate resulting from the final step above exactly equals $\widehat{\delta}^g$, the group-g part of the sample omitted variables bias formula. Aside from computational convenience,[10] this interpretation of the decomposition has a nice intuitive flavor. In the wage gap example, $H_{(i)}$ is the component of i's wage that is explained by i's value of the vector of covariates, $X_{2(i)}$. Overall differences in the simple and conditional wage gaps can then be seen to be due to black-white differences in the average of the heterogeneity term $H_{(i)}$. When X_1 contains no variables other than a race dummy, the part of this difference attributable to variation in group-g covariates is the simple cross-group difference in the mean of $H^g_{(i)}$, the wage heterogeneity due to variables in group-g. When X_1 also includes other variables, as in the NLSY example I consider in this paper, the relevant parameter is the conditional cross-group difference in $H^g_{(i)}$, after partialling out these other variables.

[10]Notice that we need estimate only G auxiliary equations, rather than k_2. In the NLSY example above, this distinction was irrelevant, because there were two covariates with each being the only member of its grouop. Thus, $k_2 = G = 2$ in that example. However, there are many cases in which researchers include large numbers of covariates, especially large sets of dummy variables, in the X_2 covariate set. I will enlarge the covariate set below to illustrate this point.

To conduct inference, we must be able to estimate the asymptotic distribution of $\sqrt{n}(\widehat{\delta}^1 - \delta^1, \widehat{\delta}^2 - \delta^2, \ldots, \widehat{\delta}^G - \delta^G)$. In general, all estimators involved in my decomposition are asymptotically normal. Since the decomposition involves continuously differentiable functions of these estimators, joint asymptotic normality of the decomposition components follows from the delta method. Accounting for the variance of all estimators is straightforward, but the derivation of the proper formula is lengthy, so I relegate the full derivation of the asymptotic variance matrix for $\widehat{\delta}$ to Appendix B.[11]

To illustrate the usefulness of the grouping approach just discussed, I now observe that the NLSY has information on additional potentially important explanatory variables. In particular, I add 12 occupation dummies and 12 industry dummies to the covariate set, so that X_2 now includes 26 variables. I continue to treat the years of education and AFQT variables as their own groups, while adding another group each for occupation and industry. Thus $G = 4$ with this enlarged covariate set.

Table 6 reports results from the full specification that includes a constant, race and Hispanic-background dummies, a linear age control, and all 26 X_2 covariates. The black-white wage gap in this full specification is 9.26 log points, so that a total of $\widehat{\delta}_{\text{race}} = 12.85$ log points of the base specification's 22.11 log-point gap is explained by the 26 X_2 covariates. The estimated standard error for $\widehat{\delta}_{\text{race}}$ is 2.26 log points.

The bottom part of Table 6 reports my detailed decomposition of the part of the black-white wage gap that can be explained by each of four covariate sets, conditional on all of them simultaneously. Variation in AFQT explains 7.06 log points, which is more than half of the explained part of the gap, and nearly a third of the total gap. On the other hand, this component of the black-white wage gap is less than half of the amount that Neal and Johnson's specification would imply. Variation in educational attainment explains another 1.56 log points. This component does not explain a large share of the gap, though it is precisely estimated. Variation in occupation, as measured by

[11] I note, though, that a key requirement for the auxiliary-model approach to work is that at least one element of Γ^g or β_2^g be nonzero. If this condition does not hold, then the variance of $\sqrt{n}(\widehat{\delta}^g - \delta^g)$ converges to zero, and estimated test statistics will have non-standard behavior. For any covariate set such that the researcher allows the possibility that X_1 and X_2^g are uncorrelated, $\Gamma^g = 0$ is possible. To rule out the zero-variance case, it is then important to verify that not all elements of β_2^g are zero. This can be done by estimating the full model and testing the joint null $H_0 : \beta_2^g = 0$ before undertaking the decomposition. If this null cannot be rejected, then X_2^g should be excluded from the full model anyway, since these variables have no systematic partial association with Y.

the 12 occupation dummies, explains 4.34 log points of the black-white wage gap. This represents roughly a third of the part of the gap that is explained by the 26 covariates, and about 20 percent of the overall black-white wage gap.

Variation in industry plays essentially zero role in explaining the black-white wage gap. The small and imprecisely estimated industry effect suggests that industry variation has no effect on the black-white wage gap. This result is perhpas surprising, because both wages and race vary systematically across industry. For example, the range of the estimated industry coefficients in the full specification is over 40 log points. Moreover, a test of the null hypothesis that the vector of 12 industry dummies all have coefficients equal to zero results in a p-value of zero up to four decimal places. Similarly, results from a multivariate regression of the 12 industry dummies on race and Hispanic background dummies as well as age yield a p-value of 0.0062 against the null hypothesis that race is unrelated to industry. In sum:

- wages vary systematically across industry, even after partialling out all other variables;
- industry of employment varies systematically across race, conditional on Hispanic background and age;
- yet, the black-white wage gap is unaffected by controlling for industry of employment.

This seemingly anomalous result has a simple explanation: the correlation between race and industry is orthogonal to the conditional correlation between wages and industry. That is, even though $\Gamma^{\text{industry}} \neq 0$ and $\beta_2^{\text{industry}} \neq 0$, it is nonetheless true that $\Gamma^{\text{industry}}\beta_2^{\text{industry}} = 0$. As a consequence, wage heterogeneity due to industry variation, H^{industry}, has the same mean among blacks as whites, conditional on Hispanic background and age. This example points to the importance of accounting directly for the effects of X_2 covariates: the simple facts that X_1 and X_2 are correlated, and that X_2 enters the model of interest importantly, are necessary but not sufficient conditions for the coefficient on X_1 to be affected by partialling out X_2.

4.2 Discussion

My decomposition has several key properties. First, it exactly accounts for the difference $\widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}}$. Second, the group-specific components $\widehat{\delta}^g$ are consistent for their population analogues δ^g.

As a result, elements of my sample decomposition can be regarded as consistently estimating the analogous components of the difference in population parameters β_1^{base} and β_1. Third, my decomposition is conditional, rather than sequential.

One lesson from my results is that, in the future, empirical researchers who use linear estimators and are interested in robustness and accounting exercises should proceed by taking the following steps.

1. Determine the base specification of interest. In some situations, this specification will include a single variable of interest, as when the goal is to understand what explains the unconditional difference in means across race, sex, and so on. In other cases, researchers might believe the relevant base specification should include additional controls. This is the approach I have taken here, since both Neal and Johnson (1996) and Lang and Manove (2006) include Hispanic background and age in all specifications. It is easy to think up other such cases. For example, changes in state policies are rarely assumed to be unconditionally exogenous. In repeated cross-section studies, researchers therefore sometimes include state and year dummies in all specifications. These dummies naturally are part of X_1, using my notation.
2. Determine the entire list of control variables of interest, which together make up X_2.
3. Determine how to collect the different X_2 variables into G groups.
4. Use my decomposition to estimate how much of the base-full coefficient difference is due to each of the G groups of covariates.

Only the base and full specifications need be estimated to carry out this procedure. As such, the order in which covariates should be added plays no role in my decomposition and so cannot cause problems. In fact, there is no reason to add covariates sequentially at all. To the extent that I do so in this paper, the purpose is to illustrate the decomposition method I propose.

A key conclusion of my argument is that there is no justification for intermediate specifications. It is obviously important, then, that researchers have the "right" full specification in mind, in the sense that the β_1 parameter of interest in (1) can reasonably be interpreted causally. The plausibility of this assumption is arguable in some cases. But that issue has little to do with the appropriateness of the decomposition method I propose here. If researchers aren't willing to make assumptions sufficient to make β_1 interesting, then they have bigger fish to fry than understanding the differences between two problematic specifications. For this reason, I assume in this paper that sufficient data are available to estimate a full specification that is uncontroversial.

There are three further points to discuss concerning the full specification. The first concerns sensitivity analysis of the form advocated by Leamer (1985). Because I take the full specification as known, my purpose here is not to develop or even to advocate methods for selecting the "right" specification. Thus, methods like Leamer's extreme-bounds analysis concern a different, prior question. My focus here is on why covariates in X_2 matter in comparing base and full specifications, not on which of the 2^{k_2} specifications involving these covariates is best in some particular metric.

A second point concerns potentially endogenous, mismeasured, or otherwise problematic covariates in X_2. If a researcher believes that some X_2 is not exogenous, then either she must find a valid instrument for X_2 or focus on a reduced form specification. Nothing in my results does or is intended to address such problematic variables. Again, though, if researchers are wary of their chosen full specification, the first-order problem is to address any problematic X_2 variables, not worry about differences between $\widehat{\beta}_1^{\text{base}}$ and a suspect estimate of β_1.

The third point concerns a situation in which part of the effect of X_1 operates through covariates in X_2. For example, Ayres (2005) considers the case of racial differences in organ transplantation practices, emphasizing the distinction between racially disparate impacts and treatment. To illustrate, suppose hospitals required would-be transplant recipients to have high school diplomas. Given that the dropout rate is higher among blacks than whites, such a policy would lead to racial differences in the unconditional rate of transplant receipt. Consider regressing, say, race-specific state-by-year transplant rates on a race dummy in the base specification that includes no other covariates would then be nonzero.

If we suppose the diploma requirement is the only factor that causes black-white differences in transplantation, then controlling for a dummy indicating dropout status would drive the race dummy's coefficient to zero. Yet it would be wrong to conclude that the hospital's policy does not affect blacks differently from whites: by requiring a high school diploma, the hospital creates a situation in which blacks are less likely to receive transplants. Whether the hospital's policy is justifiable is an interesting quesiton, but it is also irrelevant to the issue of disparate impact. This example again shows the importance of having a clear understanding of the relevant empirical question before going about robustness or accounting exercises.

5 Extensions

In this section, I briefly consider three extensions of the analytical results above. Details for all three extensions can be found in Gelbach (2009).

5.1 Instrumental Variables Estimators

My first extension involves instrumental variables (IV) models. It can be shown that when β_1 from equation (1) is exactly identified, the equivalence between the sample omitted variables bias formula and $\widehat{\beta}_1^{\text{base}}$ and $\widehat{\beta}_1^{\text{full}}$ holds when all coefficients are estimated using two-stage least squares (2SLS). When β_1 is over-identified, the equivalence does not hold in fixed samples, but it does hold asymptotically (provided all instruments are valid). Thus, differences in IV estimates of β_1 due to addition of exogenous covariates X_2 can be meaningfully decomposed, as with OLS estimates.[12]

In Gelbach (2009), I use the same data used by Gelbach (2002b) to study the effects of public schooling on maternal labor supply. Gelbach (2002b) uses quarter of birth dummies to instrument for five-year-old's enrollment in public school. He finds that public school enrollment is associated with sizable increases in labor supply among single mothers whose youngest child is five. In Gelbach (2009), I use the IV version of the conditional decomposition discussed above to assess how sensitive Gelbach's original results are to controlling for covariates. I find that (a) controlling for covariates does not systematically affect the public school enrollment impact, even though (b) different groups of covariates do affect this impact. This is possible because groups of covariates have approximately equal, opposite-signed bias effects, up to sampling variation. This is an interesting finding in light of recent work by Buckles and Hungerman (2008), who find that IV estimates of the returns to schooling increase by 20 to 50 percent when they add family background controls to the model, though they do not test for statistical significance. Because my asymptotic variance formulas can be adopted to cover the IV case, the present paper will allow one to answer whether covariates importantly affect IV estimates in applications like estimating the returns to schooling.

[12]While for often adjust the instrument sets across specifications, I take the instrument set as fixed for purposes of this paper.

5.2 The Oaxaca-Blinder Decomposition

The second extension concerns the Oaxaca-Blinder decomposition (OBD). Suppose we are interested in decomposing the simple mean difference in black and white wages. Let D be a race dummy, and let X be the full set of other explanatory variables, not including a constant. Finally, let $D \times X$ be the interaction of these variables. It is straightforward to show that the explained and unexplained components of the OBD, also known as quantities and prices, can be found using the following procedure:

1. Use least-squares to estimate the coefficients on a constant, the race dummy, the explanatory variables, and their interaction, yielding $\widehat{\beta}_0$, $\widehat{\beta}_d$, $\widehat{\beta}_x$, and $\widehat{\beta}_{dx}$.

2. For each observation i, construct the estimated heterogeneity term $\widehat{H}^X_{(i)} \equiv X_{(i)}\widehat{\beta}_x$. The explained component of the OBD equals the cross-group difference in mean values of $\widehat{H}^X$. This is true because regressing $\widehat{H}^X$ on a constant and D yields $\Delta\overline{X}\widehat{\beta}_x$, which is the explained component valued using the estimated wage-equation coefficients for the group with $D = 0$.

3. Construct the heterogeneity estimate $\widehat{H}^{DX}_{(i)} \equiv \widehat{\beta}_d + \sum_k X_{k(i)}\widehat{\beta}_{k,dx}$, where k indexes the variables in X. The unexplained component of the OBD equals the cross-group difference in mean values of $\widehat{H}^{DX}$. This is true because regressing $\widehat{H}^{DX}$ on a constant and D yields $\widehat{\beta}_d + \overline{X}_1 \times \widehat{\beta}_{dx}$, where $\overline{X}_1$ is the vector of explanatory-variable means for the group with $D = 1$.

This procedure works because (i) the simple difference in mean wages equals the coefficient on a race dummy from the projection of individual wages on a constant and the race dummy, while (ii) the fully interacted model estimated in the procedure's first step yields the estimated coefficients from race-specific wage equations. Together, $\widehat{\beta}_0$ and $\widehat{\beta}_x$ are the wage-equation coefficient estimates for the group with $D = 0$. Adding those coefficients to $\widehat{\beta}_d$ and $\widehat{\beta}_{dx}$, respectively, then yields the coefficient estimates for the group with $D = 1$. The second two steps in the procedure above are just the auxiliary regression steps used in my groupwise decomposition, with the groups of covariates being X and DX. The only twist is the addition of $\widehat{\beta}_d$ to the heterogeneity term in the third step.[13]

[13]The OBD developed above uses the estimated coefficients for the group with $D = 0$ to measure the explained part of the wage gap. It is well known that the OBD is not unique. To measure the explained part of the wage gap with the other group's coefficient estimates, one just replaces D with $1 - D$ everywhere above. Other decompositions can also be defined; discussions of this issue appear in Neumark (1988) and Haider, Elder and Goddeeris (2009), for example.

This relationship between the OBD and my decomposition may be practically useful. First, my variance formulas can be used to conduct inference on OBD estimates, which is very rarely done.[14] Second, my auxiliary regression method can be used to compute detailed as well as overall Oaxaca-Blinder decompositions.Given the popularity of the OBD, and the interest in questions like, "Does industry or occupation explain more of the wage gap in an OBD?", my variance formulas will be practically very useful for applied practice.[15]

5.3 Using the Hausman Test

A third extension involves testing for statistically significant changes in the coefficient on X_1 under the null hypothesis that X_2 covariates are orthogonal to X_1. It is of course possible to test for orthogonality directly, but this essentially requires estimating a separate least-squares coefficient vector for every covariate in X_2. In many applications, the dimension of X_2 is large enough that this will be a less than entertaining task. It can be shown that when X_1 is orthogonal to all covariates in X_2, Hausman's (1978) variance-of-differences result holds (even though the sufficient conditions of his theorem need not hold, as Hausman states them). When a single coefficient is of interest, then, one can use already published results to test whether added covariates matter: using reported coefficient estimates and their estimated standard errors, one can construct a test statistic whose asymptotic distribution is standard normal under the null that X_1 and X_2 are orthogonal. Only the estimated coefficients and standard errors from the two specifications are needed to carry out such a test.

6 Conclusion

This paper has two primary results. First, comparing coefficient estimates from specifications in which covariates are added sequentially to a base model does not generally identify population

[14] Horrace and Oaxaca (2001) provide variance formulas that treat sample means as non-stochastic, whereas my more general variance formulas account for this source of covariance.

[15] Oaxaca and Ransom (1999) argue that detailed components of $\widehat{Q}$ suffer from an identification problem, due to the fact that adding an arbitrary constant c to the k^{th} variable in Z will change its sample mean while leaving its estimated coefficient unchanged. As a consequence, the detailed price component for the group including covariate k will rise by $c\widehat{\beta}_{2DZk}$. However, as Gelbach (2002a) argues, the issue they raise is not really an identification problem. Instead, it concerns where in the covariate distribution one chooses to evaluate the effect of differences in coefficients. I refer interested readers to Gelbach's paper for more discussion.

parameters of interest. Therefore, this common practice does not generally answer economically or econometrically interesting questions. Second, by using the sample version of the omitted variables bias formula, researchers can answer the questions that have led them to use sequential covariate addition in the first place. By contrast to sequential covariate addition, elements of my decomposition are estimated conditional on all covariates. As such, no sequencing problems arise. My analytical discussion shows that my decomposition is easily implemented using intuitively meaningful auxiliary regressions.

I have also presented empirical findings related to the black-white wage gap, using Neal and Johnson (1996) and Lang and Manove (2006) as points of departure. I follow the latter authors and treat education as a necessary covariate in a full wage equation specification. In my final results, I also treat occupation and industry as covariates that belong in the wage equation's full specification. Conditional on educational attainment and dummies for industry and occupation, I find that variation in test scores explain more than half of the wage gap. This is a large share, but considerably less than what Neal and Johnson calculate using more parsimonious specifications. Variation in occupation explains about a fifth of the overall black-white wage gap, while variation in educational attainment variation explains less than a tenth. Industry variation has no effect on the black-white wage gap, even though industry dummies are significantly correlated with both wages and race. This example shows that while nonzero values of β_2 and the correlation of X_1 and covariates in X_2 are necessary for a given set of covariates to contribute to the base-full difference in coefficients, they are not sufficient, even together.

I have also discussed a number of extensions. First, my decomposition nests the Oaxaca-Blinder decomposition (OBD). This result is part curiosum, but it is also practically useful, since it means my auxiliary-regression approach and variance formulas can be used to compute OBD results. Second, my results carry over exactly to the case of just-identified instrumental variables estimators. The decomposition method also works in large samples in overidentified cases. Extensions to instrumental variables is important due to the common use of linear IV estimators together with sequential covariate addition. Third, I show that cross-specification testing can be done using a Hausman test under the null hypothesis that X_1 and X_2 are uncorrelated, which is sufficient for there to be no difference across specifications in the coefficient on X_1. As a result, when only one coefficient is of interest, one can carry out specification tests using only the information typically reported in tables provided in published and working papers.

I conclude with one final observation: there really is no reason at all to sequentially add X_2 covariates to a base model. Sequential addition can obscure, overstate, or understate the true part of δ that can be attributed to variation in any given set of X_2 variables. The only meaningful way to estimate the sensitivity of β_1 to covariates is to add all the covariates at once and then compare $\widehat{\beta}_1^{\text{base}}$ and $\widehat{\beta}_1^{\text{full}}$. Providing tables with subsets of X_2 added sequentially across columns or rows thus makes little sense, and this practice should simply be abandoned. By contrast, my decomposition method has all the advantages of existing practice. Its only disadvantage has to do with variance computation. Since conventional practice has not included any systematic testing of differences across columns, even this difficulty isn't really a disadvantage by comparison to current practice. Nonetheless, I have programmed consistent variance estimators in Stata in publicly available code, so this disadvantage will be practically irrelevant.

Appendix A Sample Construction

Appendix A.1 NLSY Sample

To construct my NLSY sample, I began with all 5,579 males appearing in the non-military NLSY79 sample for 1979. I code as black those who are listed as black and non-Hispanic, and similarly for whites. My educational attainment is the highest grade ever completed as of 1990. The AFQT variable is the sum of the variables ASVAB_SEC2_81, ASVAB_SEC3_81, ASVAB_SEC4_81, and ASVAB_SEC8_81, which I then standardize to have mean 0 and standard deviation 1 in my sample. The wage variable I use is HRP_J1_1990. I use only observations on those aged 18 or younger in 1980.

I form twelve industry dummies from the following 13 groups:

1. Agriculture, Forestry, and Fisheries
2. Mining
3. Construction
4. Durable Manufacturing
5. Non-durable Manufacturing
6. Transportation, Communications, and Other Public Utilities
7. Wholesale and Retail Trade

8. Finance, Insurance, and Real Estate
9. Business and Repair Services
10. Personal Services
11. Entertainment and Recreation Services
12. Professional and Related Services
13. Public Administration

I form twelve occupation dummies from the following 13 groups:

1. Professional, Technical, and Kindred Workers
2. Managers and Administrators, except Farm
3. Sales Workers
4. Clerical and Unskilled Workers
5. Craftsmen and Kindred Workers
6. Operatives, except Transport
7. Transport Equipment Operatives
8. Laborers, except Farm
9. Farmers and Farm Managers
10. Farm Laborers and Farm Foremen
11. Service Workers, except Private Household
12. Private Household Workers
13. Additional NLSY79 Occupation Codes

All models are estimated without sampling weights, following the approach in Neal and Johnson (1996). Estimated standard errors are computed under the assumption of conditional sphericality; heteroskedasticity-robust estimated standard errors were very similar to those reported here.

Appendix B Covariance Derivation and Estimation

In this appendix, I derive and discuss consistent estimation of the covariance of the vector of decomposition elements. This vector can be written as

$$\widehat{\Delta} = \begin{pmatrix} \widehat{\delta}^1 \\ \widehat{\delta}^2 \\ \vdots \\ \widehat{\delta}^G \end{pmatrix} = \begin{pmatrix} \widehat{\Gamma}^1\widehat{\beta}_2^1 \\ \widehat{\Gamma}^2\widehat{\beta}_2^2 \\ \vdots \\ \widehat{\Gamma}^G\widehat{\beta}_2^G \end{pmatrix},$$

where we recall that there are k_1 rows in each $\widehat{\delta}^g$, $\widehat{\Gamma}^g$ is $k_1 \times k^g$, and $\widehat{\beta}_2^g$ is a $k^g \times 1$ column vector. Let $\Delta = (\delta^{1\prime}, \delta^{2\prime}, \ldots, \delta^{G\prime})'$. Consider the g^{th} component of $\sqrt{n}(\widehat{\Delta} - \Delta)$:

$$\begin{aligned} \sqrt{n}(\widehat{\delta}^g - \delta^g) &= \widehat{\Gamma}^g \sqrt{n}(\widehat{\beta}_2^g - \beta_2^g) + \sqrt{n}(\widehat{\Gamma}^g - \Gamma^g)\beta_2^g \\ &= \widehat{\Gamma}^g \sqrt{n}(\widehat{\beta}_2^g - \beta_2^g) + \left(\frac{X_1'X_1}{n}\right)^{-1} \frac{X_1'v^g}{\sqrt{n}}, \end{aligned} \tag{16}$$

where $v^g = W^g\beta_2^g$ is the population residual from taking the projection relationship between the columns of X_2^g on X_1 and post-multiplying by β_2^g, i.e., $X_2^g\beta_2^g = X_1\Gamma^g\beta_2^g + W^g\beta_2^g$. The asymptotic covariance between $\widehat{\delta}^g$ and $\widehat{\delta}^h$ involves the expectation of the product of (16) and its transposed analogue when we replace g with h:[16]

$$\begin{aligned} \text{ACov}(\widehat{\delta}^g, \widehat{\delta}^{h\prime}) = \text{plim} \Bigg\{ & \widehat{\Gamma}^g \sqrt{n}(\widehat{\beta}_2^g - \beta_2^g)\sqrt{n}(\widehat{\beta}_2^h - \beta_2^h)'\widehat{\Gamma}^{h\prime} \\ & + \left(\frac{X_1'X_1}{n}\right)^{-1} \frac{X_1'v^g v^{h\prime} X_1}{n} \left(\frac{X_1'X_1}{n}\right)^{-1} \\ & + \widehat{\Gamma}^g \sqrt{n}(\widehat{\beta}_2^g - \beta_2^g) \frac{v^{h\prime}X_1}{\sqrt{n}} \left(\frac{X_1'X_1}{n}\right)^{-1} + \left(\frac{X_1'X_1}{n}\right)^{-1} \frac{X_1'v^g}{\sqrt{n}} \sqrt{n}(\widehat{\beta}_2^h - \beta_2^h)'\widehat{\Gamma}^{h\prime} \Bigg\}. \end{aligned}$$

$$\begin{aligned} &= \Gamma^g \text{ACov}(\widehat{\beta}_2^g, \widehat{\beta}_2^{h\prime})\Gamma^h + Q_{11}^{-1} \text{plim} \left(\frac{X_1'v^g v^{h\prime}X_1}{n}\right) Q_{11}^{-1} \\ &+ \Gamma^g \text{ACov}\left[\sqrt{n}(\widehat{\beta}_2^g - \beta_2^g), \frac{v^{h\prime}X_1}{\sqrt{n}}\right] Q_{11}^{-1} + Q_{11}^{-1}\text{ACov}\left[\frac{X_1'v^g}{\sqrt{n}}, \sqrt{n}(\widehat{\beta}_2^h - \beta_2^h)',\right]\Gamma^{h\prime}. \end{aligned} \tag{17}$$

The first matrix simplifies to $\Gamma^g V_\beta^{gh}\Gamma^h$, where V_β^{gh} is the $k^g \times k^h$ submatrix of the limiting the covariance matrix $\text{AV}(\widehat{\beta}_2)$; it involves the rows corresponding to $\widehat{\beta}_2^g$ and columns corresponding to $\widehat{\beta}_2^h$. The first matrix in (17) can thus be estimated by replacing Γ^g and Γ^h with consistent estimates, $(X_1'X_1)^{-1}X_1'X_2^g$ and $(X_1'X_1)^{-1}X_1'X_2^h$, and V_β^{gh} with the corresponding submatrix from a consistent

[16] Note that we can find the asymptotic variance matrix for any $\widehat{\delta}^g$ by setting $g = h$ in (17).

estimate of the asymptotic variance matrix of the full estimator. To estimate the rest of the matrices in (17), it will help to consider the stacked system

$$\begin{bmatrix} Y \\ X_2^1\beta_2^1 \\ X_2^1\beta_2^2 \\ \vdots \\ X_2^1\beta_2^G \end{bmatrix} = \begin{bmatrix} X & 0 & 0 & \dots & 0 \\ 0 & X_1 & 0 & \dots & 0 \\ 0 & 0 & X_1 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & X_1 \end{bmatrix} \begin{bmatrix} \beta \\ \Gamma^1\beta_2^1 \\ \Gamma^2\beta_2^2 \\ \vdots \\ \Gamma^G\beta_2^G \end{bmatrix} + \begin{bmatrix} \varepsilon \\ v^1 \\ v^2 \\ \vdots \\ v^G \end{bmatrix}. \tag{18}$$

I derive the covariance matrix for OLS estimation of this system under the assumption that β_2 is observed, because this assumption turns out not to matter for the purposes of estimating covariance terms. Letting θ be the population parameters in (18), its OLS estimator can be written,

$$\widehat{\theta} = \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & \mathrm{I}_G \otimes (X_1'X_1)^{-1} \end{bmatrix} \begin{bmatrix} X'Y \\ \mathrm{I}_G \otimes X_1'X_2^g\beta_2^g \end{bmatrix}, \tag{19}$$

with covariance matrix

$$V_\theta = \begin{bmatrix} V_\beta & C'_{\beta\delta} \\ C_{\beta\delta} & V_{\text{aux}} \end{bmatrix}. \tag{20}$$

Here V_β is the covariance matrix for $\widehat{\beta}$, the parameters of the full model. The matrix in the lower-right block of V_θ is

$$V_{\text{aux}} = \left[\mathrm{I}_G \otimes \text{ plim } \left(\frac{X_1'X_1}{n}\right)^{-1}\right] \text{plim } \left(\frac{(\mathrm{I}_G \otimes X_1{}')\,\widetilde{v}\widetilde{v}'\,(\mathrm{I}_G \otimes X_1)}{n}\right) \left[\mathrm{I}_G \otimes \text{ plim } \left(\frac{X_1'X_1}{n}\right)^{-1}\right], \tag{21}$$

where $\widetilde{v}$ is the stacked vector of all residuals from the system (18). The covariance of estimated coefficients from the full model for β and the auxiliary model for Δ is $C_{\beta\delta}$. This matrix can be formed by stacking the G matrices $C_{1\beta}, C_{2\beta}, \dots, C_{G\beta}$ on top of each other, where each of these $k_1 \times (k_1 + k_2)$ matrices has the form

$$C_{g\beta} = \text{ plim } \left(\frac{X_1'X_1}{n}\right)^{-1} \text{ plim } \frac{X_1'W^g\beta_2^g\varepsilon'X}{n} \text{ plim } \left(\frac{X'X}{n}\right)^{-1}.$$

To estimate plims of $(X_1'X_1/n)$ or $(X'X/n)$, which I will call Q_{11} and Q, we can use the usual sample analogues. To estimate ε, we use the usual OLS fitted residual $\widehat{\varepsilon} = Y - X\widehat{\beta}$ from the full specification of interest. To estimate v^g, we can use the fitted residual from the g^{th} auxiliary equation in (18). As discussed in the main text, we must estimate this equation using $\widehat{\beta}_2^g$ in place of the unknown β_2^g. However, because β_2^g is consistent for β_2, nothing is lost in estimating covariance terms that involve v^g: even if we knew β_2^g, we would have to estimate v^g with

$H^g - (X_1'X_1)^{-1} X_1'X_2^g \beta_2^g$; $\widehat{v}^g = \widehat{H}^g - (X_1'X_1)^{-1} X_1'X_2^g \widehat{\beta}_2^g$ is consistent and therefore just as good in large samples. When conditional heteroskedasticity or group-level clustering is suspected, the usual robust covariance estimators can be used to estimate the middle matrices in V_β, V_{aux} and $C_{g\beta}$.

Appendix B.1 Conditional sphericality

When conditional sphericality holds, all the matrices just defined will simplify, since the middle matrix of each then reduces to a scalar multiple of the outer matrices. In this case the limiting variance of $\sqrt{n}(\widehat{\beta} - \beta)$ is $V_\beta = \sigma_{\varepsilon\varepsilon} Q^{-1}$, where $\sigma_{\varepsilon\varepsilon} = E[\varepsilon_i^2 | X_i]$ and $\text{V}_{\text{aux}} = \Sigma_{vv} \otimes (X'X)^{-1}$, where Σ_{vv} is $G \times G$ with scalar gh element $\sigma_{gh} = E[v_i^g v_i^h | X_1]$ for all i. When $\sigma_{g\varepsilon} = E[v_i^g \varepsilon_i | X]$ for all i, $C_{g\beta}$ simplifies to $\sigma_{g\varepsilon} \begin{bmatrix} \text{I}_{k_1} & \Gamma \end{bmatrix} Q^{-1}$. Using the partitioned inverse formula, it can be shown that the columns of this matrix corresponding to $\widehat{\beta}_2$ are necessarily zero, so that $\widehat{\delta}^g$ and $\widehat{\beta}^g$ have zero covariance in this case.

We now have the ingredients to estimate the four matrices in $\text{ACov}(\widehat{\delta}^g, \widehat{\delta}^{h\prime})$ from (17); in this discussion, I use a hat to denote a consistent estimate of any matrix. We have already seen that the first matrix can be estimated with $\widehat{\Gamma}^g \widehat{V}_\beta^{gh} \widehat{\Gamma}^{h\prime}$. The second matrix can be estimated with the block of $\widehat{V}_{\text{aux}}$ that corresponds to the rows of the g^{th} auxiliary equation and the columns of the h^{th}. To estimate the third and fourth matrices, let the $k_1 \times k^h$ matrix $C_{g\beta}(h)$ be the k^h columns of $C_{g\beta}$ that correspond to the X_2 covariates in group h; this submatrix is the covariance of $(X_1'X_1)^{-1} X_1'H^g$ and $\widehat{\beta}_2^h$. The fourth matrix in (17) can be estimated with $\widehat{C}_{g\beta}(h)\widehat{\Gamma}^{h\prime}$, and the third matrix can be estimated with $\widehat{\Gamma}^g \widehat{C}_{h\beta}'(g)$. As discussed above, $\widehat{C}_{g\beta}(h)$ and $\widehat{C}_{h\beta}'(g)$ equal zero matrices when conditional sphericality holds, so under sphericality, there is no need to estimate either covariance term.

References

Altonji, Joseph and Rebecca Blank, "Race and Gender in the Labor Market," in Orley Ashenfelter and David Card, eds., Handbook of Labor Economics, Vol. 3C, Amsterdam, New York, and Oxford: Elsevier Science, North-Holland, 1999.

Ayres, Ian, "Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of "Included Variable" Bias," Perspectives in Biology and Medicine, 2005, 48 (1), S68–S87.

Bound, John and Richard B. Freeman, "What Went Wrong? The Erosion of relative earnings and employment among young black men in the 1980s," Quarterly Journal of Economics, February 1992, 107 (1), 201–32.

Buckles, Kasey and Daniel M. Hungerman, "Season of Birth and Later Outcomes: Old Questions, New Answers," Technical Report 14573, NBER December 2008.

Caceres-Delpiano, Julio, "The Impacts of Family Size on Investment in Child Quality," Journal of Human Resources, 2006, XLI (Fall), 738–54.

Card, David and Alan B. Krueger, Myth and Measurement, Princeton, NJ: Princeton University Press, 1995.

Charles, Kerwin Kofi and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination," Journal of Political Economy, 2008, 116 (5).

Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt, "The impact of school choice on student outcomes: an analysis of the Chicago Public Schools," Journal of Public Economics, 2005, 89 (5-6), 729 – 760.

Deaton, Angus, "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll," Journal of Economic Perspectives, Spring 2008, 22 (2), 53–72.

DiNardo, John E. and Jorn-Steffen Pischke, "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?," Quarterly Journal of Economics, February 1997, 112 (1), 291–303.

Donohue, John J. III and Steven D. Levitt, "The Impact of Race on Policing and Arrests," Journal of Law and Economics, October 2001, 44 (2), 367–94.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," American Economic Review, September 2001, 91 (4).

Edlund, Lena and Rohini Pande, "Why Have Women Become Left-Wing? The Political Gender Gap and the Decline in Marriage," Quarterly Journal of Economics, August 2002, 117 (3), 917–61.

Gelbach, Jonah B., "Identified Heterogeneity in Detailed Wage Decompositions," December 2002. Available at http://glue.umd.edu/~gelbach/papers/comment-or/submitted-12-16-02-comment-or.pdf.

____ , "Public Schooling for Young Children and Maternal Labor Supply," American Economic Review, March 2002, 92 (2), 307–322.

____ , "When Do Covariates Matter? And Which Ones, and How Much?," June 2009. University of Arizona Department of Economics Working Paper #09-07, available at http://papers.ssrn.com/abstract=1425737.

Goldberger, Arthur S., A Course in Econometrics, Cambridge, MA: Harvard University Press, 1991.

Goldin, Claudia and Lawrence F. Katz, "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," Journal of Political Economy, August 2002, 110 (4), 730–70.

Greenstone, Michael, "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures," Journal of Political Economy, December 2002, 110 (6), 1175–1219.

Guryan, Jonathan, "Desegregation and Black Dropout Rates," American Economic Review, September 2004, 94 (4), 919–43.

Haider, Steven J., Robert F. Schoeni, Yuhua Bao, and Caroline Danielson, "Immigrants, Welfare Reform, and the Economy," Journal of Policy Analysis and Management, 2004, 23 (4), 745–64.

____ , Todd E. Elder, and John H. Goddeeris, "Unexplained Gaps and Oaxaca-Blinder Decompositions," Technical Report, Michigan State University April 2009.

Hanushek, Eric A. and Ludger Woessmann, "The Role of Cognitive Skills in Economic Development," Journal of Economic Literature, 2008, 46 (3), 607–68.

Hausman, Jerry, "Specification Tests in Econometrics," Econometrica, November 1978, 46 (6), 1251–71.

Hellerstein, Judith K. and David Neumark, "Workplace Segregation in the United States: Race, Ethnicity, and Skill," Review of Economics and Statistics, August 2008, 90 (3), 459–77.

Horrace, William C. and Ronald L. Oaxaca, "Interindustry wage differentials and the gender wage gap: an identification problem," Industrial and Labor Relations Review, 2001, 54, 611–18.

Kane, Thomas J., Stephanie K. Riegg, and Douglas O. Staiger, "School Quality, Neighborhoods, and Housing Prices," American Law and Economics Review, 2006, 8 (2), 183–212.

Kessler, Daniel and Mark McClellan, "Do Doctors Practice Defensive Medicine?," Quarterly Journal of Economics, May 1996, 111 (2), 353–90.

Khwaja, Asim Ijaz and Atif Mian, “Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market,” American Economic Review, 2008, 98 (4), 1413–42.

Klick, Jonathan and Thomas Stratmann, “Medical Malpractice Reform and Physicians in High-Risk Specialties,” Journal of Legal Studies, June 2007, 36.

Krueger, Alan B., “How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989,” Quarterly Journal of Economics, February 1993, 108 (1), 33–60.

Lang, Kevin and Michael Manove, “Education and Labor-Market Discrimination,” Technical Report 12257, NBER May 2006.

Leamer, Edward E., “Sensitivity Analyses Would Help,” American Economic Review, June 1985, 75 (3), 308–13.

Lee, David S. and Justin McCrary, “Crime, Punishment, and Myopia,” Technical Report 11491, NBER June 2005.

Lefgren, Lars and Frank McIntyre, “Explaining the Puzzle of Cross-State Differences in Bankruptcy Rates,” Journal of Law and Economics, May 2009, 52.

Levitt, Steven D. and Chad Syverson, “Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions,” Review of Economics and Statistics, August 2008, 90, 599–611.

Moore, William J., Robert J. Newman, and Geoffrey K. Turnbull, “Do Academic Salaries Decline with Seniority?,” Journal of Labor Economics, April 1998, 16 (2), 352–66.

Neal, Derek A. and William R. Johnson, “The Role of Premarket Factors in Black-White Wage Differences,” Journal of Political Economy, October 1996, 104 (5), 869–95.

Neumark, David, “Employers’ Discriminatory Behavior and the Estimation of Wage Discrimination,” Journal of Human Resources, 1988, pp. 279–295.

Oaxaca, Ronald L. and Michael R. Ransom, “Identification in detailed wage decompositions,” Review of Economics and Statistics, February 1999, 81 (1).

Sharkey, Catherine M., “Unintended Consequences Of Medical Malpractice Damages Caps,” New York University Law Review, May 2005, 80 (2).

Stevenson, Betsey and Justin Wolfers, “Bargaining in the Shadow of the Law: Divorce Laws and Family Distress,” Quarterly Journal of Economics, February 2006, 121 (1).

Wooldridge, Jeffrey M., Econometric analysis of cross section and panel data, Cambridge, MA: The MIT Press, 2002.

Table 1: Results from Table 4 of Altonji & Blank (1999)

	Model:		
	(4)	(5)	(6)
Coefficient on black indicator	-0.207	-0.117	-0.089
Included covariates:			
Hispanic and female indicators	Y	Y	Y
Education, experience, & region	N	Y	Y
Occupation, industry, job characteristics	N	N	Y

Table 2: Results from Tables 1 and 2 of Levitt & Syverson (2008)

	Model:				
	Table 1	Table 2			
	Column (4)	(1)	(2)	(3)	(4)
Coefficient on agent dummy	0.128	0.048	0.042	0.038	0.037
Included covariates:					
City-by-year interactions	Y	Y	Y	Y	Y
Basic house characteristics	N	Y	Y	Y	Y
Indicators of house quality	N	N	Y	Y	Y
Keywords in description	N	N	N	Y	Y
Block fixed effects	N	N	N	N	Y

Table 3: NLSY Wage Results for Young Men, Similar to Table 1 of Neal & Johnson (1996)

Panel A: Black-white wage gap results

	(1)	(2)	(3)	(4)
Coefficient on black indicator	-22.1	-19.6	-7.2	-11.7
(log points)	(3.6)	(3.5)	(3.8)	(3.9)
Included covariates:				
Hispanic indicator, and age	Y	Y	Y	Y
Education	N	Y	N	Y
AFQT	N	N	Y	Y
N	1,749	1,749	1,749	1,749

Panel B: Implied decomposition components

	Sequence	
	E-A	A-E
Education completed as of 1990	-2.5	4.5
AFQT	-7.9	-14.9
Total	-10.4	-10.4

Note: All equations include the same sample used throughout and include 1,749 observations. Estimated standard errors appear in parentheses.

Table 4: The Black-White Gap in Years of Schooling Completed by 1990s

	(1)	(2)
Coefficient on black indicator	-0.37	1.10
(log points)	(0.14)	(0.11)
N	1,749	1,749
Included covariates:		
Hispanic indicator, and age	Y	Y
AFQT	N	Y

Note: Reported coefficients are estimates of $\Gamma_{\text{race}}^{\text{educ}}$ and $\Gamma_{\text{race}}^{\text{AFQT}}$, computed as the coefficient on a black indicator from least-squares estimation of equations with education/AFQT as the dependent variable and a constant and Hispanic and age controls included. All equations include the same sample used throughout and include 1,749 observations. Estimated standard errors appear in parentheses.

Table 5: Understanding the Decomposition of $\widehat{\delta}_{\text{race}}$

Panel A. Conditional Decomposition Components

	Education	AFQT	Total
$\widehat{\Gamma}_{\text{race}}$	-0.37	-0.87	
$\widehat{\beta}_2$	0.041	0.1013	
$\widehat{\delta}_{\text{race}}$ $(=\widehat{\Gamma}_{\text{race}} \times \widehat{\beta}_2)$	-0.0155	-0.0881	-0.104

Panel B. Sequential Decomposition Components

	Education	AFQT	Total
Sequence 1:			
Education-AFQT	-0.0249	-0.0786	-0.104
Bias	-0.0095	0.0095	0.00
Sequence 2:			
AFQT-Education	0.0453	-0.1489	-0.104
Bias $(=\widehat{\Gamma}_{\text{race}} \times \widehat{\beta}_2)$	0.0608	-0.0608	0.00

Note: Estimates of $\Gamma_{\text{race}}^{\text{educ}}$ and $\Gamma_{\text{race}}^{\text{AFQT}}$ are computed as the coefficient on a black indicator from least-squares estimation of equations with education/AFQT as the dependent variable and a constant and Hispanic and age controls included. Estimates of $\beta_{2,\text{educ}}$ and $\beta_{2,\text{AFQT}}$ are taken from the full specification of the wage equation, which includes all X_1 variables (a constant, black and Hispanic indicators, and age) and both X_2 variables (AFQT and years of schooling). All equations include the same sample used above and include 1,749 observations.

Table 6: Decomposing the Black-White Wage Gap Into Education, AFQT, Occupation and Industry Components

	Specification		
	Base	Full	Explained
Black-white gap	-22.11	-9.26	-12.85
	(3.63)	(3.81)	(2.26)
Covariates			
AFQT	N	Y	-7.06
			(1.84)
Education	N	Y	-1.56
			(0.66)
12 Occupation dummies	N	Y	-4.34
			(1.03)
12 Industry dummies	N	Y	0.11
			(0.90)

Note: Standard errors in parentheses (computed under the assumption of conditional sphericality). Both models include Hispanic-background and linear age controls, and both models use the same 1,749 observations.