

# Omitted Variables, Countervailing Effects, and the Possibility of Overadjustment\*

KEVIN A. CLARKE, BRENTON KENKEL AND MIGUEL R. RUEDA

*The effect of conditioning on an additional covariate on confounding bias depends, in part, on covariates that are unobserved. We characterize the conditions under which the interaction between a covariate that is available for conditioning and one that is not can affect bias. When the confounding effects of two covariates, one of which is observed, are countervailing (in opposite directions), conditioning on the observed covariate can increase bias. We demonstrate this possibility analytically, and then show that these conditions are not rare in actual data. We also consider whether balance tests or sensitivity analysis can be used to justify the inclusion of an additional covariate. Our results indicate that neither provide protection against overadjustment.*

Empirical researchers working with observational data rely on covariate adjustment to analyze causal effects. Selecting just the right covariates for conditioning is essential for the elimination of confounding bias. Until quite recently, the prevailing wisdom in political science and much of economics held that one should condition on all available covariates in the hopes of minimizing the bias caused by omitted confounders. Research conducted across a variety of domains has eroded that belief, and many now admit that overadjustment—conditioning on covariates that increase bias—is a real possibility. Nevertheless, there remain scholars, Pearl (2010) writes of those in the “experimentalist” camp, who deride the notion that in some cases a researcher should not condition on all available covariates (see e.g., Rubin 2009).

The problem of covariate selection is made more difficult by the realization that the set of covariates available for adjustment is a subset of the set of all relevant covariates. That is, the effect on the bias of including an additional covariate in the conditioning set may be determined, in part, by variables unavailable to the analyst (Clarke 2005; Clarke 2009; DeLuca, Magnus and Peracchi 2015). Our goal in this paper is to characterize more fully the conditions under which the interaction between a covariate that is available for conditioning and a covariate that is not can affect confounding bias.

Our discussion takes place within the potential outcomes framework (the online Appendix contains a brief primer) and is related to the bias amplification literature (Pearl 2010; Pearl 2011; White and Lu 2011). “Bias amplifying” refers to covariates that, if conditioned on, will increase existing bias. Such variables tend to be those that have greater effects on the treatment

---

\* Kevin A. Clarke is an Associate Professor in the Department of Political Science, University of Rochester, Rochester, NY 14627-0146 (kevin.clarke@rochester.edu). Brenton Kenkel is an Assistant Professor in the Department of Political Science, Vanderbilt University, Nashville, TN 37203 (brenton.kenkel@vanderbilt.edu). Miguel R. Rueda is an Assistant Professor in the Department of Political Science, Emory University, Atlanta, GA 30233 (miguel.rueda@emory.edu). A previous version of this paper was given at the 27th Annual Summer Meeting of the Society for Political Methodology. The authors thank the participants for their comments. The authors thank Jake Bowers, John Jackson, Michael Peress, the editor, and two anonymous reviewers for helpful comments and discussion. Brad Smith provided excellent research assistance. Errors remain the authors own. To view supplementary material for this article, please visit <https://doi.org/10.1017/psrm.2016.46>

than on the outcome. Pearl (2010), building on work by Bhattacharya and Vogt (2007) and Wooldridge (2009), demonstrates that instrumental variables (IVs) are bias amplifying. That is, by including an IV in a conditioning set, one will increase any existing confounding bias.

We analyze a different situation where an available confounder and an unavailable confounder have *countervailing effects*. That is, we consider the case where the confounding effects of the two variables are in opposite directions, but do not offset each other exactly. Under these conditions, including the available confounder in the conditioning set increases the bias. We first demonstrate this possibility analytically, and then we show that these conditions occur in applications. In addition, we consider whether balance tests or sensitivity analysis can be used to justify the inclusion of additional covariates. Our results show that it is possible for a covariate to improve balance while increasing bias. Finally, we demonstrate that sensitivity analysis cannot alert us to the possibility of countervailing effects because sensitivity analysis addresses a different question.

Our findings lend little credence to the claim that a researcher should condition on all available pretreatment covariates. Which variables should be included in a data analysis depends on factors that vary from situation to situation. We can tackle the problem using theory, judgment, and common sense, and we end with a discussion of how our results can be helpful to researchers.

## ANALYSIS

We use a simple analytical example to illustrate the conditions under which conditioning on an observed variable can increase the bias in treatment effect estimation. Our analysis closely follows that of Pearl (2010). The main difference is that we assume the treatment variable is binary, as in the prototypical examples of causal inference in political science (e.g., Ho et al. 2007).<sup>1</sup> Whereas Pearl focuses on the bias amplification properties of IVs—those that affect treatment assignment but not the outcome—we characterize the conditions under which adjusting for a true confounder can increase bias. If an observed confounder and an unobserved confounder have *countervailing effects*, a condition we define formally below, then controlling for the observed variable may worsen the bias of a treatment effect estimator.

Assume there are two covariates:  $X$ , which is observed, and  $U$ , which is not. Each is drawn independently from a Bernoulli distribution with probability  $\frac{1}{2}$ . There is a binary treatment  $T$ , whose probability is a linear function of the covariates:

$$\Pr(T=1) = \frac{1}{2} + \gamma_X X + \gamma_U U, \quad (1)$$

where  $|\gamma_X| < \frac{1}{2}$ ,  $|\gamma_U| < \frac{1}{2}$ , and  $|\gamma_X + \gamma_U| < \frac{1}{2}$ . The outcome  $Y$  is a linear function of the treatment, the covariates, and a white noise random variable  $\varepsilon$

$$Y = \alpha + \tau T + \beta_X X + \beta_U U + \varepsilon \quad (2)$$

where  $E[\varepsilon | T, X, U] = 0$ .

The analyst's goal is to estimate the average treatment effect  $\tau$  from a sequence of observations of  $(Y, T, X)$ . Because  $U$  is unobserved, no estimator of  $\tau$  exists that is unbiased for

<sup>1</sup> We ran several Monte Carlo experiments that use continuous explanatory variables that are consistent with our analytic results.

all possible sets of parameters ( $\tau, \alpha, \gamma_X, \gamma_U, \beta_X, \beta_U$ ). The analyst faces a choice between adjusting for  $X$  via some matching or weighting scheme, or estimating  $\tau$  by an unadjusted difference of means. The expected value of the naive difference of means estimator is

$$\begin{aligned}
 E[\hat{\tau}_0] &= E[Y|T=1] - E[Y|T=0] \\
 &= \tau + \frac{\beta_X \gamma_X + \beta_U \gamma_U}{1 - (\gamma_X + \gamma_U)^2}.
 \end{aligned}
 \tag{3}$$

The denominator of the second term is strictly positive, so the magnitude of the bias of this estimator is

$$|\text{Bias}(\hat{\tau}_0)| = \frac{|\beta_X \gamma_X + \beta_U \gamma_U|}{1 - (\gamma_X + \gamma_U)^2}.
 \tag{4}$$

This estimator is unbiased if and only if neither covariate is a confounder ( $\beta_X \gamma_X = \beta_U \gamma_U = 0$ ), or their contributions to the bias are exactly offsetting ( $\beta_X \gamma_X = -\beta_U \gamma_U$ ). This latter condition proves important when we look at when it is worse, in terms of bias, to control for  $X$ .

Now consider an estimator that conditions on the observed variable  $X$ . Because  $X$  is binary, a natural way to estimate the treatment effect is by subclassification (Rosenbaum and Rubin 1983a): take the average of the within-group differences of means, where the groups are defined by the values of  $X$ .<sup>2</sup> The expected value of this estimator is

$$\begin{aligned}
 E[\hat{\tau}_X] &= \frac{1}{2}(E[Y|T=1, X=1] - E[Y|T=0, X=1]) \\
 &\quad + \frac{1}{2}(E[Y|T=1, X=0] - E[Y|T=0, X=0]) \\
 &= \tau + \frac{\beta_U \gamma_U}{2} \left( \frac{1}{1 - (2\gamma_X + \gamma_U)^2} + \frac{1}{1 - \gamma_U^2} \right).
 \end{aligned}
 \tag{5}$$

Therefore, the magnitude of the bias when adjusting for the observed variable  $X$  via subclassification is

$$|\text{Bias}(\hat{\tau}_X)| = \frac{|\beta_U \gamma_U|}{2} \left( \frac{1}{1 - (2\gamma_X + \gamma_U)^2} + \frac{1}{1 - \gamma_U^2} \right).
 \tag{6}$$

This estimator is unbiased if and only if  $U$  is not a confounder ( $\beta_U \gamma_U = 0$ ).

We can now find the parameters under which the naive estimator is less biased than the one that conditions on the observed covariate  $X$ . Controlling for  $X$  worsens the bias when the following inequality holds:

$$\frac{|\beta_U \gamma_U|}{2} \left( \frac{1}{1 - (2\gamma_X + \gamma_U)^2} + \frac{1}{1 - \gamma_U^2} \right) \geq \frac{|\beta_X \gamma_X + \beta_U \gamma_U|}{1 - (\gamma_X + \gamma_U)^2}.
 \tag{7}$$

From this expression, we can derive a set of simple sufficient conditions for when it is worse to condition on  $X$ .

<sup>2</sup> If we weighted the average by the observed proportion of observations in each group ( $X = 0$  and  $X = 1$ ), then Equation 5 only gives the asymptotic bias.

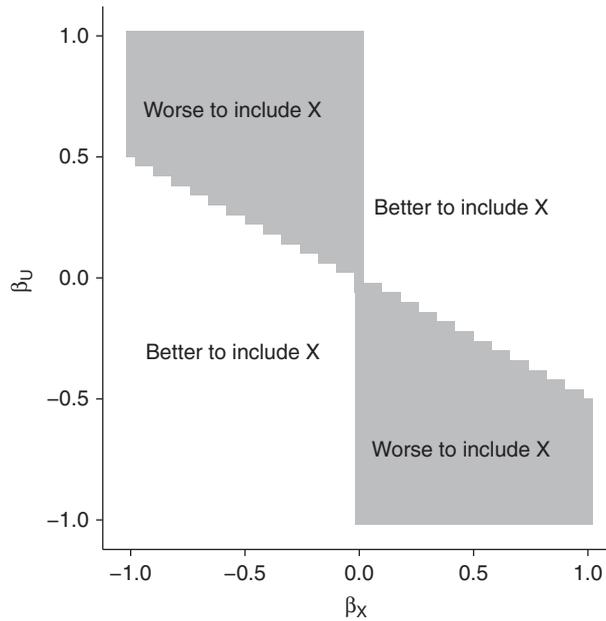


Fig. 1. Numerical illustration of the conditions under which the naive difference of means estimator is less biased than one that conditions on the observed variable  $X$ . The effects on treatment assignment are held fixed at  $\gamma_X = \gamma_U = 1/8$ , so the countervailing effects condition holds when  $\beta_X$  and  $\beta_U$  have opposite signs.

PROPOSITION 1: If all of the following conditions hold, then  $|\text{Bias}(\hat{\tau}_X)| > |\text{Bias}(\hat{\tau}_0)|$ .

- $U$  is a confounder:  $\beta_U \gamma_U \neq 0$ .
- $U$  and  $X$  have countervailing effects:  $\beta_U \gamma_U$  and  $\beta_X \gamma_X$  have opposite signs.
- The confounding effect of  $U$  is at least as great as that of  $X$ :  $|\beta_U \gamma_U| \geq |\beta_X \gamma_X|$ .

The intuition behind the countervailing effects condition is straightforward. If a confounding variable increases both the chance of treatment assignment and the expected value of the outcome, then failing to control for it causes the resulting treatment effect estimate to be biased upward on average. Conditioning on such a variable will, in expectation, decrease the estimate of the treatment effect. The same is true of a variable that has a negative relationship with both treatment assignment and the outcome. Conversely, if a variable reduces the chance of treatment and increases the expected outcome (or vice versa), then failing to control for it leads to a downward bias in the treatment effect estimate. Countervailing effects means that the confounding effects of  $X$  and  $U$  go in opposite directions—that omitting one biases the estimated treatment effect downward, while omitting the other biases it upward.

When the confounding effects of the two variables are countervailing and the magnitude of the unobserved variable  $U$ 's effect is greater, it is worse to condition on the observed variable  $X$ . To see why, imagine that  $U$  has a strong positive effect on both treatment assignment and the outcome, while  $X$  has a weak negative effect on treatment assignment and a weak positive effect on the outcome. In isolation, failing to control only for  $U$  would bias the estimated treatment effect upward, while failing to control only for  $X$  would bias it slightly downward. Failing to control for either results in a moderate upward bias. Controlling for  $X$  without controlling for  $U$  (because the latter is unobserved) would on average cause the estimated treatment effect to go up—the wrong direction.

## EMPIRICAL EXAMPLES

Our analytic results identify conditions under which adjusting for all available pretreatment variables could lead to an increase in the bias of estimated treatment effects. When considering whether to condition on a variable, a researcher must take two sets of factors into account. The first is the potential effects of an unobserved confounder on treatment and outcome, and the second is the size of those effects relative to those of the conditioning variable.

In this section, we go beyond our analytic results and demonstrate that countervailing effects are not a mathematical curiosity, but a problem that occurs with regularity in applications. We use two well-known data sets to make the point. At the same time, we consider whether balance tests, frequently used to justify the inclusion of covariates, provide a false sense of security when countervailing effects are present.

We examine these issues using two different data sets. The first is the well-known study of the impact of the National Support Work Demonstration (NSW) labor training program on post-intervention income (LaLonde 1986). The NSW was implemented in the 1970s to provide work experience to the poor. The data set is used widely to evaluate the performance of treatment effect estimators for observational data because the effects of NSW on income were evaluated experimentally. Researchers can therefore compare their estimates with the experimental benchmark (LaLonde 1986; Dehejia and Wahba 1999). In his original study, LaLonde used individuals from the Current Population Survey (CPS) as control units for comparison. We use the CPS control units plus the original treated ones for our demonstration.<sup>3</sup>

The data set contains pairs of variables that could be  $X$  and  $U$  and that simultaneously satisfy three conditions: (1) they have countervailing effects, (2) balance improves when conditioning on  $X$ , and (3) the bias on the estimated average treatment effect on the treated (ATT) increases when  $X$  is included in the conditioning set. Given that we are using observational data, we do not know the true effects of a given pair  $X$  and  $U$  on the outcome and treatment, which prevent us from knowing whether such effects are countervailing. To circumvent this problem, we estimate these effects by relying on the propensity score specification that gave Dehejia and Wahba (1999) the closest ATT estimate to the experimental benchmark.<sup>4</sup>

To clarify, consider the pair of variables  $u74$  (1 if there are no reported earnings in 1974 and 0 otherwise) and  $black$  (1 if the person is black and 0 otherwise). Let  $u74$  be the potential conditioning variable  $X$ , and let  $black$  be  $U$ , the unobserved variable. In this case, a researcher interested in evaluating the training program must decide whether to include  $u74$  in the conditioning set when information on race,  $black$ , is not available. Note that it is entirely reasonable to control for whether a person reported zero earnings previously, as this variable could proxy for unobserved characteristics that determine her future salaries.

Using Dehejia and Wahba's specification (the one that gives the best estimate of the ATT with these data), we find that the effect of  $black$  on income is negative, while its effect on treatment assignment is positive. The estimated effect of  $u74$  is positive for both income and treatment, although the coefficient in the income equation is small and not significant.<sup>5</sup> Importantly, we find that the confounding effect of  $black$  is larger than that of  $u74$ . We expect that the bias to increase

<sup>3</sup> The data set is included in the R package of Random Recursive Partitioning (Iacus 2007). It has a total of 16,177 observations with no missingness on the variables from the original study. For a description of the data set, see Dehejia and Wahba (1999, 1054).

<sup>4</sup> The specification includes the variables:  $age$ ,  $age^2$ ,  $age^3$ ,  $education$ ,  $education^2$ ,  $no\ degree$ ,  $married$ ,  $black$ ,  $hispanic$ ,  $re74$ ,  $re75$  (real earnings in 1974 and 1975),  $u74$ ,  $u75$  (indicators of zero earnings in 1974 and 1975), and  $education \times re74$ .

<sup>5</sup> The estimated linear regression coefficients are:  $\hat{\beta}_U = -574.4$ ,  $\hat{\gamma}_U = 0.103$ ,  $\hat{\beta}_X = 365.9$ , and  $\hat{\gamma}_X = 0.045$ .

TABLE 1 *Labor Training Program Example (Countervailing Effects and Balance Test)*

	$\text{sgn}(\hat{\beta}_U \hat{\gamma}_U)$	$\text{sgn}(\hat{\beta}_X \hat{\gamma}_X)$	Unobserved ( $U$ )	Added Variable ( $X$ )	$\Delta$ Bias/ATT	Vars. p-Val Increased/Total
1	-	+	<i>black</i>	<i>u74</i>	0.385	7/12
2	+	-	<i>no degree</i>	<i>married</i>	0.145	7/12
3	+	-	<i>no degree</i>	<i>education</i> <sup>2</sup>	0.114	7/12
4	+	-	<i>no degree</i>	<i>re74</i>	0.11	7/12
5	+	-	<i>education</i>	<i>age</i>	0.096	4/10
6	-	+	<i>u75</i>	<i>no degree</i>	0.084	6/11
7	-	+	<i>re75</i>	<i>no degree</i>	0.084	6/11
8	+	-	<i>education</i>	<i>age</i> <sup>2</sup>	0.077	7/10
9	-	+	<i>black</i>	<i>no degree</i>	0.063	8/12
10	-	+	<i>married</i>	<i>no degree</i>	0.06	9/12
11	-	+	<i>age</i>	<i>education</i>	0.059	8/10
12	-	+	<i>age</i>	<i>u74</i>	0.053	8/10
13	+	-	<i>education</i>	<i>age</i> <sup>3</sup>	0.048	6/10
14	+	-	<i>no degree</i>	<i>age</i> <sup>2</sup>	0.015	7/12
15	-	+	<i>married</i>	<i>u74</i>	0.014	9/12
16	+	-	<i>no degree</i>	<i>education</i> $\times$ <i>re74</i>	0.008	5/12
17	-	+	<i>black</i>	<i>education</i>	0.007	8/12

The column  $\Delta$  Bias/ATT is the increase in bias after adjusting on  $X$  as a fraction of the experimental average treatment effect. The last column gives the fraction of control variables that had an increase in the p-value of an equality of means test between treated and control units after controlling for  $X$ .

*black* = 1 if black, 0 otherwise; *no degree* = 1 if no high school degree, 0 otherwise; *married* = 1 if married, 0 otherwise; *age* = age in years; *re74* = real earnings in 1974; *re75* = real earnings in 1975; *u74* = 1 if earnings in 1974 are 0, 0 otherwise; *u75* = 1 if earnings in 1975 are 0, 0 otherwise.

after adjusting for  $u74$ , and in concert with our analytic results, that is what we observe when estimating the ATT using caliper matching.<sup>6</sup> The bias on the estimated ATT is \$691 larger when we include  $u74$  in the propensity score than when we do not. This bias is 38.50 percent of the experimental treatment effect. We also perform balance tests after matching with and without  $u74$ . We find that by including  $u74$ , the p-values on the equality of means  $t$ -tests between treated and untreated matched units increases for seven variables (out of a total of 12 covariates common to both specifications). These balance tests results could be used as justification for including  $u74$  when in fact it increases bias.

*black* and  $u74$  are not a unique pair. Table 1 shows 16 other cases where a countervailing effect is present and including  $X$  improves balance for a number of covariates, but increases bias. The first two columns contain the sign of the product of the estimated effects of  $U$  and  $X$  on treatment and outcome that identify the countervailing effect. The remaining columns contain the variables that stand in for  $X$  and  $U$ , the increase in bias as a fraction of the ATT when adjusting for  $X$ , and the fraction of variables whose p-values, in an equality of means test between treated and control matched units, increase when adjusting for  $X$ . In half of the cases, the increase in bias is larger than 7 percent of the true ATT (rows are ordered by the relative size of the bias from largest to smallest). Moreover, the average fraction of the covariates whose balance improves after adjusting for the additional variable is 0.62. These results indicate that countervailing effects associated with increases in bias are not rare in these data, and that improving balance by conditioning on an additional covariate does not necessarily mean reducing bias.

We ran the same analysis using a separate data set and found similar results. As before, we need a data set that includes an experimental benchmark. The data come from Mackenzie,

<sup>6</sup> The treatment units are matched with the closest control unit that is within 0.25 SD of the linear prediction of the propensity score.

Gibson and Stillman (2010), in which the authors study the effect of migration on income.<sup>7</sup> They focus on New Zealand, which uses a random ballot to choose among the excess number of Tongan immigration applicants. Unlike the previous example, the authors find that non-experimental methods (other than IVs estimation) overstate the effect of migration by 20–82 percent. For our purposes, we use the specification that gave them the closest estimate of the experimental effect using observational methods.<sup>8</sup>

The results are in Table 2. There are 28 cases where bias increases after adjusting for a covariate when countervailing effects are present. The increases in bias are generally smaller than in the labor training example reaching a maximum of 8 percent of the ATT. For the 28 cases, the average fraction of control variables where balance improved after adjusting for an additional covariate was 0.45.

Note that for some combinations of  $X$  and  $U$  it seems intuitive to adjust for the additional regressor. Consider row 17 where *age* is the unobserved variable, and the researcher is considering whether to include an indicator of being born in Tongatapu—the most populous island in Tonga. Living in Tongatapu is likely to affect both income and the likelihood of migration positively, as its residents benefit from a more dynamic economy and are more exposed to the outside world than residents of other Tongan islands. If we ignore other potential confounders, it would appear that not adjusting for this variable would overstate the effect of migration. However, the fact that age has a large positive effect on income, but a negative effect on migration makes conditioning on being born in Tongatapu bad for estimation. Conditioning on *born Tongatapu* increases bias by further decreasing an already understated ATT.

Whether countervailing effects exist with magnitudes of the correct size depends on the application at hand. The demonstrations above show that in at least two applications, it is easy to find the conditions highlighted in our theoretic results. Moreover, we show that adjusting for an additional covariate can improve balance without reducing bias. The intuition is that even if balance improves among most of the observed covariates when adjusting for an additional variable, nothing guarantees that the same would happen for unobserved confounders. The differences in the distributions of relevant unobservables between treated and untreated units that remain continue biasing the estimates of interest regardless of the improved balance among observables.

#### SENSITIVITY ANALYSIS

Discussions concerning causal analysis and the effects of unobserved confounders naturally lead to calls for sensitivity analysis. An unobserved covariate that has countervailing effects with a covariate for which we have adjusted could, if we observed it, change our estimate of the treatment effect. If we could show that an unobserved covariate does not exist, then we would have more confidence in our estimate of the treatment effect.

<sup>7</sup> To the best of our knowledge, the only studies with political science applications that include experimental and non-experimental measures of treatment effects are Arceneaux, Gerber and Green (2006) and Arceneaux, Gerber and Green (2010). These papers, however, did not find that observational methods were able to approximate the experimental estimates. Although observational methods did not recover the experimental estimate in Mackenzie, Gibson and Stillman (2010) either, their results under some specifications of the propensity score were closer to the experimental estimate than the results in the Arceneaux, Gerber and Green papers.

<sup>8</sup> The specification includes the variables: *age*, *age*<sup>2</sup>, *age*<sup>3</sup>, *age*<sup>4</sup>, *born Tongatapu*, *height*, *male*, *married*, *past income*, *education*, *education*<sup>2</sup>, *education*<sup>3</sup>, *education*<sup>4</sup>, *male* × *age*, *male* × *education*, *male* × *married*, *male* × *born Tongatapu*, *male* × *past income*, *male* × *height*, *age* × *education*. For a complete description of the data see Mackenzie, Gibson and Stillman (2010, 918).

TABLE 2 Migration Example (Countervailing Effects and Balance Test)

	$sgn(\hat{\beta}_U \hat{\gamma}_U)$	$sgn(\hat{\beta}_X \hat{\gamma}_X)$	Unobserved ( $U$ )	Added Variable ( $X$ )	$\Delta$ Bias/ATT	Vars. p-Val Increased/Total
1	-	+	height	male $\times$ education	0.078	2/17
2	-	+	height	male	0.063	1/17
3	+	-	education	age	0.058	8/13
4	+	-	education	age <sup>2</sup>	0.054	8/13
5	+	-	education	past income	0.049	10/13
6	+	-	education	age <sup>3</sup>	0.048	6/13
7	-	+	past income	male $\times$ education	0.048	9/17
8	-	+	past income	male $\times$ married	0.047	4/17
9	-	+	height	education <sup>4</sup>	0.045	3/17
10	-	+	past income	male $\times$ height	0.039	11/17
11	-	+	male $\times$ age	age <sup>4</sup>	0.037	8/12
12	+	-	male	age <sup>3</sup>	0.032	6/12
13	+	-	male	age	0.03	9/12
14	+	-	male	age <sup>2</sup>	0.03	11/12
15	+	-	married	past income	0.026	9/17
16	-	+	height	married	0.025	7/17
17	-	+	age	born Tongatapu	0.024	12/13
18	-	+	male $\times$ age	born Tongatapu	0.021	3/12
19	-	+	height	education	0.019	3/17
20	-	+	height	education <sup>3</sup>	0.014	2/17
21	-	+	height	born Tongatapu	0.01	8/17
22	+	-	born Tongatapu	past income	0.007	7/17
23	+	-	age <sup>4</sup>	past income	0.006	4/13
24	-	+	height	age $\times$ education	0.004	9/17
25	+	-	born Tongatapu	male $\times$ age	0.003	6/17
26	-	+	age	male $\times$ height	0.002	12/13
27	+	-	education	male $\times$ age	0.001	3/13
28	+	-	education <sup>3</sup>	male $\times$ age	0.001	3/13

The column  $\Delta$ Bias/ATT is the increase in bias after adjusting on  $X$  as a fraction of the experimental average treatment effect. The last column gives the fraction of control variables that had an increase in the p-value of an equality of means test between treated and control units after controlling for  $X$ .  
*age* = age in years; *born Tongatapu* = 1 if born in Tongatapu, 0 otherwise; *education* = years of education; *male* = 1 if male, 0 otherwise; *married* = 1 if married, 0 otherwise; *past income* = past income in New Zealand dollars per week.

Unfortunately, sensitivity analysis, which goes back to Cornfield et al. (1959) and was further developed by Rosenbaum and Rubin (1983b) and Rosenbaum (1988), cannot tell us whether an unobserved variable exists, and it cannot tell us whether that variable has countervailing effects. Sensitivity analysis addresses a different, but related, question: how large an effect must an unobserved covariate have before it changes our treatment estimate. The distinction is subtle. The question sensitivity analysis addresses is not whether an unobserved covariate exists, but how powerful it would be *if it existed*.

To make this point concrete, we apply sensitivity analysis to the Dehejia and Wahba (1999) specification that comes closest to the LaLonde experimental benchmark (see footnote 3). The method we use works with either regression or propensity scores and comes from Hosman, Hansen and Holland (2010). First, we briefly describe the approach.

Consider a regression

$$y = \alpha + \tau_U T + \mathbf{X}\beta + \beta_U U + \varepsilon,$$

where  $\tau$  is the coefficient of interest,  $\mathbf{X}$  the conditioning set and includes the newly added covariate. As before,  $U$  is unobserved. The researcher, however, can only run

$$y = \alpha + \tau_{no\ U} T + \mathbf{X}\beta + \varepsilon,$$

where the unobserved variable  $U$  is omitted. The sensitivity analysis quantifies how large an effect  $U$  must have, when included in the regression, before the true treatment effect,  $\hat{\tau}$ , changes substantively.<sup>9</sup>

The bias on  $\hat{\tau}$  caused by a possibly omitted variable  $U$  is a function of  $U$ 's confounding with the treatment and  $U$ 's effect on the dependent variable. The Hosman, Hansen and Holland (2010) method generates sensitivity intervals for  $\hat{\tau}$  that are a function of these two effects. Confounding is measured by the  $t$ -statistics from a regression of  $T$  on the other regressors. We denote confoundedness of  $U$  with the treatment of interest as  $t_U$ .  $U$ 's effect on the dependent variable is measured by the proportionate reduction in unexplained variance when  $U$  is included in the regression

$$\rho_{y-u|tx}^2 = \frac{(1 - R_{noU}^2) - (1 - R_U^2)}{(1 - R_{noU}^2)}.$$

Note that neither the  $t$ -statistics nor  $\rho_{y-u|tx}^2$  are used for inferential purposes. Both values simply describe the relationships between the possibly omitted variable  $U$  and either the treatment or the dependent variable.

Hosman, Hansen and Holland (2010) prove that the omitted variable bias can be written as a product of the two effects described above and the standard error on  $\hat{\tau}$

$$\hat{\tau}_{noU} - \hat{\tau}_U = SE(\hat{\tau}_{noU}) t_U \rho_{y-u|tx},$$

provided  $R_{y,tx}^2 < 1$  and  $t_U$  is finite. They go on to prove, under the same conditions, that the same statistics can be used to express the effect of omitting  $U$  on the standard error

$$SE(\hat{\tau}_U) = SE(\hat{\tau}_{noU}) \sqrt{1 + \frac{1 + t_U^2}{df - 1} \sqrt{1 - \rho_{y-u|tx}^2}},$$

where  $df = n - \text{rank}(\mathbf{X}) - 1$ , the residual degrees of freedom after  $Y$  is regressed on  $X$  and the treatment. Taken together, these results allow the specification of a union of interval estimates

$$\{\hat{\tau}_U \pm qSE(\hat{\tau}_U) : |t_U| \leq Z, \rho_{y-u|tx}^2 \leq R\},$$

for any non-negative limits  $Z$  and  $R$ .<sup>10</sup> The union is the collection of  $\hat{\tau}_U$  values falling into the interval after adding the omitted variable  $U$ .

Hosman, Hansen and Holland (2010) suggest choosing values for  $t_U$  and  $\rho_{y-u|tx}^2$  by benchmarking: treating the observed covariates one at a time as being the unobserved covariate  $U$  and collecting values for  $t_U$  and  $\rho_{y-u|tx}^2$  to use as guides. When using propensity scores, they suggest removing the covariates one at a time and then resubclassifying the sample using the modified propensity score.  $T$  is then regressed on the withheld covariate and on the propensity strata to get a  $t$ -statistic for that covariate.

The results of the sensitivity analysis are in Table 3. For each variable, the sensitivity interval comfortably brackets 0, which means that if an unobserved variable existed that had similar-sized effects on the treatment and the outcome as, for example, *married*, its inclusion in the conditioning set would change our beliefs about the treatment effect. Note that we get this result despite using the Dehejia and Wahba's specification. The sensitivity analysis is not telling us that unobserved covariates exist that might change our findings, but that if unobserved

<sup>9</sup> We write of a single omitted variable,  $U$ , for convenience. We can, however, think of  $U$  as being a combination of two or more omitted variables without doing damage to the argument. See Hosman, Hansen and Holland (2010, 861).

<sup>10</sup> See Hosman, Hansen and Holland (2010) for derivations and proofs.

TABLE 3 *95 Percent Sensitivity Intervals With the Unobserved Variable's Treatment Confounding Hypothesized to be No Greater Than the Treatment Confounding of the Variables Deliberately Omitted Below. The decrease in unexplained variance is hypothesized to be no greater than 5 percent*

Variables	5%
Age	-2628.5, 293.3
Education	-2599.7, 382.3
Black	-2458.3, 899.6
Hispanic	-2906.7, 582.4
Married	-3358.8, 1104.7
No degree	-3181.7, 676.0
Income 74	-3622.5, 1985.8
Income 75	-8840.5, 4660.6

covariates exist, and they had effects similar in size to the variables in Table 3, then our findings would be in jeopardy. Sensitivity analysis cannot tell us about countervailing effects because it addresses a different question.

#### DISCUSSION

Researchers working with observational data have to make decisions regarding covariate adjustment. When making these decisions, researchers have to consider the covariates to which they have access as well as the covariates to which they do not. Our analytic results show that if two variables, one observed and one unobserved, have countervailing effects, and we condition on the observed variable, we may increase confounding bias. Our empirical results show, using two data sets, that pairs of variables having countervailing effects are not rare. Finally, we showed that balance tests cannot be used to justify the inclusion of additional covariates and that sensitivity analysis cannot alert us to the presence of countervailing effects. It is possible to increase balance by conditioning on a covariate while at the same time increasing bias. Sensitivity analysis answers a different question.

We have yet to address how researchers can best make use of our findings. Our results indicate that researchers cannot rely on advice such as condition on all pretreatment covariates or on balance and sensitivity tests. Some progress can be made if we consider the two kinds of unobserved covariates that plague empirical analyses. To paraphrase Donald Rumsfeld (Morris 2010), there are known unknowns and unknown unknowns. That is, there are covariates, perhaps suggested by theory, that cannot be measured or perhaps measurement is infeasible. These are the known unknown covariates. A researcher can hypothesize about the relationships of such a covariate with previously included variables and any variables that are candidates for inclusion. Our results provide some guidance in such a situation. If the candidate covariate and the unobserved covariate have countervailing effects, a case can be made for leaving the candidate covariate unadjusted.

On the other hand, there exist, in Rumsfeldian terms, unknown unknown covariates. These are variables that have not been suggested by theory and have not crossed the mind of the researcher in question (or anyone else). In such a case, no theorizing can take place, and our

results demonstrate that including a new covariate in a conditioning set may increase or decrease the bias on the treatment estimate. Sensitivity analysis that explicitly takes unobserved covariates into account is of little use. The only surefire response a researcher has to the problem discussed in this paper is to be modest in the claims she makes based on her results. Scientific progress is rarely the result of a single study, and empirical generalizations are accepted only after many repeated demonstrations across varying spatial and temporal domains.

## REFERENCES

- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. 'Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment'. *Political Analysis* 14 (1):37–62.
- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2010. 'A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark'. *Sociological Methods and Research* 39(2):256–82.
- Bhattacharya, Jay, and William B. Vogt. 2007. 'Do Instrumental Variables Belong in Propensity Scores?'. Technical Report No. 343, NBER, MA.
- Clarke, Kevin A. 2005. 'The Phantom Menace: Omitted Variable Bias in Econometric Research'. *Conflict Management and Peace Science* 22(4):341–52.
- Clarke, Kevin A. 2009. 'Return of the Phantom Menace: Omitted Variable Bias in Econometric Research'. *Conflict Management and Peace Science* 26(1):46–66.
- Cornfield, Jerome, William Haenszel, E. Cuyler Hammond, Abraham M. Lilienfeld, Michael B. Shimkin, and Ernst L. Wynder. 1959. 'Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions'. *Journal of the National Cancer Institute* 22(1):173–203.
- Dehejia, Rajeev H., and Sadek Wahba. 1999. 'Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs'. *Journal of the American Statistical Association* 94 (448):1053–62.
- DeLuca, Giuseppe, Jan R. Magnus, and Franco Peracchi. 2015. 'On the Ambiguous Consequences of Omitting Variables'. Technical report, Einaudi Institute for Economics and Finance Rome, Italy.
- Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. 'Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference'. *Political Analysis* 15(3):199–236.
- Hosman, Carrie A., Ben B. Hansen, and Paul W. Holland. 2010. 'The Sensitivity of Linear Regression Coefficients' Confidence Limits to the Omission of a Confounder'. *The Annals of Applied Statistics* 4(2):849–70.
- Iacus, Stefano M. 2007. 'rrp: Random Recursive Partitioning'. R package version 0.7. Available at <http://cran.r-project.org/web/packages/rrp/>.
- LaLonde, Robert. 1986. 'Evaluating the Econometric Evaluations of Training Programs With Experimental Data'. *American Economic Review* 76(4):604–20.
- Mackenzie, David, John Gibson, and Steven Stillman. 2010. 'How Important is Selection? Experimental Vs Non-Experimental Measures of Income Gains from Migration'. *Journal of the European Economic Association* 8(4):913–45.
- Morris, Errol. 2010. 'The Anosognosic's Dilemma: Something's Wrong But You'll Never Know What It Is (Part 1)'. The New York Times. Available at [http://opinionator.blogs.nytimes.com/2010/06/20/the-anosognosics-dilemma-1/?\\_r=0](http://opinionator.blogs.nytimes.com/2010/06/20/the-anosognosics-dilemma-1/?_r=0), accessed 20 June 2010.
- Pearl, Judea. 2010. 'On a Class of Bias-Amplifying Variables That Endanger Effect Estimates'. In Peter Grunwald and Peter Spirtes (eds), Proceedings of the Twenty-Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-10), 417–24. Corvallis, OR: AUAI Press.
- Pearl, Judea. 2011. 'Invited Commentary: Understanding Bias Amplification'. *American Journal of Epidemiology* 174(11):1223–227.
- Rosenbaum, Paul R. 1988. 'Sensitivity Analysis for Matching With Multiple Controls'. *Biometrika* 75(3):577–81.

- Rosenbaum, Paul R., and Donald B. Rubin. 1983a. 'The Central Role of the Propensity Score in Observational Studies for Causal Effect'. *Biometrika* 70(1):41–55.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983b. 'Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study With Binary Outcome'. *Journal of the Royal Statistical Society, Series B* 45(2):212–18.
- Rubin, Donald B. 2009. 'Should Observational Studies be Designed to Allow Lack of Balance in Covariate Distributions Across Treatment Groups?'. *Statistics in Medicine* 28(9):1420–423.
- White, Halbert, and Xun Lu. 2011. 'Causal Diagrams for Treatment Effect Estimation With Application to Efficient Covariate Selection'. *The Review of Economics and Statistics* 93(4):1453–459.
- Wooldridge, J. 2009. 'Should Instrumental Variables be Used as Matching Variables?'. Technical report, Michigan State University, Lansing, MI.