University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Economics Department Working Paper Series

Economics

2021

Bias-Adjusted Treatment Effects Under Equal Selection

Deepankar Basu

Bias-Adjusted Treatment Effects Under Equal Selection

Deepankar Basu[*]

February 28, 2021

Abstract

In a recent contribution, Oster (2019) has proposed a method to generate bounds on treatment effects in the presence of unobservable confounders. The method can only be implemented if a crucial problem of non-uniqueness is addressed. In this paper I demonstrate that one of the proposed methods to address non-uniqueness that relies on computing bias-adjusted treatment effects under the assumption of equal selection on observables and unobservables, is problematic on several counts. First, additional assumptions, which cannot be justified on theoretical grounds, are needed to ensure a unique solution; second, the method will not work when estimate of the treatment effect declines with the addition of controls; and third, the solution, and therefore conclusions about bias, can change dramatically if we deviate from equal selection even by a small magnitude.
Keywords: treatment effect, omitted variable bias.
JEL Codes: C21.

1 Introduction

Researchers are often interested in estimating treatment effect in models where there are clear problems of unobserved or unobservable confounders.

[*]Department of Economics, University of Massachusetts Amherst. Email: dbasu@econs.umass.edu

To fix ideas, consider the following model,

$$Y = \beta X + \Psi\omega^0 + W_2 + \varepsilon, \tag{1}$$

where Y is the outcome variable, X is the treatment variable of interest, ω^0 is a $J \times 1$ vector of observed controls, and W_2 is an unobserved confounder. Suppose a researcher is interested in consistently estimating β, but is unable to do so because of the presence of the unobservable confounder, W_2 (which can be thought of as an index of a set of unobservable variables), int his hypothetical 'long' regression model.

Faced with this problem, researchers often compare the ordinary least square (OLS) estimate of β between a 'short' and an 'intermediate' regression, where the short regression is given by

$$Y = \bar{\beta} X + \bar{\varepsilon}, \tag{2}$$

in which both the observable and unobservable controls, i.e. ω^0 and W_2, are missing from the model, and the intermediate regression is given by

$$Y = \tilde{\beta} X + \tilde{\Psi}\omega^0 + \tilde{\varepsilon}, \tag{3}$$

in which only the unobservable control, W_2, is missing from the model. If the numerical magnitude of $\tilde{\beta}$ and $\bar{\beta}$ are roughly similar, i.e. the estimate of the treatment effect is 'stable', researchers conclude that the bias from the omitted, unobservable confounder is small.

In a recent, innovative contribution, Oster (2019) has demonstrated that such 'coefficient stability' arguments to deal with possible omitted variable bias is misleading.[1] In fact, what is needed to draw conclusions about the magnitude of possible bias due to the unobservable confounder is not the raw change in the estimate of the treatment effect, but an R-squared scaled change in the estimate of the treatment effect between the short and intermediate regressions. This becomes clear when we write the expression for the omitted variable bias in the OLS estimate of the treatment effect in the intermediate regression in terms of the R-squared in the short, intermediate and long regressions, and relevant coefficients in the long regression. A little algebraic manipulation generates a cubic equation in the bias (of the OLS estimate of the treatment effect in the intermediate regression).

A cubic equation with real coefficients will have either one or three real roots. When the cubic equation has a unique real root, the researcher is able to identify the bias, and hence the bias-adjusted treatment effect, without any ambiguity. When the cubic equation has three real roots, the researcher is confronted with the problem of non-uniqueness. She will need a reliable, theoretically valid method to choose between the multiple solutions. If there is no way of choosing between the multiple solutions, then the proposed methodology will not work.

Oster (2019, pp. 193) is aware of this issue and proposes two approaches to deal with the problem of non-uniqueness. To understand her proposals, let us denote by δ, a measure of proportional selection, i.e. a suitably defined ratio of the importance of the unobservable confounder and the observed controls in explaining the variation of the treatment variable. Oster (2019)

[1] Oster (2019) extends previous work by Altonji et al. (2005).

argues that there are two scenarios under which we will get unique solutions. She recommends researchers to use either of these methods to pin down the unique solution and make informed statements about the magnitude of omitted variable bias.

The first method involves computing the bias-adjusted treatment effect under the twin assumptions of $\delta = 1$ (equal selection on observables and unobservables) and a sign restriction (which is stated as Assumption 3 in her paper). In this case, Oster (2019, pp. 194) argues, we can arrive at a unique solution for the bias in the treatment effect and can therefore compute a unique bias-adjusted treatment effect. In actual examples, Oster (2019) shows how this can be useful for putting bounds on the 'true' treatment effect. For instance, if moving from the short to the intermediate regression, a researcher notes that the estimate of the treatment effect moves towards zero, then an interesting question would be to see if the interval formed by the estimate from the intermediate regression and the bias-adjusted estimate of the treatment effect includes zero. If it does, then that would raise questions about any research that reports a non-zero treatment effect.

The second method relies on choosing some value of R_{max} (the magnitude of R-squared in the hypothetical long regression), and calculating the magnitude of δ, i.e. proportion of selection due to unobservables, that would be consistent with $\beta = 0$ (no treatment effect). In this case, Oster (2019) shows that we are able to find a unique magnitude of proportional selection that would make the treatment effect vanish. For instance, if the computed magnitude of δ is 2, it means that a reported non-zero treatment effect would, in reality, be zero only if the unobservable confounder was twice as important as the observed controls in explaining the variation in the treatment variable. In many cases, it might be possible to rule out such large effects for unobservable confounders on intuitive grounds and thus assert the robustness of the reported results.

Both these methods promise to be enormously useful for applied researchers because they provide workable solutions for the pervasive and rather intractable problem of omitted variable bias (Basu, 2020). That is why the method proposed by Oster (2019) has been widely noted in economics and the social sciences.[2] Unfortunately, as I demonstrate in this note, the first method to deal with non-uniqueness, i.e. computation of bias-adjusted treatment effect under the assumption of equi-proportional selection, $\delta = 1$, is fraught with serious problems. First, without additional assumptions, it is not possible to ensure the existence of a unique solution. But these assumptions cannot be justified either on theoretical or empirical grounds. Second, once these assumptions are imposed, there is no leeway for researchers to experiment with different values of R_{max} because a specific value of R_{max} gets pinned down. Third, the method will not work for cases where estimates of the treatment effect declines with the addition of control variables. Finally, there is a sharp discontinuity at $\delta = 1$, i.e. conclusions can change dramatically if δ is perturbed even slightly from the value of unity.

Given these problems, my conclusion is that if researchers wish to use

[2]On Professor Oster's google scholar page, the paper shows 1611 citations. Here are just a few examples: Galor and Ozak (2016); Michalopoulos and Papaioannou (2016); Goldsmith-Pinkham et al. (2019); Jaschke and Keita (2021). Papers published before 2019 cite different working paper versions of Oster (2019).

the methodology proposed by Oster (2019) to address the problem of bias in treatment effect they should avoid computing bias-adjusted treatment effect under equal selection. Instead, they should use one of the following alternatives: (a) compute roots of the cubic equation for a plausible range of combinations of R_{max} (R-squared in the hypothetical long regression) and δ (measure of proportional selection between unobservables and observables), keeping δ bounded away from unity, see if a unique real root emerges, and investigate robustness using the set of unique real roots; (b) use the second method proposed in Oster (2019), i.e. compute δ that is consistent with $\beta = 0$ (treatment effect is zero) for a plausible range of values of R_{max}, and argue on intuitive or theoretical grounds about the plausibility of such a δ. When they use this second method, it is important that they do not use $\delta = 1$ as a benchmark.

The rest of the paper is organized as follows. In the next section, I discuss the basic set-up; in the following section I investigate the case of equi-proportional selection; in the final section I conclude with some suggestions for applied researchers who intend to use the innovative methodology proposed by Oster (2019).

2 Basic Set-Up

Consider once again the hypothetical 'long' regression,

$$Y = \beta X + \Psi\omega^0 + W_2 + \varepsilon, \tag{4}$$

and denote by R_{max}, the R-squared from the long regression. Consider the 'short' regression,

$$Y = \bar{\beta}X + \bar{\varepsilon}, \tag{5}$$

and denote as $\bar{R}$, the R-squared from the short regression. In a similar manner, consider an intermediate regression,

$$Y = \tilde{\beta}X + \tilde{\Psi}\omega^0 + \tilde{\varepsilon}, \tag{6}$$

and denote by $\tilde{R}$, the R-squared from the intermediate regression. Note that $R_{max} \geq \tilde{R} \geq \bar{R}$ (Greene, 2012, Theorem 3.6, pp. 42). Finally, consider an auxiliary regression,

$$X = \alpha\omega^0 + u, \tag{7}$$

and denote by $\tilde{X}$, the residual from this auxiliary regression. Let $\hat{\tau}_X$ denote the variance of $\tilde{X}$, σ_X^2 denote the variance of X and σ_Y^2 denote the variance of Y, and note that $\sigma_X^2 > \tau_X$. Following Oster (2019), let us define the measure of proportional selection as,

$$\delta = \frac{\sigma_{2X}/\sigma_2^2}{\sigma_{1X}/\sigma_1^2} \tag{8}$$

where $\sigma_{1X} = cov\,(W_1, X)$, $\sigma_{2X} = cov\,(W_2, X)$, $\sigma_1^2 = var\,(W_1)$, and $\sigma_2^2 = var(W_2)$, and $W_1 = \Psi\omega^0$ (an index of the observable controls).

Using the well-known formula for omitted variable bias in the short and intermediate regressions, we can write expressions for the asymptotic bias in $\bar{\beta}$ and $\tilde{\beta}$ in terms of coefficients in the hypothetical long regression and coefficients in several auxiliary regressions. The innovation introduced by Oster (2019) is to rewrite the expression for the bias in the intermediate regression using the R-squared in the short, intermediate and long regressions. A little algebraic manipulation then generates a system of 3 equations in 3 unknowns: σ_1^2, the variance of W_1; σ_{1X}, the covariance of W_1 and X (treatment variable); and ν (the bias of the treatment effect in the intermediate regression). Finally, we can reduce the three equations into a single cubic equation in ν given by,

$$a\nu^3 + b\nu^2 + c\nu + d = 0, \tag{9}$$

where

$$a = (\delta - 1)\left(\tau_X \sigma_X^2 - \tau_X^2\right) \tag{10}$$

$$b = \tau_X \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2 (\delta - 2) \tag{11}$$

$$c = \delta \left(R_{max} - \tilde{R}\right) \sigma_Y^2 \left(\sigma_X^2 - \tau_X\right) - \left(\tilde{R} - \bar{R}\right) \sigma_Y^2 \tau_X - \sigma_X^2 \tau_X \left(\bar{\beta} - \tilde{\beta}\right)^2 \tag{12}$$

$$d = \delta \left(R_{max} - \tilde{R}\right) \sigma_Y^2 \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2 \tag{13}$$

The cubic equation in (9) will have either one real root or three real roots.[3] In the case when there is only one real root, denote it by ν_1. If $\beta^* = \tilde{\beta} - \nu_1$, then $\beta^* \xrightarrow{p} \beta$, so that ν_1 is the asymptotic bias in the treatment effect estimated by the intermediate regression. Hence, $\tilde{\beta} - \nu_1$ is the bias-

[3] In any specific analysis, both R_{max} and δ are unobserved. All other parameters that determine the coefficients of the cubic will come from the output of the short and intermediate regressions. Hence, researchers will have to choose specific values of R_{max} and δ to compute the roots of the cubic.

adjusted treatment effect (which will converge in probability to the true treatment effect, as proved in Proposition 2 in Oster (2019)). In the case when the cubic equation has three real roots, ν_1, ν_2, ν_3, then only one of these will give us the asymptotic bias in the treatment effect. Hence, only one of the following, $\tilde{\beta} - \nu_1$, $\tilde{\beta} - \nu_2$ and $\tilde{\beta} - \nu_2$, will be the bias-adjusted treatment effect. Without more information or assumptions, a researcher will not be able to unambiguously find the bias-adjusted treatment effect or develop a meaningful bounding argument (because there will be multiple bounding sets to choose from and the union of all these sets is likely to be too large to be informative).

Oster (2019, pp. 194) proposes two approaches to deal with the problem of non-uniqueness. First, if we calculate the magnitude of δ that would be consistent with $\beta = 0$, we would be able to find a unique magnitude of proportional selection that would make the treatment effect vanish. This result is proved in Proposition 3 in Oster (2019). The second approach asks us to compute the bias-adjusted treatment effect, i.e. $\beta^* = \tilde{\beta} - \nu_1$, under the assumption that $\delta = 1$ and the additional sign restriction that,

$$\operatorname{sign}\left(cov\left(X, \hat{W}_1\right)\right) = \operatorname{sign}\left(cov\left(X, W_1\right)\right), \tag{14}$$

where $\hat{W}_1$ is the estimated value of W_1. Oster (2019, pp. 194) argues that we will arrive at a unique solution following this second approach, i.e. there will be a unique solution ν_1, so that $\tilde{\beta} - \nu_1$ will be the unique bias-adjusted treatment effect. The paper does not offer a proof of this important claim. So, let us investigate it in detail.

3 Equi-Proportional Selection

3.1 What is the Solution?

When there is equal selection on observables and unobservables, we will have $\delta = 1$. What will be the solution for bias under equal selection? If we impose the restriction that $\delta = 1$ on the coefficients of the cubic in (9) we get,

$$
\begin{aligned}
a &= 0 \\
b &= -\tau_X \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2 \\
c &= \left(R_{max} - \tilde{R}\right) \sigma_Y^2 \left(\sigma_X^2 - \tau_X\right) - \left(\tilde{R} - \bar{R}\right) \sigma_Y^2 \tau_X - \sigma_X^2 \tau_X \left(\bar{\beta} - \tilde{\beta}\right)^2 \\
d &= \left(R_{max} - \tilde{R}\right) \sigma_Y^2 \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2,
\end{aligned}
$$

which converts the cubic in (9) to a quadratic equation in ν,

$$b_1 \nu^2 + c_1 \nu + d_1 = 0, \tag{15}$$

where the coefficients of this quadratic are given by,

$$b_1 = -\tau_X \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2 \tag{16}$$

$$c_1 = \left(R_{max} - \tilde{R}\right) \sigma_Y^2 \left(\sigma_X^2 - \tau_X\right) - \left(\tilde{R} - \bar{R}\right) \sigma_Y^2 \tau_X - \sigma_X^2 \tau_X \left(\bar{\beta} - \tilde{\beta}\right)^2 \tag{17}$$

$$d_1 = \left(R_{max} - \tilde{R}\right) \sigma_Y^2 \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2. \tag{18}$$

The solutions of the quadratic in (15) are given by

$$\nu = \frac{-c_1 \pm \sqrt{c_1^2 - 4d_1 b_1}}{2b_1},$$

which are noted in Corollary 1 in Oster (2019, pp. 193). Our first result is that the solution of the quadratic equation in (15) is always real.

Proposition 1. *The quadratic equation in (15) either has a unique real root or two distinct real roots. It does not have any complex roots.*

Proof. The proof follows by noting that the discriminant of this quadratic equation is non-negative, i.e. $c_1^2 - 4d_1 b_1 \geq 0$, because $c_1^2 \geq 0$, and

$$\begin{aligned}
-4d_1 b_1 &= -4 \left\{ \left(R_{max} - \tilde{R} \right) \sigma_Y^2 \left(\bar{\beta} - \tilde{\beta} \right) \sigma_X^2 \right\} \left\{ -\tau_X \left(\bar{\beta} - \tilde{\beta} \right) \sigma_X^2 \right\} \\
&= 4 \left(R_{max} - \tilde{R} \right) \sigma_X^4 \sigma_Y^2 \tau_X \left(\bar{\beta} - \tilde{\beta} \right)^2 \\
&\geq 0
\end{aligned}$$

where the last inequality follows because $R_{max} \geq \tilde{R}$. □

The implication of this result is that, in general, there will be two real roots of the quadratic equation in (15). Hence, in general, there will be two values of the bias in the treatment effect, and hence two values of the bias-adjusted treatment effect, when there is equal selection on observables and unobservables. Without further assumptions, it is not possible to arrive at a unique solution for the bias or the bias-adjusted treatment effect. So, we need to investigate the following question: what conditions are necessary to give us an unique solution for the quadratic in (15)? The unique root will arise if and only if the discriminant of the quadratic equation is identically equal to zero. I now show that the discriminant can be zero only if we impose additional assumptions. These assumptions are difficult to justify on either theoretical or empirical grounds, and therefore raise questions about the logic of the bounding argument.

3.2 Condition for Unique Solution

For the quadratic equation in (15) to have a unique real solution, the discriminant must be zero, i.e.,

$$\left\{\left(R_{max}-\tilde{R}\right)\sigma_Y^2\left(\sigma_X^2-\tau_X\right)-\left(\tilde{R}-\bar{R}\right)\sigma_Y^2\tau_X-\sigma_X^2\tau_X\left(\bar{\beta}-\tilde{\beta}\right)^2\right\}^2 + 4\left(R_{max}-\tilde{R}\right)\sigma_Y^4\tau_X\left(\bar{\beta}-\tilde{\beta}\right)^2=0.$$

Defining $Z = R_{max} - \tilde{R}$, we can write the above condition as a quadratic equation in Z,

$$A^2Z^2+\left(2AB+4C\right)Z+B^2=0, \tag{19}$$

where

$$A=\sigma_Y^2\left(\sigma_X^2-\tau_X\right)>0 \tag{20}$$

$$B=-\left[\left(\tilde{R}-\bar{R}\right)\sigma_X^2\tau_X+\sigma_X^2\tau_X\left(\bar{\beta}-\tilde{\beta}\right)^2\right]<0 \tag{21}$$

$$C=\sigma_Y^2\left(\bar{\beta}-\tilde{\beta}\right)\sigma_X^2. \tag{22}$$

The two roots of (19) are given by

$$Z_1, Z_2 = \frac{-\left(2AB + 4C\right) \pm \sqrt{\left(2AB + 4C\right)^2 - 4A^2B^2}}{2A^2}. \tag{23}$$

Note that the discriminant of the quadratic equation in (19) reduces to $16C^2 + 16ABC$. Since $B < 0$, it is possible, though not necessary, for the discriminant, $16C^2 + 16ABC$, to be negative.[4] Hence, there are two cases to consider.

Case 1. If the discriminant is negative, then both the roots of (19), Z_1, Z_2, are complex numbers. In this case, the uniqueness analysis falls through. This is because it is meaningless to entertain the possibility that $Z = R_{max} - \tilde{R}$ is a complex number. What does this mean? Since $\tilde{R}$ is a known real number, this implies that there is no real value of R_{max} that would make the discriminant of the quadratic equation in (15) to be zero. Hence, in this case, there does not exist a unique magnitude of the bias in the treatment effect, ν, and hence, it is not possible to find a unique bias-adjusted treatment effect, β^*.

Case 2. If the discriminant is nonnegative, then both the roots of (19), Z_1, Z_2, are real. Thus, there exists real values of R_{max} which would give a unique value of the bias, and hence, the bias-adjusted treatment effect. But not all possible values of R_{max} are permissible. We know that R_{max} is never smaller than $\bar{R}$. Hence, we need necessary conditions to ensure that the solutions of (19) are nonnegative. This is given in

[4]Since $\tilde{R} \geq \bar{R}$, the term in the square bracket in the definition of B is positive. Hence, $B < 0$.

Proposition 2. *If $\bar{\beta} < \tilde{\beta}$, then then both roots of (19) are real. One of these roots will be nonnegative only if*

$$\sigma_Y^2 \left(\sigma_X^2 - \tau_X\right) \left[\left(\tilde{R} - \bar{R}\right) \sigma_X^2 \tau_X + \sigma_X^2 \tau_X \left(\bar{\beta} - \tilde{\beta}\right)^2\right] + 2\sigma_Y^2 \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2 \leq 0. \tag{24}$$

Proof. To see the first part, note that if $\bar{\beta} < \tilde{\beta}$, then $C < 0$. Hence $C^2 + ABC \geq 0$. Hence, the discriminant of (19) is nonnegative. To see the second part, note that since the denominator of the expression for the roots in (23) is always positive, the sign of the roots are the same as the sign of the numerator. If $2AB + 4C > 0$, then the numerator is negative because the expression within the square root in (23) is nonnegative and less than $2AB + 4C$. Hence, we have the following: $2AB + 4C > 0 \implies Z_1 < 0$ and $Z_2 < 0$. The contrapositive of this statement gives us: $Z_1 \geq 0$ or $Z_2 \geq 0 \implies 2AB + 4C \leq 0$. Hence, $2AB + 4C \leq 0$ is the necessary condition for at least one root being nonnegative. Plugging the expression for A, B and C, this becomes

$$\sigma_Y^2 \left(\sigma_X^2 - \tau_X\right) \left[\left(\tilde{R} - \bar{R}\right) \sigma_X^2 \tau_X + \sigma_X^2 \tau_X \left(\bar{\beta} - \tilde{\beta}\right)^2\right] + 2\sigma_Y^2 \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2 \leq 0,$$

which is the expression in (24). $\square$

The above analysis has important implications. If $\delta = 1$, i.e. there is equal selection on observables and unobservables, and the condition in (24) is satisfied, then either $R_{max} - \tilde{R}$ will be given by the positive root of (19) when one of the two roots is negative, or $R_{max} - \tilde{R}$ will attain two positive values given by both the roots of (19) when both roots are positive. In either case, once we choose to impose the restriction that $\delta = 1$, then there will either be a unique value of R_{max} or two possible values of R_{max} that are permissible. The choice of $\delta = 1$ implies these *specific* values of R_{max}. Researchers are no longer at liberty to choose any other value of R_{max}.

This raises questions about some of the examples discussed in Oster (2019). For instance, in section 4.2, the bounding argument about treatment effects is explained through a discussion of the impact of maternal behaviour on child outcomes. The results reported in Table 3 include, in column 5, computations of $\beta^* (R_{max}, 1)$, where R_{max} is chosen as 0.61 in panel A and as 0.53 in panel B. These specific values of R_{max} are chosen from existing studies that have reported regressions with sibling fixed effects when investigating the effect of maternal behaviour on child outcomes. In a similar manner, the discussion in section 5 computes bias-adjusted treatment effects for $\delta = 1$ under different choices of R_{max}. For instance, the results reported in Table 5 use two different values of R_{max}. In column 3, $R_{max} = \tilde{R} + \left(\tilde{R} - \bar{R}\right)$, and in column 4, $R_{max} = 1.3\tilde{R}$.

The results reported in this paper in proposition 1 and proposition 2 show that researchers cannot choose values for R_{max} arbitrarily once they choose to fix the value of δ at 1. Meaningful values of R_{max} can only arise from using the positive, real roots of the quadratic equation in (15). It is not clear that the choice of R_{max} used by Oster (2019) respects this restriction. This raises questions about the conclusions of the paper.

There is an additional angle to consider with regard to the analysis of bias under the assumption of equal selection and this is highlighted in the next result.

Corollary 1. *If the estimate of the treatment effect declines with the addition of controls, i.e. if $\bar{\beta} > \tilde{\beta}$, then a meaningful bias-adjusted treatment effect cannot be computed under the assumption of equal selection.*

Proof. Note, first, that if $\bar{\beta} > \tilde{\beta}$, it is no longer guaranteed that (19) will have real roots. Consider, further, the necessary condition in (24) and note that if $\bar{\beta} > \tilde{\beta}$, then the condition cannot be satisfied. This is because the second term

$$2\sigma_Y^2 \left(\bar{\beta} - \tilde{\beta}\right) \sigma_X^2$$

is positive. Since $\sigma_X^2 > \tau_X$ and $R_{max} \geq \tilde{R}$, the first term

$$\sigma_Y^2 \left(\sigma_X^2 - \tau_X\right) \left[\left(\tilde{R} - \bar{R}\right) \sigma_X^2 \tau_X + \sigma_X^2 \tau_X \left(\bar{\beta} - \tilde{\beta}\right)^2\right]$$

is always positive. Hence the expression on the left hand side of the condition in (24) is positive. An application of proposition 2 then shows that the quadratic equation in (19) cannot have meaningful roots. This, in turn, implies that the discriminant of the quadratic equation in (15) cannot be zero. This implies that the quadratic equation in (15) cannot have a unique root. Hence, a meaningful bias-adjusted treatment effect cannot be computed. □

The implication of this corollary is that other than in cases where the treatment effect falls with the addition of controls, i.e. $\bar{\beta} > \tilde{\beta}$, the uniqueness of the solution is impossible. Thus, if for some research it is found that the treatment effect decreases with the addition of controls, for instance from 0.75 to 0.15, then we can be sure that for this particular research a unique bias-adjusted treatment effect cannot be computed under the assumption of equal selection. It is interesting to note that in all the cases reported in Table 5 in Oster (2019, pp. 202), $\bar{\beta} < \tilde{\beta}$. That is why the reported bias-adjusted treatment effects are meaningful. If instead, we had $\bar{\beta} > \tilde{\beta}$, then those results would break down.

3.3 Discontinuity at Equal Selection

There is a further problem in using equal selection on observables and unobservables that I would like to highlight. To see this, let us explicitly write the solutions of the cubic equation in (9) using Cardano's formulas,[5]

$$\begin{aligned}\nu =& \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) + \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} \\ &+ \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) - \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}} \\ &- \frac{b}{3a},\end{aligned} \tag{25}$$

where the expressions for a, b, c and d are given in (10), (11), (12) and (13). Returning to the quadratic equation in (15), we note that its solutions are given by

$$\nu = \frac{-c_1 \pm \sqrt{c_1^2 - 4d_1 b_1}}{2b_1}, \tag{26}$$

where b_1, c_1 and d_1 are given by (16), (17), and (18).

The roots in (25) and (26) are very different because they come from two different equation systems. If a researcher works with the cubic equation

[5] See https://math.vanderbilt.edu/schectex/courses/cubic/

and uses values of δ close to unity the roots she finds will be very different from those that will arise if she works with the quadratic equation, where $\delta = 1$. Both the magnitude and sign of the real roots can be different when the researcher perturbs δ by a very small magnitude starting from a value of unity. Hence there is a discontinuity in the root of the equation system at $\delta = 1$. Since the root of the equation system, whether cubic or quadratic, gives us the treatment bias, researchers will arrive at very different conclusions depending on whether they use $\delta = 1$ or $\delta = 1 \pm \epsilon$, for a small $\epsilon > 0$. This raises serious questions about the usefulness or validity of using the condition of equal selection at all in the analysis, and especially using it as a useful benchmark.

To illustrate this point, let me present an example with the following values of the parameters: $\bar{\beta} = 1.907$, $\tilde{\beta} = 0.964$, $\bar{R} = 0.196$, $\tilde{R} = 0.497$, $\sigma_X^2 = 0.209$, $\tau_X = 0.401$, $\sigma_Y^2 = 3.809$. I choose to use $R_{max} = 0.85$. Using these parameter values, when I calculate the roots for the case of $\delta = 1$, I get two real roots: -22.137 and 2.414. Using the same parameter values, when I calculate the roots for the case of $\delta = 1.01$, I get one real root, -1.406, and two complex roots, $2.810 - 10.713i$ and $2.810 + 10.713i$. Using the same parameter values, once again, when I calculate the roots for the case of $\delta = 0.99$, I get three real roots, -62.834, 16.987 and -1.454. This shows how moving from $\delta = 0.99$ to $\delta = 1.00$ to $\delta = 1.01$ change the conclusions about the magnitude and sign of bias dramatically, raising serious questions about the robustness of the procedure.

4 Concluding Comments

There are four points to take away from the analysis in this paper. First, the claim in Oster (2019, pp. 194) that the sign restriction in (14) and the assumption of $\delta = 1$ together give a unique solution for the bias in the treatment effect is unlikely to be true. To generate a unique solution under the assumption of $\delta = 1$ requires a precise quantitative relationship between R_{max} and a host of other parameters like $\bar{R}$, $\tilde{R}$, σ_X^2, τ_X, $\bar{\beta}$, and $\tilde{\beta}$. The precise quantitative relationship is given by the roots of (19) and the additional condition captured by (24) that needs to be imposed to ensure a meaningful solution. On its own, a sign restriction like (14) cannot ensure these quantitative relationships. Hence, the claim in Oster (2019) that "... calculating the bias-adjusted effect under the assumption of $\delta = 1$, with Assumption 3 active ... will provide a unique solution" seems to be false.[6]

The second point to note is that under the assumption of $\delta = 1$, there will, in general, be two real magnitudes of the bias in the treatment effect, as demonstrated in Proposition 1. Since there is no unique magnitude of the bias-adjusted treatment effect, researchers cannot construct unique bounding intervals and investigate whether zero is contained in the bounding interval. If a researcher wants to generate a unique magnitude for the bias of the treatment effect under the assumption of equal selection, i.e. $\delta = 1$, then additional assumptions will need to be imposed. But once these assumptions are imposed they imply very specific values of R_{max}, viz. those that are captured by the roots of (19), assuming that they are real

[6] In Oster (2019), Assumption 3 refers to what we have expressed in this paper as the assumption in (14).

and positive (captured by the condition in proposition 2). There are no theoretical or empirical reasons to assume that the choice of R_{max} by any researcher in any specific analysis would satisfy this specific condition. More importantly, researchers are not at liberty to choose other values of R_{max} to conduct robustness analyses. This seriously restricts the usefulness of the bounds argument if we impose the condition of equal selection on unobservables and observable, i.e. $\delta = 1$.

The third point to note is that the method, with all its restrictive features, will work when the estimate of the treatment effect declines with the addition of controls to the regression model. This is because if $\bar{\beta} > \tilde{\beta}$, then it is no longer possible to compute a meaningful and unique magnitude of bias (see corollary 1). This rules out the usefulness of this method for a large subset of existing and future studies. The final point to note is that there is a sharp discontinuity at $\delta = 1$. This means that the conclusions of the analysis can change dramatically even when researchers change δ by a very small magnitude around the value of $\delta = 1$. The drastic change can manifest itself either in a change in the magnitude of the bias or in a reversal of sign of the bias. Taken together, the second, third and fourth points suggest that researchers should not use equal selection on observables and unobservables, i.e. $\delta = 1$, when studying the problem of bias in treatment effects due to unobserved confounders.

The upshot is that when researchers use the methodology proposed by Oster (2019) to deal with omitted variable bias, they should not use the method of computing bias-adjusted treatment effects under the assumption of equi-proportional selection, i.e. $\delta = 1$. The theoretical justification for this method is weak and it is likely to lead to misleading conclusions. Instead, applied researchers can do one of the following two things. First, they can compute the roots of the cubic equation in (9) for a range of plausible combinations of R_{max} *and* δ and see if they arrive at unique real roots. If they do, then they will be able to compute unique bias-adjusted treatment effects for each such combination. If the results do not change qualitatively across these combinations, e.g. if zero is excluded from all the bounding sets, then their results are robust to the omitted variable bias. Second, if they do not get unique real roots, they might instead use the method of computing δ (magnitude of proportional selection), as proposed by Oster (2019), that makes the treatment effect zero for a range of plausible values of R_{max} and argue why or why not such a δ is unrealistic. When using this second method, it is necessary for researchers to avoid any reference to the case of $\delta = 1$ as a benchmark. It is not a meaningful benchmark to use.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *The Journal of Human Resources*, 40(4):791–821.

Basu, D. (2020). Bias of OLS Estimators due to Exclusion of Relevant Variables and Inclusion of Irrelevant Variables. *Oxford Bulletin of Economics and Statistics*, 82(1):209–234.

Galor, O. and Ozak, O. (2016). The Agricultural Origins of Time Preference.

American Economic Review, 106(10):2586–2624.

Greene, W. H. (2012). *Econometric Analysis*. Prentice Hall.

Jaschke, P. and Keita, S. (2021). Say it like Goethe: Language learning facilities abroad and the self-selection of immigrants. *Journal of Development Economics*, 149:102597.

Michalopoulos, S. and Papaioannou, E. (2016). The Long-Run Effects of the Scramble for Africa. *American Economic Review*, 106(7):1802–1848.

Oster, E. (2019). Unobservable Selection and Coefficient Stability. *Journal of Business and Economic Statistics*, 37(2):187–204.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2019). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.