Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities

Macartan Humphreys[*]
Columbia University
mh2245@columbia.edu

December 8, 2009

Abstract

In many contexts, treatment assignment probabilities differ across strata or are correlated with some observable third variables. Regression with covariate adjustment is often used to account for these features. It is known however that in the presence of heterogeneous treatment effects this approach does not yield unbiased estimates of average treatment effects. But it is not well known how estimates generated in this way diverge from unbiased estimates of average treatment effects. Here we show that biases can be large, even in large samples. However we also find conditions under which the usual approach provides interpretable estimates and we identify a monotonicity condition that ensures that least squares estimates lie between estimates of the average treatment effects for the treated and the average treatment effects for the controls. The monotonicity condition can be satisfied for example with Roy-type selection and is guaranteed in the two stratum case.

[*] My thanks to Craig McIntosh, Andy Gelman, Joshua Angrist, Kosuke Imai, Laura Paler, Neelan Sircar, and Guido Imbens for generous comments on this paper.

1 Introduction

Consider a treatment in which a technology is distributed randomly to children and one is interested in the effects of the intervention on their parents. Randomization notwithstanding, simple comparison of outcomes between parents with and without treated children may not yield unbiased estimates of the effect of the technology on parents. The problem is that parents with more children are more likely to be exposed to the technology. If characteristics of strata (here, parents with the same number of children) are related to the outcome of interest, then simple differences between outcomes in treated and control units may yield biased estimates of treatment effects.

More broadly this is an example of a situation in which assignment to the treatment of interest is "ignorable" only conditional upon observables (Rosenbaum and Rubin, 1983).[1] As discussed in section 2, this kind of problem is common and arises both for experimental and observational studies.

In such settings unbiased estimates of treatment effects can be obtained within each stratum; in this case by comparing outcomes among parents with a given number of children. Average treatment effects can then be estimated by averaging such stratum level treatment effects over strata. This is in effect a matching procedure and it can be implemented using a range of available software packages (MatchIt in R (Ho, Imai, King, and Stuart, 2007) and nnmatch (Abadie, Herr, Imbens, and Drukker, 2004) and CEM for R and Stata (Iacus, King, and Porro, 2008)).

In practice, however, many researchers seek to estimate treatment effects using some form of regression and 'conditioning' upon strata. For the spillover problem described above for example, this might be done by regressing the outcome on parental exposure plus a measure of the number of children a parent has, or by using a more flexible procedure in which an indicator variable is entered for each stratum. The approach is almost universal in observational studies in which researchers seek to control for confounds and it is recommended for experimental work in Duflo, Glennerster, and Kremer (2007).

This procedure, though common, can however produce problems if there are heterogeneous effects. It is known that in such cases regression with covariate adjustment produces estimates that can deviate from average treat-

[1] Implicitly we assume that child-to-parent transmission is the only way that parents are exposed.

ment effects (Angrist (1998); see also Freedman (2008)). Less well understood is when these biases arise and how important they are likely to be. Our contribution is to provide an interpretation of the least squares (OLS) estimate, note conditions under which OLS estimates are 'close' to causal quantities of interest, and identify a simple monotonicity condition that ensures that OLS estimates are bounded by such causal quantities of interest. We provide illustrations from real and simulated data that show when and how standard approaches can yield inaccurate or accurate results.

2 Applications

The situation in which assignment to treatment is random only conditional upon stratum arises in a diverse range of experimental and observational settings.

One family of cases, highlighted above, arises when researchers are interested in spillover effects. In this case the treatment of interest is often not the same as the treatment that is directly under the control of the researcher and treatment assignment may be related to covariates. Say for example n individuals are linked through friendship relations and some technology is randomly distributed to k of these. Say that the treatment of interest is second-hand exposure to the technology, which occurs if one or any of one's immediate friends receives the technology. In this case an individual j with s_j friends (s_j small relative to n) is assigned to treatment with probability $p_x = 1 - (\frac{n-k}{n})^{s_j+1}$. The assignment to the treatment of interest is thus correlated with the number of friends one has, but random conditional upon the number of friends. A problem of this form is examined by Miguel and Kremer (2004) and Oster and Thornton (2009).

A second set of applications can arise when the definition of the treatment depends on features of units. Thus for example, in studies in which individuals are randomly matched with partners, the probability that an individual is matched with a same sex or same ethnicity partner depends on the distribution of sexes and ethnicities in a population (Habyarimana, Humphreys, Posner, and Weinstein, 2007). In this case an individual from a group with relative size s_j will encounter an individual from the same group with probability $p_j = s_j$.

Two other sets of applications arise from the manner in which randomization is conducted. First, if two or more treatments are randomly assigned using correlated probabilities then assignment to one treatment may be correlated with assignment to another but each can be random conditional upon the other. A similar logic arises if individuals are assigned to treatment through two or more lotteries with different probabilities associated with each.

A final, and perhaps most common set of applications, occurs in observational data when third factors are plausibly correlated with both an independent variable of interest and the outcome variable and in which researchers claim that they can identify and measure all such variables. An important example is the situation in which individuals self-select into treatment on the basis of expected gains (Roy, 1951). Say for example that each individual in group j expects to gain τ_j from a treatment but faces an individual cost to participating of ϵ_i where ϵ_i is independent and identically distributed for all individuals in the population according to density F. In this case, if individuals select into treatment whenever benefits exceed costs, the probability of treatment for an individual in group j is $p_j = F(\tau_j)$.

In such cases researchers often seek to estimate treatment effects after controlling for third variables or strata. We now describe the quantities that are estimated using such an approach and how they relate to causal quantities of interest.

3 Results

We consider the case in which assignment to a binary treatment is ignorable conditional upon membership in a stratum. Let n_x denote the number of units in stratum x, w_x the share of all units that are in stratum x and $p_x \in (0, 1)$ the share of units in stratum x that receive some binary treatment. Note that with $p_x \in (0, 1)$ we assume that there is "overlap" in the sense that each stratum contains treated and control units. We take the collection (p_x) to be equivalent to the stratum level propensities of assignment to treatment. Employing the potential outcomes framework (Rosenbaum and Rubin, 1983), let y_{ixt} and y_{ixc} denote the value on some outcome variable that unit i in stratum x would take if allocated to treatment and control conditions respectively. The causal effect of the treatment on unit i is given by $\tau_i = y_{ixt} - y_{ixc}$.

Unfortunately τ_i cannot be estimated since only one of y_{ixt} and y_{ixc} is observed. However, under conditions described by Rosenbaum and Rubin

(1983) and others the average treatment effect for units in stratum x, τ_x can be estimated without bias by $\widehat{\tau}_x = \bar{y}_{xt} - \bar{y}_{xc}$, where $\bar{y}_{xt}$ (resp $\bar{y}_{xc}$) is the average value of y for treated (resp. control) units in stratum x; these are well defined under our assumption that there are both treatment and control units in each stratum. Under these conditions unbiased estimates of the average treatment effect ($\widehat{\tau}_{ATE}$), the average treatment effect on the treated ($\widehat{\tau}_{ATT}$), and the average treatment effect on the controls ($\widehat{\tau}_{ATC}$) (sometimes referred to as the ATU) are given by:

$$\widehat{\tau}_{ATE} = \sum_x \frac{w_x}{\sum_j w_j} \widehat{\tau}_x \tag{1}$$

$$\widehat{\tau}_{ATT} = \sum_x \frac{p_x w_x}{\sum_j p_j w_j} \widehat{\tau}_x \tag{2}$$

$$\widehat{\tau}_{ATC} = \sum_x \frac{(1-p_x) w_x}{\sum_j (1-p_j) w_j} \widehat{\tau}_x \tag{3}$$

In each case, these estimates of average treatment effects are weighted averages of estimates of the stratum level treatment effects, $\widehat{\tau}_x$. What differs is the weighting: the $\widehat{\tau}_{ATT}$ places more weight on the treatment effect of strata with many treated units, the $\widehat{\tau}_{ATC}$ places more weight on the treatment effect of strata with few treated units. Clearly if $p_x = p_j$ for all x, j, then $\widehat{\tau}_{ATE} = \widehat{\tau}_{ATT} = \widehat{\tau}_{ATC}$.

3.1 Least squares estimates

Now consider an estimate of treatment effects resulting from a model in which the outcome is regressed on treatment and a set of indicator variables for each of the strata using OLS. In this case OLS also returns a weighted average of the stratum level treatment effects, however, the weights used by OLS reflect the variance in treatment, not the degree of treatment, within each stratum (Angrist, 1998). In particular:

$$b_{OLS} = \sum_x \frac{p_x (1-p_x) w_x}{\sum_j p_j (1-p_j) w_j} \widehat{\tau}_x \tag{4}$$

We see that b_{OLS} can differ from $\widehat{\tau}_{ATT}$, $\widehat{\tau}_{ATC}$ and $\widehat{\tau}_{ATE}$; indeed since OLS weights can take any value between 0 and 1 for any stratum, depending only on the values taken by the collection (p_x), for a given collection of estimated stratum treatment effects ($\widehat{\tau}_x$) and sizes (w_x), b_{OLS} can take any value between $\min(\widehat{\tau}_x)$ and $\max(\widehat{\tau}_x)$.

Thus b_{OLS} does not correspond to estimates of average treatment effects; does this mean that it is incorrect? In fact, as we see from (4), b_{OLS}, like each of the treatment effects in (1)-(3), is a weighted average of estimates of the fundamental causal quantities of interest, the within-stratum treatment effects (τ_x). Each of these averages provides a way to summarize these multiple, possibly heterogeneous, causal quantities, but does not generate new causal information. Determining what type of averaging, if any, is appropriate, is a substantive, not an inferential, question. The $\widehat{\tau}_{ATT}$ reports the historical effects; for a program implementer for example, the ATT provides information on what the effects of an implemented program were. The $\widehat{\tau}_{ATC}$ reports an estimate of potential effects in untreated units; for a program implementer, it provides information on what would happen in the control areas if the program were expanded. The $\widehat{\tau}_{ATE}$ combines estimates of historical and hypothetical effects to provide an estimate of what the expected effect would for a typical unit in the study population.

In the same way b_{OLS} provides another average, one that is based on variance rather than prevalence. Indeed b_{OLS} corresponds exactly to what Crump, Joseph, Imbens, and Mitnik (2006) identify as the most precisely estimable average treatment effect (under homoskedasticity). This consideration provides a statistical but not a substantive justification for a focus on b_{OLS}. One substantive interpretation of the quantity estimated by b_{OLS} is the following. Say that stratum level treatment effects are independently drawn from some distribution of possible effects, f, with expectation τ^*. Making statements about average treatment effects given a realized distribution of τ_xs requires estimating $\widehat{\tau}_{ATE} = \sum_x \frac{w_x}{\sum_j w_j}\widehat{\tau}_x$; but making statements about the expected treatment effects for some new (out of sample) stratum requires estimating τ^*. Thus there is independent interest in estimating τ^*. Clearly if estimates of stratum level treatment effects are unbiased then $\widehat{\tau}_{ATE}$ gives an unbiased estimate of τ^*; b_{OLS} could however give a lower variance unbiased estimate of τ^* if (p_x) and (τ_x) are independent. In fact, as given in our first proposition, with independence and homoskedasticity, b_{OLS} is the minimum variance unbiased estimate of τ^* among all averages of ($\widehat{\tau}_x$).

Proposition 1. *If outcomes are distributed with constant variance conditional upon observables and* (p_x) *and* (τ_x) *are independent, then in the class of convex combinations of* ($\widehat{\tau}_x$) *the minimum variance unbiased estimate of*

τ^* *is* $\widehat{\tau^*} = b_{OLS} = \sum_x \frac{p_x(1-p_x)w_x}{\sum_j p_j(1-p_j)w_j}\widehat{\tau}_x$.

Proof. See appendix. □

An application of this result might be the following. Say that a subset of provinces (strata) are randomly selected for study and that lotteries are held in each one, with selection probabilities in each stratum that may differ but for reasons unrelated to potential outcomes. Then, assuming constant variance, the least squares estimate provides an efficient unbiased estimate of treatment effects in unsampled provinces.

The proof of the proposition proceeds similarly to that in Crump, Joseph, Imbens, and Mitnik (2006). Independence is important for the result. If (p_x) and (τ_x) are not independent then $\widehat{\tau^*}$ may not be unbiased, even though treatment is ignorable conditional upon observables. To illustrate, say there are two equally sized strata, that $\tau_x \sim U[0,1]$ and that $p_x = \frac{\tau_x}{2}$. Then $\tau^* = 0.50$ but: $\mathsf{E}(\widehat{\tau^*}) = \mathsf{E}\left(\frac{\frac{1}{2}\tau_1(1-\frac{1}{2}\tau_1)\tau_1 + \frac{1}{2}\tau_2(1-\frac{1}{2}\tau_2)\tau_2}{\frac{1}{2}\tau_1(1-\frac{1}{2}\tau_1) + \frac{1}{2}\tau_2(1-\frac{1}{2}\tau_2)}\right) \approx 0.57$. More generally, whereas the expected value of $\widehat{\tau}_{ATE}$ does not depend upon assignment probabilities, confidence in the b_{OLS} estimate requires confidence that there is no correlation between (τ_x) and $(p_x(1-p_x))$. This may be difficult to defend when neither the stratum treatment effects nor the assignment probabilities are homogeneous.

While under some circumstances there may be substantive arguments for examining the particular OLS averaging, when assignment probabilities and treatment effects are not thought to be independent, b_{OLS} may not be a fundamental quantity of interest. In such cases it is of interest to know how b_{OLS} compares to estimates of average treatment effects. Before introducing our main result we note three immediate results that relate b_{OLS} to $\widehat{\tau}_{ATE}$, $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$.

- First, if for all x, j, $\widehat{\tau}_x = \widehat{\tau}_j$ then $b_{OLS} = \widehat{\tau}_{ATE} = \widehat{\tau}_{ATT} = \widehat{\tau}_{ATC}$. With the same estimates of treatment effects in each stratum, different averagings yield the same result.

- Second, if there is some p such that for all x either $p_x = p$ or $p_x = 1-p$ then again $b_{OLS} = \widehat{\tau}_{ATE}$. The case in which $p_x = p$ for all x corresponds to a situation in which assignment to treatment is orthogonal to strata and in this case $b_{OLS} = \widehat{\tau}_{ATE} = \widehat{\tau}_{ATT} = \widehat{\tau}_{ATC}$. Moreover in this case least squares returns $\widehat{\tau}_{ATE}$ even if strata are ignored. In the other cases, in which assignment to treatment may not be orthogonal to strata (but for some x, j, $p_x = 1 - p_j$), b_{OLS} still corresponds to the $\widehat{\tau}_{ATE}$ when strata effects are included, although failing to condition on strata can produce bias.

- Third, since $\frac{p_x(1-p_x)}{\sum_j p_j(1-p_j)} \approx \frac{p_x}{\sum_j p_j}$ if p_x is small for all x and $\frac{p_x(1-p_x)}{\sum_j p_j(1-p_j)} \approx \frac{1-p_x}{\sum_j (1-p_j)}$ if p_x is large for all x we have that $b_{OLS} \approx \widehat{\tau}_{ATT}$ for 'rare' treatments and $b_{OLS} \approx \widehat{\tau}_{ATC}$ for 'common' treatments.

These three results provide some guidelines for when b_{OLS} is close to quantities of interest. However, when these conditions do not hold the differences between b_{OLS} and treatment effects of interest can be large. In addition since neither b_{OLS} or $\widehat{\tau}_{ATE}$ depend on sample sizes, systematic differences between them can obtain for arbitrarily large amounts of data. It is natural to ask then under what conditions are there tighter bounds than $\min(\widehat{\tau}_x)$ and $\max(\widehat{\tau}_x)$ on b_{OLS}.

3.2 Monotonicity

Our main result establishes that if assignment probabilities are monotonic in within-stratum treatment effects, then the OLS estimate lies between $\widehat{\tau}_{ATC}$ and $\widehat{\tau}_{ATT}$. As we note below, an implication is that if $\widehat{\tau}_{ATC}$ and $\widehat{\tau}_{ATT}$ are themselves unbiased then under the conditions of the proposition, the least squares estimate is expected to lie between the expected values of τ_{ATC} and τ_{ATT}.

Proposition 2. *If for all x, j, $p_x \geq p_j \leftrightarrow \widehat{\tau}_x \geq \widehat{\tau}_j$, or if for all x, j, $p_x \leq p_j \leftrightarrow \widehat{\tau}_x \geq \widehat{\tau}_j$, then $b_{OLS} \in Conv\{\widehat{\tau}_{ATC}, \widehat{\tau}_{ATT}\}$*

Proof. See appendix. □

Monotonicity is thus a sufficient condition for b_{OLS} to lie between $\widehat{\tau}_{ATC}$ and $\widehat{\tau}_{ATT}$ (the proof of the proposition identifies sharp conditions that are weaker than monotonicity).

We note that there is in general no *ex ante* reason to expect a monotonic relation between the $\widehat{\tau}_x$ and the p_x. However, such relations may arise if both τ_x and p_x reflect some systematic feature of units. In the case of spillovers considered above for example, a monotonic relation could hold if more connected subjects are not just more likely to receive treatment but are also more (or less) likely to be influenced by a new technology. Monotonicity is also natural in the case of a Roy-type selection model as described above: one subject is more likely than another to select into treatment precisely because he or she predicts greater benefits from treatment.

Finally we note that while monotonicity ensures that b_{OLS} lies between $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$ there is no guarantee that $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$ are close to each other. Indeed, all else equal, the difference between these two is greatest under monotonicity.[2]

3.3 Special case with two strata

The relation between b_{OLS} and treatment effects is particularly simple in the two stratum case. In this case, since monotonicity is guaranteed, we have the following corollary.

Corollary 1. *If there are two strata then* $b_{OLS} \in Conv\,\{\widehat{\tau}_{ATC}, \widehat{\tau}_{ATT}\}$

Moreover a somewhat stronger statement is possible in this case. With two strata, b_{OLS} is a convex combination of $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$, with weights that depend only on the probability of assignment to treatment. In particular, for p_1, $p_2 \in (0, 1)$, the weight on $\widehat{\tau}_{ATT}$ is given by

$$\lambda_{ATT} = \frac{p_1 w_1 + p_2 w_2}{\frac{p_1 w_1}{1-p_2} + \frac{p_2 w_2}{1-p_1}} \tag{5}$$

In general however with more than two strata there is no guarantee that b_{OLS} lies between $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$. An implication is that even when it does, the weights on $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$ do not depend on propensities alone, but also on the particular values taken by the collection $(\widehat{\tau}_j)$.

3.4 Estimators and estimands

Proposition 2 provides a condition under which least squares estimates of treatment effects lie between estimates of the ATC, $\widehat{\tau}_{ATC}$, and estimates of the ATT, $\widehat{\tau}_{ATT}$. There is no guarantee however, even in the simplest case, that least squares estimates will lie between the *estimands* ATT and ATC.

[2] In particular, given sets $(\tau_x)_{x=1}^k$ and $(p_x)_{x=1}^k$ for k equal sized strata $\mathsf{E}(ATT - ATC)$ is maximized (resp. minimized) by a (bijective) mapping $h : \{1, 2, ...k\} \rightarrow \{1, 2, ...k\}$ for which $(\tau_x)_{x=1}^k$ is monotonically increasing (resp. decreasing) in $(p_{h(i)})_{x=1}^k$.

To illustrate, consider a case with one stratum and two units, one of which is to be assigned to treatment, with $y_{1c} = y_{1t} = 0$ and $y_{2c} = y_{2t} = 1$. Clearly $\tau_i = 0$ for each i and so whatever the assignment to treatment $ATT = ATC = ATE = 0$, however if unit 1 is assigned then $b_{OLS} = \widehat{\tau}_{ATE} = -1$ and if unit 2 is assigned then $b_{OLS} = \widehat{\tau}_{ATE} = 1$.

Nevertheless, there is an immediate version of Proposition 2 for estimands; specifically, fixing the share of units in each stratum that receives treatment according to (p_x), if for each stratum x, $\widehat{\tau}_x$ provides an unbiased estimate for τ_x then under the conditions of the proposition, the least squares estimate lies between the expected value of ATT and the expected value of ATC in expectation. That is $\mathsf{E}(b_{OLS}) \in Conv(\mathsf{E}(ATC), \mathsf{E}(ATT))$.

4 Illustrations

We provide four illustrations of these relations.

4.1 Two strata illustration

Figure 1 illustrates the relations between b_{OLS} and treatment effects for a case with two equally sized strata. We set treatment effects at 0 and 1 all units in stratum 1 and stratum 2; we set the propensity to 0.25 for the first stratum and let the propensity for the second vary between 0 and 1.[3] The figure shows that in this case b_{OLS} can diverge substantially from ATE but, since there are only two strata, it always lies between ATT and ATC.

[3] For this and the next two illustrations we assume constant treatment effects within strata and ignore the distinction between $\widehat{\tau}_x$ and τ_x.

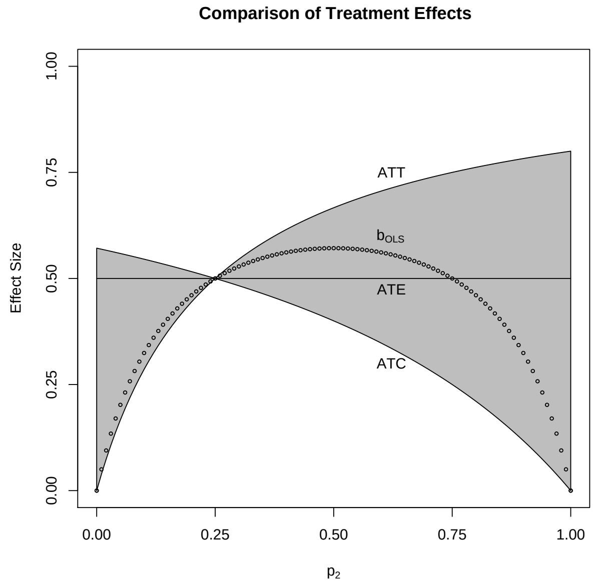

Figure 1 Treatment effects for $p_1 = .25$, $w_1 = w_2$, $\tau_1 = 0$ and $\tau_2 = 1$ for p_2 in $(0, 1)$. We see: (i) All estimates coincide when $p_1 = p_2 = .25$ (ii) $b_{OLS} = ATE$ when $p_2 = 1 - p_1$ (iii) $b_{OLS} = 0 = \tau_1$ (and data on stratum 2 is ignored) at extreme values of p_2 (iv) b_{OLS} is 'close' to ATT when p_2 is small and (v) b_{OLS} lies between ATC and ATT for all values of p_2.

4.2 Failure of monotonicity with three strata

Now consider a case with three equal sized strata in which:

$$\begin{array}{rcrl} \tau_1 & = & 3 & p_1 = \frac{1}{2} - \frac{\sqrt{3}}{4} \\ \tau_2 & = & -3 & p_2 = \frac{1}{2} \\ \tau_3 & = & 3 & p_3 = \frac{1}{2} + \frac{\sqrt{3}}{4} \end{array}$$

In this case it is easy to verify that $ATE = ATT = ATC = 1$ but $b_{OLS} = -1$. The divergence arises from the fact that stratum 2 has the greatest treatment variance and so is weighted more heavily by b_{OLS} than by ATT and ATC.

The next two examples show situations in which b_{OLS} sometimes does and sometimes does not lie between ATT and ATC depending on the relation between (p_j) and (τ_j); in the first case we examine a spillover problem, in the second we revisit data on the effects of participation in the military on earnings of applicants.

4.3 Approximate monotonicity with a continuum of strata

Consider next the problem in which assignment to treatment is related to how connected one is on a network. Let $k \sim U[0,1]$ denote the connectedness of a subject and assume that more connected individuals are more likely to receive treatment according to $p(k) = \sqrt{k}$. Say that connectedness is associated with worse outcomes in general and that treatment has a positive effect but the effect is strongest for people with close to some optimal degree of connectedness, γ. In particular we assume $y(T,k|\gamma) = T \times (1-(\gamma-k)^2) - k$. In this case, τ_j is decreasing in k for $\gamma = 0$ and increasing for $\gamma = 1$; it is non-monotonic (concave) for intermediate values of γ but 'closer' to monotonic for extreme values.

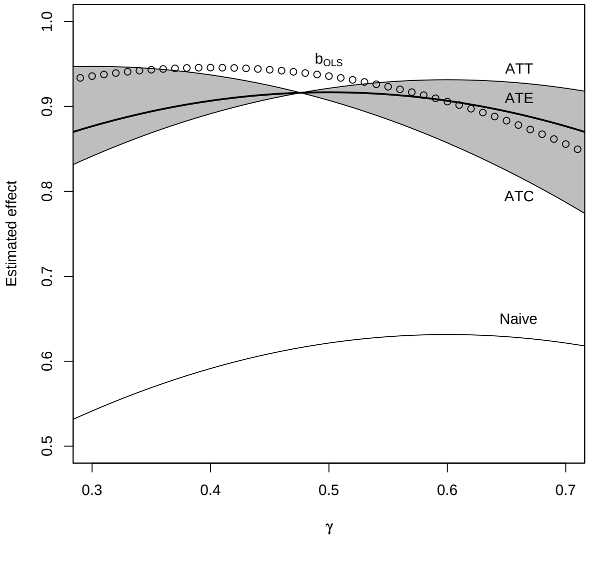

Figure 2 Treatment effects for $p(k) = \sqrt{k}$,
$y(T,k|\gamma) = T \times (((1-\gamma-k)^2) - k$ and $k \sim U[0,1]$. In this case $ATE(k)$ is closer to monotonic in $p(k)$ for γ close to 0 and 1. The shaded band in the figure marks the region between ATT and ATC. The width of this band is greatest for extreme values of γ. We see that the naive estimate diverges from all treatment effects, b_{OLS} lies between ATT and ATC when γ is close to 0 or 1 but lies above both for intermediate values of γ.

The ATE, ATT, ATC, b_{OLS} and naive estimates (simple difference in means) for this case are as shown in Figure 2. The naive estimate is biased down since it fails to take account of the fact that more connected individuals (who are also more likely to receive treatment) fare more poorly, independent of the treatment. The OLS estimate lies between ATT and ATC for more extreme values of γ, but not for intermediate values. Finally we see that the difference between ATT and ATC is greatest for extreme values of γ.

4.4 Estimates of the effects of military service on earnings

Our final illustration revisits the analysis of Angrist (1998). Angrist examines the effects of military service on subsequent earnings, for subjects broken down by race. Ignorability is defended on the basis that the population under study consists entirely of applicants and most of the data employed by the military to select individuals from this pool (age, schooling and test scores) are available to the researchers. Plausibly, then, participation in the military is random conditional upon these features.

The outcome of interest is earnings which was gathered from social security earnings data and matched to the profiles of 750,000 applicants to the military, themselves grouped into 5654 strata /treatment combinations.[4]

An interesting feature of this study is that outcome data is available for multiple years, from the mid 1970s to the early 1990s. The analysis suggests that service had a positive effect on earnings for white veterans in the early 1980s (when they were insulated from the economic recession) but has negative effects in later periods. Nonwhite veterans experienced positive gains throughout the 1980s. In terms of our examination of treatment effects, the multiple outcome measures mean that associated with each individual there is a single assignment probability p_j, but multiple stratum treatment effects, τ_j. It is possible that the relations between (p_j) and (τ_j) vary over time, possibly resulting in different relations between regression estimates and treatment effects.

[4] Due to lack of balance there are 3,167 strata 2,487 of which contain both treatment and control cells. In the replication below we limit attention to the strata examined by Angrist which excludes early applicants and certain education profiles. After removing these plus strata with null or missing data we are left with an average of 820 strata per year.

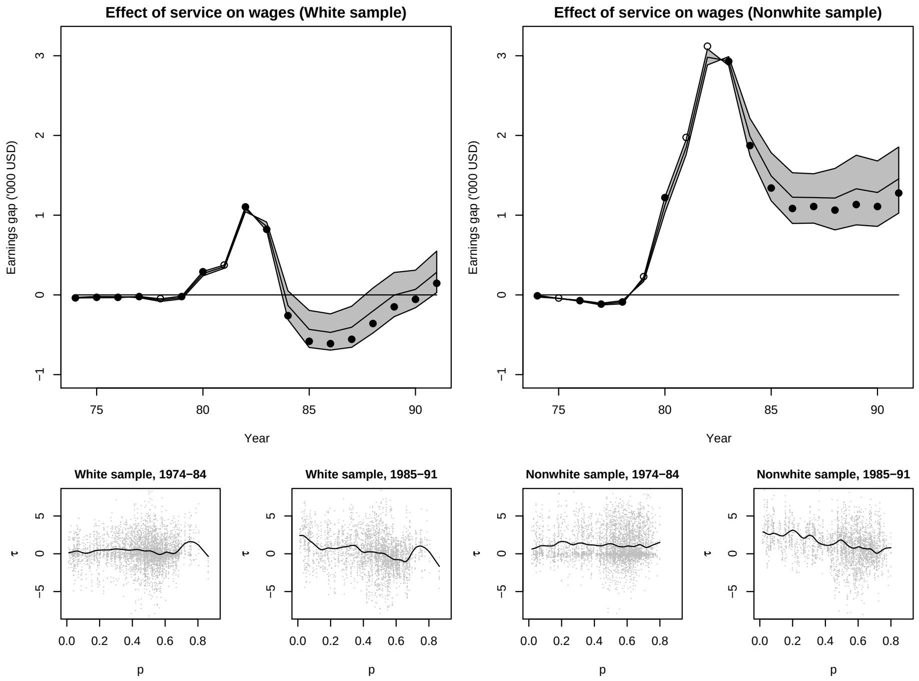

Figure 3 The top panel shows effects of participation in the military on earnings of white and nonwhite veterans for years 1974-91. The shaded band marks the region between $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$, the center line gives the $\widehat{\tau}_{ATE}$. The OLS estimate is marked with circles, which are filled whenever b_{OLS} lies between $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$. Bottom panels show the relation between p_x and $\widehat{\tau}_x$ for each group for two time periods. In early periods these relations are approximately flat and so $\widehat{\tau}_{ATC} \approx \widehat{\tau}_{ATT}$; in later periods there is a negative (near) monotonic relation and so $\widehat{\tau}_{ATC} > \widehat{\tau}_{ATT}$. In these later periods b_{OLS} always lies between $\widehat{\tau}_{ATE}$ and $\widehat{\tau}_{ATT}$.

Figure 3 shows that this is indeed the case. In early periods the relation between (p_x) and $(\widehat{\tau}_x)$ was approximately flat. In these periods, $\widehat{\tau}_{ATC}$ always lies close to $\widehat{\tau}_{ATT}$ for both white and nonwhite applicants. In these cases b_{OLS} lies close to these treatment effects, but it sometimes lies marginally outside of $conv(\widehat{\tau}_{ATT}, \widehat{\tau}_{ATC})$. In later years however there is a marked negative relation between probability of treatment and average treatment effects—those individuals most likely to be selected (in many cases those with the highest AFQT scores) were the ones for which participation in the military would have the most adverse effects on income in the long term. This is true for both white and non-white groups. For these cases we find, consistent with our results above, that the difference between $\widehat{\tau}_{ATT}$ and $\widehat{\tau}_{ATC}$ is larger, but that b_{OLS} lies between these bounds. Whereas Angrist found that given monotonicity in later years b_{OLS} was greater than $\widehat{\tau}_{ATT}$, our results suggest, and the data confirm, that b_{OLS} was nevertheless always bounded from above in these cases by $\widehat{\tau}_{ATC}$.

5 Conclusion

Researchers commonly seek to account for covariates that are correlated with treatment assignment probabilities. This situation can arise even in the context of a randomized trial when researchers are interested in spillover effects, when there are multiple correlated treatments, and when treatment is assigned through multiple lotteries with different assignment probabilities.

If assignment probabilities are known, then matching procedures can be used to recover average treatment effects; alternatively appropriate weights can be used; or outcomes can be regressed on strata dummies and a full a set of interactions between treatment and strata. These solutions all return the correct results. In practice however the usual approach for estimating the causal effect of a treatment is to regress outcomes on treatment and some collection of covariates, possibly allowing for flexible functional forms for the covariates. This 'naive' approach is not guaranteed to produce unbiased estimates of the average treatment effect but it is not well known how estimates generated in this manner diverge from the $\widehat{\tau}_{ATE}$.

Here we identify conditions under which estimates from the 'naive' approach still approximate quantities of interest. For 'rare' treatments the estimate lies close to the average treatment effect for the treated; for 'common' treatments it is close to the treatment effect for the controls. Under other conditions, when *ex ante* propensities are independent of treatment effects, then least squares may provide an efficient unbiased estimate for treatments in unsampled strata. Finally when a monotonicity condition is satisfied least squares estimates produce results that lie between estimates of the average treatment effect for the treated and the average treatment effect for the controls. Thus when higher values on third variables are associated both with more positive (or more negative) outcomes and with a higher (or lower) propensity to being assigned to treatment, the regression estimate is bounded by causal quantities of interest. When these conditions are not met however least squares estimates may diverge substantially and range anywhere from the lowest within-stratum treatment effect to the highest within-stratum treatment effect.

References

Abadie, A., J. L. Herr, G. W. Imbens, and D. M. Drukker (2004): "NNMATCH: Stata module to compute nearest-neighbor bias-corrected estimators," Working Paper, Harvard University.

Angrist, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66(2), 249–288.

Crump, R. K., H. V. Joseph, G. W. Imbens, and O. A. Mitnik (2006): "Moving the Goalposts: Addressing Limited Overlap in Estimation of Average Treatment Effects by Changing the Estimand," Discussion Paper, IZA DP.

Duflo, E., R. Glennerster, and M. Kremer (2007): "Chapter 61 Using Randomization in Development Economics Research: A Toolkit," vol. 4 of *Handbook of Development Economics*, pp. 3895 – 3962. Elsevier.

Freedman, D. A. (2008): "On regression adjustments to experimental data," *Advances in Applied Mathematics*, 40, 180–93.

Habyarimana, J., M. Humphreys, D. Posner, and J. Weinstein (2007): "Why Does Ethnic Diversity Undermine Public Goods Provision?," *American Political Science Review*, 101(4), 709–725.

Ho, D. E., K. Imai, G. King, and E. A. Stuart (2007): "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15(3), 199–236.

Iacus, S. M., G. King, and G. Porro (2008): "Matching for Causal Inference Without Balance Checking," Working Paper, Harvard University.

Miguel, E., and M. Kremer (2004): "Worms: identifying impacts on education and health in the presence of treatment externalities," *Econometrica*, 72(1), 159–217.

Oster, E., and R. Thornton (2009): "Determinants of Technology Adoption: Private Value and Peer Effects in Menstrual Cup Take-Up," Working Paper, Chicago University.

Rosenbaum, P., and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

6 Appendix

Proposition 1 If outcomes are distributed with constant variance conditional upon observables and (p_x) and (τ_x) are independent, then in the class of convex combinations of $(\widehat{\tau}_x)$ the minimum variance unbiased estimate of τ^* is $\widehat{\tau^*} = b_{OLS} = \sum_x \frac{p_x(1-p_x)w_x}{\sum_j p_j(1-p_j)w_j}\widehat{\tau}_x$.

Proof of Proposition 1. The problem is to choose weights $(\alpha_x)_{x=1}^k$ with $\sum \alpha_x = 1$ to minimize:

$$var\left(\sum_x \alpha_x \widehat{\tau}_x\right) = \sum_x \alpha_x^2 var\left(\widehat{\tau}_x\right)$$

Letting T and C denote the set of treatment and control units respectively, and S_x the collection of units in stratum x, we have:

$$\begin{aligned} var(\widehat{\tau}_x) =& var\left(\frac{\sum_{T\cap S_x} y_h}{p_x w_x n} - \frac{\sum_{C\cap S_x} y_h}{(1-p_x)w_x n}\right) \\ =& \frac{\sigma_{ti}^2}{p_x w_x n} + \frac{\sigma_{ci}^2}{(1-p_x)w_x n} \\ =& \frac{(1-p_x)w_x n\sigma_{ti}^2 + p_x w_x n \sigma_{ci}^2}{p_x(1-p_x)(w_x n)^2} \end{aligned}$$

where σ_{ki}^2 is the variance of outcomes in treatment group k and stratum x.

Under constant variance, $\sigma_{ti}^2 = \sigma_{ci}^2 = \sigma^2$:

$$var(\widehat{\tau}_x) = \frac{(1-p_x)w_x n\sigma_{ti}^2 + p_x w_x n \sigma_{ci}^2}{p_x(1-p_x)(w_x n)^2} = \frac{\sigma^2}{n}\frac{1}{p_x(1-p_x)w_x}$$

And so the problem is to find

$$\arg \min_{(\alpha_x)_{x=1}^k} \left(\sum_x \frac{\sigma^2}{n}\frac{\alpha_x^2}{p_x(1-p_x)w_x}\right) = \arg \min_{(\alpha_x)_{x=1}^k} \left(\sum_x \frac{\alpha_x^2}{p_x(1-p_x)w_x}\right)$$

which has solution:

$$\alpha_x = \frac{p_x(1-p_x)w_x}{\sum_j p_j(1-p_j)w_j}$$

And so the minimum variance estimate of τ^* is

$$\widehat{\tau^*} = \sum_x \frac{p_x(1-p_x)w_x}{\sum_j p_j(1-p_j)w_j}\widehat{\tau}_x = b_{OLS}$$

□

Proposition 2 If for all x, j, $p_x \geq p_j \leftrightarrow \widehat{\tau}_x \geq \widehat{\tau}_j$, or if for all x, j, $p_x \leq p_j \leftrightarrow \widehat{\tau}_x \geq \widehat{\tau}_j$, then $b_{OLS} \in Conv\left\{\widehat{\tau}_{ATC}, \widehat{\tau}_{ATT}\right\}$

Proof of Proposition 2. We consider the case in which p_x is monotonically increasing in τ_x and so $\widehat{\tau}_{ATT} \geq \widehat{\tau}_{ATC}$. The proof for the case in which p_x is monotonically decreasing in τ_x is similar.

Assume first contrary to the proposition that $b_{OLS} > \widehat{\tau}_{ATT}$. Then:

$$\frac{\sum p_x(1-p_x)w_x\widehat{\tau}_x}{\sum p_x(1-p_x)w_x} > \frac{\sum p_x w_x \widehat{\tau}_x}{\sum p_x w_x}$$

$$\leftrightarrow$$

$$\sum \widehat{\tau}_x \left(\frac{p_x w_x}{\sum p_j \widehat{\tau}_j} - \frac{p_x^2 w_x}{\sum p_j^2 w_j}\right) > 0$$

But this yields a contradiction since under monotonicity

$$\sum \tau_x \left(\frac{p_x w_x}{\sum p_j w_j} - \frac{p_x^2 w_x}{\sum p_j^2 w_j}\right) \leq 0$$

To see this note that:

$$\frac{p_x w_x}{\sum_j p_j w_j} - \frac{p_x^2 w_x}{\sum_j p_j^2 w_j} \geq 0$$

$$\leftrightarrow$$

$$\sum_j w_j p_j (p_j - p_x) \geq 0$$

Hence $\frac{p_x w_x}{\sum_j p_j w_j} - \frac{p_x^2 w_x}{\sum_j p_j^2 w_j}$ is greater than or equal to 0 for all p_x less than some value p^* and less than or equal to 0 for all p_x greater than p^*. The greatest value that $\sum \widehat{\tau}_x \left(\frac{p_x w_x}{\sum p_j w_j} - \frac{p_x^2 w_x}{\sum p_j^2 w_j}\right)$ can take is with $\widehat{\tau}_x$ as high as possible for all x with $p_x \leq p^*$ and as low as possible for all x with $p_x \geq p^*$ subject to preserving the ordering that $\widehat{\tau}_x \geq \widehat{\tau}_j$ if $p_x \geq p_j$. Thus the highest value is attained when $\widehat{\tau}_x = ATE^*$ for all x and some ATE^*. But in this case, since $\sum_x \left(\frac{p_x w_x}{\sum_j p_j w_j} - \frac{p_x^2 w_x}{\sum_j p_j^2 w_j}\right) = 0$, we have $\sum \widehat{\tau}_x \left(\frac{p_x w_x}{\sum p_j w_j} - \frac{p_x^2 w_x}{\sum p_j^2 w_j}\right) = 0$ and so in general $\sum \widehat{\tau}_x \left(\frac{p_x w_x}{\sum p_j w_j} - \frac{p_x^2 w_x}{\sum p_j^2 w_j}\right) \leq 0$.

In the same way if $b_{OLS} < \widehat{\tau}_{ATC}$. Then:

$$\frac{\sum p_x(1-p_x)w_x\widehat{\tau}_x}{\sum p_j(1-p_j)w_j} < \frac{\sum (1-p_x)w_x\widehat{\tau}_x}{\sum (1-p_j)w_j}$$

$$\leftrightarrow$$

$$\sum\nolimits_x \widehat{\tau}_x \left(\frac{(1-p_x)w_x}{\sum_j (1-p_j)w_j} - \frac{(1-p_x)^2 w_x}{\sum_j (1-p_j)^2 w_j} \right) <0$$

But again we have a contradiction since under monotonicity

$$\sum\nolimits_x \widehat{\tau}_x \left(\frac{(1-p_x)w_x}{\sum_j (1-p_j)w_j} - \frac{(1-p_x)^2 w_x}{\sum_j (1-p_j)^2 w_j} \right) \geq 0$$

This is established in the same way:

$$\frac{(1-p_x)w_x}{\sum_j (1-p_j)w_j} - \frac{(1-p_x)^2 w_x}{\sum_j (1-p_j)^2 w_j} \geq 0$$

$$\leftrightarrow$$

$$\sum\nolimits_j (1-p_j)w_j(p_x - p_j) \geq 0$$

Thus $\frac{(1-p_x)}{\sum_j (1-p_j)} - \frac{(1-p_x)^2}{\sum_j (1-p_j)^2}$ is nonnegative for high p_x and nonpositive for low p_x. The lowest value $\sum_x \widehat{\tau}_x \left(\frac{(1-p_x)w_x}{\sum (1-p_j)w_j} - \frac{(1-p_x)^2 w_x}{\sum (1-p_j)^2 w_j} \right)$ can take is when strata with high (resp. low) p_x have as low (resp. high) a value of $\widehat{\tau}_x$ as is consistent with monotonicity, in which case $\sum_x \widehat{\tau}_x \left(\frac{(1-p_x)w_x}{\sum (1-p_j)w_j} - \frac{(1-p_x)^2 w_x}{\sum (1-p_j)^2 w_j} \right) = 0$.
□