forthcoming in BIOMETRIKA.

Dealing with limited overlap in estimation of average treatment effects

By RICHARD K. CRUMP

Department of Economics, University of California at Berkeley, Berkeley, California 94720, U.S.A.

crump@econ.berkeley.edu

V. JOSEPH HOTZ

Department of Economics, Duke University, Durham, North Carolina 27708, U.S.A.

hotz@econ.duke.edu

GUIDO W. IMBENS

Department of Economics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

imbens@harvard.edu

and OSCAR A. MITNIK

Department of Economics, University of Miami, Coral Gables, Florida 33124, U.S.A.

omitnik@miami.edu

Summary

Estimation of average treatment effects under unconfounded or ignorable treatment assignment is often hampered by lack of overlap in the covariate distributions between treatment groups. This lack of overlap can lead to imprecise estimates, and can make commonly used estimators sensitive to the choice of specification. In such cases researchers have often used ad hoc methods for trimming the sample. In this paper we develop a systematic approach to addressing lack of overlap. We characterize optimal subsamples for which the average treatment effect can be estimated most precisely. Under some conditions, the optimal selection rules depend solely on the propensity score. For a wide range of distributions, a good approximation to the optimal rule is provided by the simple rule of thumb to discard all units with estimated propensity scores outside the range $[0.1, 0.9]$.

Some key words: Average treatment effect; Causality; Ignorable treatment assignment; Overlap; Propensity score; Treatment effect heterogeneity; Unconfoundedness.

1. Introduction

There is a large literature on estimating average treatment effects under assumptions of unconfoundedness or ignorability, following the seminal work by Rubin (1974, 1997), Rosenbaum & Rubin (1983), and Rosenbaum (1989). Researchers have developed estimators based on regression methods (Hahn, 1998; Heckman et al., 1998), matching (Rosenbaum, 1989; Abadie & Imbens, 2006), and methods based on the propensity score (Rosenbaum & Rubin, 1983; Hirano et al., 2003). Related methods for missing data problems are discussed in Robins & Rotnitzky (1995); see Rosenbaum (2001) and Imbens (2004) for general surveys. An important practical concern in implementing these methods is the need for overlap in the covariate distributions in the treated and control subpopulations. Even if the supports of the two covariate distributions are identical, there may be parts of the covariate space with limited numbers of observations for either the treatment or control group. Such areas of limited overlap can lead to conventional estimators of average treatment effects having substantial bias and large variances. Often researchers discard units for whom there are no close counterparts in the subsample with the opposite treatment. The implementation of these methods is typically ad hoc, with, for example, researchers discarding units for whom they cannot find a match that is identical in terms of the propensity score up to 1, 2 or even 8 digits; see for example Grzybowski et al. (2003) and Vincent et al. (2002).

We propose a systematic approach to dealing with samples with limited overlap in the covariate distributions in the two treatment arms. Our proposed method is not tied or limited to a specific estimator. It has some optimality properties and is straightforward to implement in practice. We focus on average treatment effects within a selected subpopulation, defined solely in terms of covariate values, and look for the subpopulation that allows for the most precise estimation of the average treatment effect. We show that this problem is, in general, well defined, and, under some conditions, leads to discarding observations with propensity scores outside an interval

$[\alpha, 1-\alpha]$, with the optimal cut-off value α determined by the marginal distribution of the propensity score. Our approach is consistent with the common practice of dropping units with extreme values of the propensity score, with two differences. First, the role of the propensity score in the selection rule is not imposed a priori, but emerges as a consequence of the criterion, and, second, there is a principled way of choosing the cut-off value α. The subset of observations is defined solely in terms of the joint distribution of covariates and the treatment indicator, and does not depend on the distribution of the outcomes. As a result, we avoid introducing deliberate bias with respect to the treatment effects being analysed. The precision gain from this approach can be substantial, with most of the gain captured by using a rule of thumb to discard observations with the estimated propensity score outside the range $[0.1, 0.9]$. The main cost is that potentially some external validity is lost by changing the focus to average treatment effects for a subset of the original sample.

We illustrate these methods using data on right heart catheterization from the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments.

2. The framework and a simple example

2·1. Underlying framework

The framework we use is that of Rosenbaum & Rubin (1983). We have a random sample of size N from a large population. For each unit i in the sample, let W_i indicate whether or not the treatment of interest was received, with $W_i = 1$ if unit i receives the treatment of interest, and $W_i = 0$ if unit i receives the control treatment. Let $Y_i(0)$ denote the outcome for unit i under control and $Y_i(1)$ the outcome under treatment. We observe W_i and Y_i, where

$$Y_i = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases}$$

In addition, we observe a K-dimensional vector of pre-treatment variables, or covariates, denoted by X_i, with support $\mathbb{X} \subset \mathbb{R}^K$. Define the two conditional mean func-

tions, $\mu_w(x) = E\{Y_i(w)|X_i = x\}$, the two conditional variance functions, $\sigma_w^2(x) = \text{var}\{Y_i(w)|X_i = x\}$, the conditional average treatment effect $\tau(x) = E\{Y_i(1) - Y_i(0)|X_i = x\} = \mu_1(x) - \mu_0(x)$, and the propensity score, the probability of selection into the treatment, $e(x) = \text{pr}(W_i = 1|X_i = x) = E(W_i|X_i = x)$.

We primarily focus on the sample and population average treatment effects

$$\tau_{\text{S}} = \frac{1}{N}\sum_{i=1}^{N}\tau(X_i), \qquad \tau_{\text{P}} = E\left\{Y_i(1) - Y_i(0)\right\}.$$

The difference between these estimands is important for our analyses, and we return to this issue in Remark 1 below. For sets $\mathbb{A} \subset \mathbb{X}$, let $1_{X_i\in\mathbb{A}}$ be an indicator for the event that X_i is an element of the set $\mathbb{A}$, and define the subsample average treatment effect

$$\tau_{\text{S},\mathbb{A}} = \frac{1}{N_{\mathbb{A}}}\sum_{i:X_i\in\mathbb{A}}\tau(X_i), \qquad N_{\mathbb{A}} = \sum_{i=1}^{N}1_{X_i\in\mathbb{A}},$$

so that $\tau_{\text{S},\mathbb{X}} = \tau_{\text{S}}$. We denote estimators for the sample and population average treatment effects by $\hat{\tau}$ and for the subsample average treatment effect by $\hat{\tau}_{\mathbb{A}}$. There is no need to index the estimators by S or P because estimators for the sample average treatment effect are estimators for the population average treatment effect as well.

To solve the identification problem, we maintain throughout the paper the following two assumptions. The first, the unconfoundedness assumption (Rosenbaum & Rubin, 1983), asserts that conditional on the pre-treatment variables, the treatment indicator is independent of the potential outcomes. The second assumption ensures overlap in the covariate distributions.

Assumption 1. We assume that $W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid X_i$.

Assumption 2. For some $c > 0$, and all $x \in \mathbb{X}$, $c \leq e(x) \leq 1 - c$.

The combination of these two assumptions is referred to as strong ignorability (Rosenbaum & Rubin, 1983).

2·2. A simple example

Consider the following example in which the covariate is a binary scalar. Suppose that $X_i = f$, female, or $X_i = m$, male, so that $\mathbb{X} = \{f, m\}$. For $x = f, m$, let N_x be the sample size for the subsample with $X_i = x$, and let $N = N_f + N_m$ be the overall sample size. Also, let $p = \text{pr}(X_i = m)$ be the population proportion of men, with $\hat{p} = N_m/N$. We use the shorthand τ_x for $\tau_{\{x\}}$, for $x = f, m$. Let N_{xw} be the number of observations with covariate $X_i = x$ and treatment indicator $W_i = w$, and let $\hat{e}_x = N_{x1}/N_x$ denote the value of the estimated propensity score, for $x = f, m$. Finally, let $\bar{Y}_{xw} = \sum_{i:X_i=x,W_i=w} Y_i/N_{xw}$ be the average outcome for the four subsamples. We assume that the distribution of the outcomes is homoskedastic, so that $\text{var}\{Y_i(w)|X_i = x\} = \sigma^2$ for all $x = f, m$ and $w = 0, 1$. The sample and population average effects can be written as

$$\tau_{\text{S}} = \hat{p}\tau_m + (1 - \hat{p})\tau_f, \qquad \tau_{\text{P}} = p\tau_m + (1 - p)\tau_f.$$

If the unconfoundedness assumption is maintained, the natural estimators for the average treatment effects for each of the two subpopulations are

$$\hat{\tau}_f = \bar{Y}_{f1} - \bar{Y}_{f0}, \qquad \hat{\tau}_m = \bar{Y}_{m1} - \bar{Y}_{m0}.$$

These estimators are unbiased and conditional on the covariates and treatment indicators their exact variances are

$$\text{var}\left(\hat{\tau}_f|X, W\right) = \sigma^2\left(\frac{1}{N_{f0}} + \frac{1}{N_{f1}}\right) = \frac{1}{N}\frac{\sigma^2}{\hat{e}_f(1 - \hat{e}_f)(1 - \hat{p})},$$

and

$$\text{var}\left(\hat{\tau}_m|X, W\right) = \sigma^2\left(\frac{1}{N_{m0}} + \frac{1}{N_{m1}}\right) = \frac{1}{N}\frac{\sigma^2}{\hat{e}_m(1 - \hat{e}_m)\hat{p}}.$$

respectively. The natural estimator for the sample, as well as the population, average treatment effect is

$$\hat{\tau} = \hat{p}\hat{\tau}_m + (1 - \hat{p})\hat{\tau}_f.$$

This estimator is unbiased for $\tau_{\rm S}$, conditional on X and W, and unbiased, unconditionally, for $\tau_{\rm P}$. The conditional variance of $\hat{\tau}_{\rm S}$ is

$$\text{var}\left(\hat{\tau}|X,W\right)=\frac{\sigma^2}{N}\left\{\frac{\hat{p}}{\hat{e}_m(1-\hat{e}_m)}+\frac{1-\hat{p}}{\hat{e}_f(1-\hat{e}_f)}\right\}.$$

It follows that the asymptotic variance of $N^{1/2}(\hat{\tau}-\tau_{\rm S})$ converges to

$$\text{avar}\left\{N^{1/2}\left(\hat{\tau}-\tau_{\rm S}\right)\right\}=\sigma^2\left\{\frac{p}{e_m(1-e_m)}+\frac{1-p}{e_f(1-e_f)}\right\}=E\left\{\frac{\sigma^2}{e_X(1-e_X)}\right\}. \quad (1)$$

The asymptotic unconditional variance of $\hat{\tau}$, that is, the asymptotic variance of $N^{1/2}(\hat{\tau}-\tau_{\rm P})$, is

$$\text{avar}\left\{N^{1/2}\left(\hat{\tau}-\tau_{\rm P}\right)\right\}=E\left\{\frac{\sigma^2}{e_X(1-e_X)}+\left(\tau_X-\tau_{\rm P}\right)^2\right\}. \quad (2)$$

Now let us turn to estimators for subpopulation average treatment effects of the type $\tau_{S,\mathbb{A}}$. The key result of the paper concerns the comparison of $\text{var}(\hat{\tau}_{\mathbb{A}}|X,W)$ for different sets $\mathbb{A}$, according to a variance minimization criterion. Let $\mathcal{A}$ be the set of all subsets of $\mathbb{X}$, excluding the empty set. Then we are interested in the set $\hat{\mathbb{A}}$ that solves

$$\text{var}(\hat{\tau}_{\hat{\mathbb{A}}}|X,W)=\inf_{\mathbb{A}\in\mathcal{A}}\text{var}(\hat{\tau}_{{\rm S},\mathbb{A}}|X,W). \quad (3)$$

In the binary covariate example considered in this section, $\mathbb{X}=\{f,m\}$, so that $\mathcal{A}=\{\{f\},\{m\},\{f,m\}\}$, and the problem simplifies to finding the set $\hat{\mathbb{A}}$ that solves

$$\text{var}(\hat{\tau}_{\hat{\mathbb{A}}}|X,W)=\min\left\{\text{var}(\hat{\tau}|X,W),\text{var}(\hat{\tau}_f|X,W),\text{var}(\hat{\tau}_m|X,W)\right\}.$$

In this case the solution is given by

$$\hat{\mathbb{A}}=\begin{cases}\{f\} & \text{if} \quad \left\{\hat{e}_m(1-\hat{e}_m)\right\}/\left\{\hat{e}_f(1-\hat{e}_f)\right\} < (1-\hat{p})/(2-\hat{p})\\ \mathbb{X} & \text{if} \quad (1-\hat{p})/(2-\hat{p})\le \left\{\hat{e}_m(1-\hat{e}_m)\right\}/\left\{\hat{e}_f(1-\hat{e}_f)\right\} < (1+\hat{p})/\hat{p}\\ \{m\} & \text{if} \quad (1+\hat{p})/\hat{p}\le \left\{\hat{e}_m(1-\hat{e}_m)\right\}/\left\{\hat{e}_f(1-\hat{e}_f)\right\}.\end{cases} \quad (4)$$

Remark 1. Note that we compare the conditional, not the unconditional, variances in (3), and so we compare objects like the right hand side of (1) rather than the right hand side of (2). Since the asymptotic unconditional variance of $\hat{\tau}$, given

in (2), depends on the conditional treatment effects τ_f and τ_m through the term $E\left\{(\tau_X - \tau_{\rm P})^2\right\}$, comparisons of the unconditional variances would make the optimal set depend on the value of the treatment effects. This has two disadvantages. First, it makes the optimal set depend on the distribution of the potential outcomes, rather than solely on the distribution of treatment and covariates, thus opening the door to potential biases. Secondly, implementing the implied criterion in the unconditional case would be considerably more difficult in practice because the lack of overlap that leads to the difficulties in precise estimation of $\tau_{\rm P}$ implies that some of the conditional treatment effects τ_x, and thus the unconditional variance, would be difficult to estimate precisely.

Remark 2. One can also define the population version of the set $\hat{\mathbb{A}}$, denoted by $\mathbb{A}^*$, as the equivalent of (4) with $\hat{p}$, $\hat{e}_f$, and $\hat{e}_m$ replaced by p, e_f, and e_m:

$$\mathbb{A}^* = \begin{cases} \{f\} & \text{if} & & \{e_m(1-e_m)\}/\{e_f(1-e_f)\} & < (1-p)/(2-p) \\ \mathbb{X} & \text{if} & (1-p)/(2-p) \leq & \{e_m(1-e_m)\}/\{e_f(1-e_f)\} & < (1+p)/p \\ \{m\} & \text{if} & (1+p)/p \leq & \{e_m(1-e_m)\}/\{e_f(1-e_f)\}\,. & \end{cases} \tag{5}$$

The set $\hat{\mathbb{A}}$ is a natural estimator for $\mathbb{A}^*$, and as a result $\tau_{\text{S},\hat{\mathbb{A}}}$ is a natural estimator for $\tau_{\text{S},\mathbb{A}^*}$. However, we focus on the asymptotic variance of $N^{1/2}(\hat{\tau}_{\hat{\mathbb{A}}} - \tau_{\text{S},\hat{\mathbb{A}}})$ rather than the asymptotic variance of $N^{1/2}(\hat{\tau}_{\hat{\mathbb{A}}} - \tau_{\text{S},\mathbb{A}^*})$; that is, we focus on the uncertainty of the estimator for the average effect conditional on the set we selected. This greatly simplifies the subsequent analysis: we can select the sample and then proceed to estimate the average treatment effect and its uncertainty, ignoring the first stage in which the sample was selected. In the binary covariate case it is again straightforward to see why this simplifies the analysis. Denote the estimated conditional variances for $\hat{\tau}_f$, $\hat{\tau}_m$ and $\hat{\tau}$ by

$$\hat{V}_f = \frac{1}{N}\frac{\hat{\sigma}^2}{\hat{e}_f(1-\hat{e}_f)(1-\hat{p})}, \qquad \hat{V}_m = \frac{1}{N}\frac{\hat{\sigma}^2}{\hat{e}_m(1-\hat{e}_m)\hat{p}},$$

and

$$\hat{V}_{\rm S} = \frac{\hat{\sigma}^2}{N}\left\{\frac{\hat{p}}{\hat{e}_m(1-\hat{e}_m)} + \frac{1-\hat{p}}{\hat{e}_f(1-\hat{e}_f)}\right\},$$

and define

$$\hat{V}_{\hat{\mathbb{A}}} = \begin{cases} \hat{V}_f & \text{if} \quad \{\hat{e}_m(1-\hat{e}_m)\}/\{\hat{e}_f(1-\hat{e}_f)\} < \zeta(\hat{p}) \\ \hat{V}_{\rm S} & \text{if} \quad \zeta(\hat{p}) \leq \{\hat{e}_m(1-\hat{e}_m)\}/\{\hat{e}_f(1-\hat{e}_f)\} < \eta(\hat{p}) \\ \hat{V}_m & \text{if} \quad \eta(\hat{p}) \leq \{\hat{e}_m(1-\hat{e}_m)\}/\{\hat{e}_f(1-\hat{e}_f)\}. \end{cases} \tag{6}$$

Then $\hat{V}_{\hat{\mathbb{A}}}^{-1}(\hat{\tau}_{\hat{\mathbb{A}}} - \tau_{{\rm S},\hat{\mathbb{A}}}) \longrightarrow \mathcal{N}(0,1)$ in distribution. In this case, $N^{1/2}(\tau_{{\rm S},\hat{\mathbb{A}}} - \tau_{{\rm S},\mathbb{A}^*})$ may in fact diverge, if, for example, one of the inequalities in (6) is an equality so that $\hat{\mathbb{A}}$ does not converge to $\mathbb{A}^*$, and $N^{1/2}(\tau_{S,\hat{\mathbb{A}}} - \tau_{S,\mathbb{A}^*})$ diverges.

In the remainder of the paper, we generalize the above analysis to the case with a vector of continuously distributed covariates. In that case the set $\mathcal{A}$ of subsets of $\mathbb{X}$ is not countable. In addition, for a particular subset $\mathbb{A} \in \mathcal{A}$ there is not a simple estimator, nor can we calculate exact variances for any estimator. We therefore compare asymptotic variances for efficient estimators. Instead of solving (3), we attempt to find the $\mathbb{A}^*$ that solves

$$\text{avar}\left\{N^{1/2}\left(\hat{\tau}^{\rm eff}_{\mathbb{A}^*} - \tau_{{\rm S},\mathbb{A}^*}\right)\right\} = \inf_{\mathbb{A}\in\mathcal{A}} \text{avar}\left\{N^{1/2}\left(\hat{\tau}^{\rm eff}_{\mathbb{A}} - \tau_{{\rm S},\mathbb{A}}\right)\right\},$$

where $\hat{\tau}^{\rm eff}_{\mathbb{A}}$ denotes any semiparametrically efficient estimator for $\tau_{{\rm S},\mathbb{A}}$. For $\mathbb{A}^*$ the average treatment effect is at least as accurately estimable as that for any other subset of the covariate space. This leads to a generalization of (5). Under some regularity conditions, this problem has a well-defined solution and, under the additional assumption of homoskedasticity, these subpopulations have a very simple characterization, namely the set of covariate values such that the propensity score is in the closed interval $[\alpha, 1-\alpha]$, or $\mathbb{A}^* = \{x \in \mathbb{X} | \alpha \leq e(x) \leq 1-\alpha\}$. The optimal value of the boundary point α is determined by the marginal distribution of the propensity score, and its calculation is straightforward. We then estimate this set by $\hat{\mathbb{A}} = \{x \in \mathbb{X} | \hat{\alpha} \leq \hat{e}(x) \leq 1-\hat{\alpha}\}$, and propose using any of the standard methods for estimation of, and inference for, average treatment effects, using only the observations with covariate values in this set, ignoring the uncertainty in the estimation of the set $\hat{\mathbb{A}}$.

3. Alternative estimands

3·1. Efficiency bounds

Section 3 contains the main results of the paper. We derive the subset of the covariate space that allows for the most precise estimation of the corresponding average treatment effect.

In this subsection we discuss some results on efficiency bounds for average treatment effects given strong ignorability and regularity conditions involving smoothness. Define the sample weighted average treatment effect,

$$\tau_{\mathrm{S},\omega} = \frac{\sum_{i=1}^{N} \tau(X_i)\omega(X_i)}{\sum_{i=1}^{N} \omega(X_i)},$$

with the weight function $\omega : \mathbb{X} \mapsto [0, \infty)$. The results in Hahn (1998), Robins & Rotnitzky (1995), and Hirano et al. (2003) imply that the efficiency bound for $\tau_{\mathrm{S},\omega}$ is

$$V_{\omega}^{\mathrm{eff}} = \frac{1}{E\left[\{\omega(X)\}^2\right]} E\left[\omega(X)^2 \left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\}\right]. \tag{7}$$

These papers also propose efficient estimators that are asymptotically linear with influence function

$$\psi_{\omega}(y, w, x) = \frac{\omega(x)}{E\left\{\omega(X)\right\}} \left\{ w\frac{y-\mu_1(x)}{e(x)} - (1-w)\frac{y-\mu_0(x)}{1-e(x)} \right\},$$

so that

$$\hat{\tau}_{\omega} = \tau_{\mathrm{S},\omega} + \frac{1}{N}\sum_{i=1}^{N} \psi_{\omega}(Y_i, W_i, X_i) + o_{\mathrm{p}}\left(N^{-1/2}\right),$$

and the efficiency bound is the variance of the influence function, $V_{\omega}^{\mathrm{eff}} = E\{\psi_{\omega}(Y, W, X)\}^2$.

3·2. The optimal subpopulation average treatment effect

We now consider the problem of selecting the estimand $\tau_{\mathrm{S},\omega}$, or equivalently, the weight function $\omega(\cdot)$, that minimizes the asymptotic variance in (7), within the set of estimands where the weight function $\omega(x)$ is an indicator function, $\omega(x) = 1_{x\in\mathbb{A}}$; in the working paper version of this paper we also consider the problem without

imposing this restriction. Formally, we choose an estimand $\tau_{\text{S},\mathbb{A}}$ by choosing the set $\mathbb{A} \subset \mathbb{X}$ that minimizes

$$V_{\mathbb{A}}^{\text{eff}} = \frac{1}{\{E\left(1_{X\in\mathbb{A}}\right)\}^2} E\left[1_{X\in\mathbb{A}}\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\}\right].$$

Defining $q(\mathbb{A}) = \text{pr}(X \in \mathbb{A}) = E(1_{X\in\mathbb{A}})$, we can write the objective function as

$$V_{\mathbb{A}}^{\text{eff}} = \frac{1}{q(\mathbb{A})} E\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\middle| X \in \mathbb{A}\right\}. \tag{8}$$

We seek $\mathbb{A} = \mathbb{A}^*$, that minimizes $V_{\mathbb{A}}^{\text{eff}}$ among all closed subsets $\mathbb{A} \subset \mathbb{X}$.

Focusing on estimands that average the treatment effect only over a subpopulation has two effects on the asymptotic variance, pusing it in opposite directions. First, by excluding units with covariate values outside the set $\mathbb{A}$, one reduces the effective sample size in expectation from N to $Nq(\mathbb{A})$. This will increase the asymptotic variance by a factor $1/q(\mathbb{A})$. Second, by discarding units with high values for $\sigma_1^2(X)/e(X) + \sigma_0^2(X)/\{1-e(X)\}$ one can lower the conditional expectation $E[\sigma_1^2(X)/e(X)+\sigma_0^2(X)/\{1-e(X)\}|X \in \mathbb{A}]$. Optimally choosing $\mathbb{A}$ involves balancing these two effects.

Theorem 1. *Suppose that Assumptions 1-2 hold, and suppose that the density of X is bounded, and bounded away from zero, and that the conditional variances of $Y_i(0)$ and $Y_i(1)$ are bounded. We consider $\tau_{\text{S},\mathbb{A}}$ where $\mathbb{A}$ is a closed subset of $\mathbb{X}$. Then the optimal subpopulation average treatment effect is $\tau_{\text{S},\mathbb{A}^*}$, where, if $\sup_{x\in\mathbb{X}} \sigma_1^2(x)/e(x) + \sigma_0^2(x)/\{1-e(x)\} \leq 2E\left[\sigma_1^2(X)/e(X) + \sigma_0^2(X)/\{1-e(X)\}\right]$, then $\mathbb{A}^* = \mathbb{X}$ and, otherwise,*

$$\mathbb{A}^* = \left\{x \in \mathbb{X}\middle| \frac{\sigma_1^2(x)}{e(x)} + \frac{\sigma_0^2(x)}{1-e(x)} \leq \frac{1}{\alpha(1-\alpha)}\right\},$$

where α is a solution to

$$\frac{1}{\alpha(1-\alpha)} = 2E\left[\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\middle| \frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)} < \frac{1}{\alpha(1-\alpha)}\right].$$

A sketch of the proof is given in the Appendix.

The result in this theorem simplifies in an interesting way under homoskedasticity. Let

$$V_{\text{H},\mathbb{A}} = \frac{1}{q(\mathbb{A})} E\left\{\frac{\sigma^2}{e(X)} + \frac{\sigma^2}{1-e(X)}\middle| X \in \mathbb{A}\right\}, \tag{9}$$

be the asymptotic variance under homoskedasticity.

COROLLARY 1. *Suppose that Assumptions 1-2 hold, and suppose that the density of X is bounded and bounded away from zero. Suppose also that $\sigma_w^2(x) = \sigma^2$ for all $w \in \{0,1\}$ and $x \in \mathbb{X}$. Then the optimal subpopulation average treatment effect is $\tau_{\text{S},\mathbb{A}_{\text{H}}^*}$, where*

$$\mathbb{A}_{\text{H}}^* = \{x \in \mathbb{X} \,|\alpha \leq e(x) \leq 1-\alpha\}\,.$$

If

$$\sup_{x\in\mathbb{X}} \frac{1}{e(x)\{1-e(x)\}} \leq 2E\left[\frac{1}{e(X)\{1-e(X)\}}\right],$$

then $\alpha = 0$ and $\mathbb{A}_{\text{H}}^* = \mathbb{X}$. Otherwise, α is a solution to

$$\frac{1}{\alpha(1-\alpha)} = 2E\left[\frac{1}{e(X)\{1-e(X)\}}\middle|\, \frac{1}{e(X)\{1-e(X)\}} \leq \frac{1}{\alpha(1-\alpha)}\right].$$

This is the key result in the paper. In practice it is more useful than the result in Theorem 1 for two reasons. First, the optimal set $\mathbb{A}_{\text{H}}^*$ depends only on the marginal distribution of the propensity score, and so its construction avoids potential biases associated with using outcome data. Second, the criterion in Corollary 1 is easier to implement because the propensity score can be precisely estimable, even in settings with little overlap, whereas the conditional variances that appear in the criterion in Theorem 1 may not be. Even when homoskedasticity does not hold, the optimal set according to this criterion may be a useful approximation.

To implement our proposed criterion, one would first estimate the propensity score. In the second step, one solve for the smallest value $\hat{\alpha} \in [0, 1/2]$ that satisfies

$$\frac{1}{\alpha(1-\alpha)} \leq 2\frac{\sum_{i=1}^N \left[1_{\hat{e}(X_i)\{1-\hat{e}(X_i)\}\geq\alpha(1-\alpha)} \big/ \hat{e}(X_i)\left\{1-\hat{e}(X_i)\right\}\right]}{\sum_{i=1}^N 1_{\hat{e}(X_i)\{1-\hat{e}(X_i)\}\geq\alpha(1-\alpha)}},$$

and use the set $\hat{\mathbb{A}} = \{x \in \mathbb{X} |\, \hat{\alpha} \leq \hat{e}(x) \leq 1-\hat{\alpha}\}$. Given this set $\hat{\mathbb{A}}$ one would use one of the standard methods for estimation of and inference for average treatment effects, such as those surveyed in Rosenbaum (2001) and Imbens (2004), ignoring the uncertainty in the estimation of $\hat{\mathbb{A}}$.

3·3. *Numerical simulations for optimal estimands when the propensity score follows a beta distribution*

In this section, we assess the implications of the results derived in the previous sections by presenting simulations for the optimal estimands, under homoskedasticity, when the true propensity score follows a Beta distribution. For a Beta distribution with parameters β and γ, the mean is $\beta/(\gamma+\beta) \in [0,1]$, and the variance is $\beta\gamma/\{(\gamma+\beta)^2(\gamma+\beta+1)\} \in [0,1/4]$. We focus on distributions for the true propensity score, with $\beta \in \{0.5,1,2,4\}$, and $\gamma \in \{\beta,\ldots,4\}$. For a given pair of values (β,γ), let $V_{\rm S}^{\rm eff}(\beta,\gamma)$ denote the asymptotic variance of the efficient estimator for the sample average treatment effect:

$$V_{\rm S}^{\rm eff}(\beta,\gamma) = \sigma^2 E\left\{ \frac{1}{e(X)} + \frac{1}{1-e(X)} \middle| e(X) \sim {\rm Be}(\beta,\gamma) \right\}.$$

In addition, let $V_{{\rm S},\alpha}^{\rm eff}(\beta,\gamma)$ denote the asymptotic variance for the efficient estimator for the sample average treatment effect, where we drop observations with the propensity score outside the interval $[\alpha,1-\alpha]$:

$$V_{{\rm S},\alpha}^{\rm eff}(\beta,\gamma) = \sigma^2 \frac{E\left\{ 1/\,e(X) + 1/\left\{1-e(X)\right\} \middle|\, \alpha \le e(X) \le 1-\alpha, e(X) \sim {\rm Be}(\beta,\gamma)\right\}}{{\rm pr}\left\{ \alpha \le e(X) \le 1-\alpha \middle|\, e(X) \sim {\rm Be}(\beta,\gamma)\right\}}$$

Let α^* denote the optimal cut-off value that minimizes $V_{{\rm S},\alpha}^{\rm eff}(\beta,\gamma)$. For each of the (β,γ) pairs we report in Table 1 the three ratios

$$\frac{V_{\rm S}(\beta,\gamma)}{V_{{\rm S},\alpha^*}(\beta,\gamma)}, \qquad \frac{V_{{\rm S},0.01}(\beta,\gamma)}{V_{{\rm S},\alpha^*}(\beta,\gamma)}, \qquad \frac{V_{{\rm S},0.10}(\beta,\gamma)}{V_{{\rm S},\alpha^*}(\beta,\gamma)}.$$

There are two main findings. First, the gain from trimming the sample can be substantial, reducing the asymptotic variance of the average treatment effect estimand by a factor of up to ten, depending on the distribution of the propensity score. Secondly, discarding observations with a propensity score outside the interval $[0.1,0.9]$ produces variances that are extremely close to those produced with optimally chosen cut-off values for the range of Beta distributions considered here. In contrast, using the smaller fixed cut-off value of 0.01 can lead to considerably larger variances than using the optimal cut-off value.

4. A re-analysis of data on right heart catheterization

Connors et al. (1996) used a propensity score matching approach to study the effectiveness of right heart catheterization in an observational setting, using data from the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. Right heart catheterization is a diagnostic procedure used for critically ill patients. The study collected data on hospitalized adult patients at 5 medical centres in the U.S.A. Based on information from a panel of experts, a rich set of variables relating to the decision to perform the right heart catheterization were collected, as well as detailed outcome data. Detailed information about the study and the nature of the variables can be found in Connors et al. (1996) and Murphy & Cluff (1990). Connors et al. (1996) found that, after adjustment for ignorable treatment assignment conditional on a range of covariates, right heart catheterization appeared to lead to lower survival rates. This conclusion contradicted popular perception among practitioners that right heart catheterization was beneficial. The primary analysis in Connors et al. (1996) matched treated and untreated patients on the basis of propensity score, with each unit matched at most once.

The study consists of data on 5735 individuals, 2184 of them assigned to the treatment group and the remaining 3551 assigned to the control group. For each individual we observe treatment status, equal to 1 if right heart catheterization was applied within 24 hours of admission, and 0 otherwise, the outcome, which is an indicator for survival at 30 days, and 72 covariates. For summary statistics on the 72 covariates see Connors et al. (1996) and Hirano & Imbens (2001). The two treatment groups differ on many of the covariates in statistically and substantially significant ways. We estimate the propensity score, using a logistic model that includes all 72 covariates. Hirano & Imbens (2001) study various methods for selecting subsets of the covariates. Figure 1 shows the distribution of estimated propensity scores. While the two groups obviously differ, the support of the estimated propensity scores in both groups is nearly the entire unit interval.

Based on the estimated propensity score, we calculate the optimal cut-off value α in Corrolary 1, obtaining $\hat{\alpha} = 0.1026$. Next, we consider three samples, (i) the full sample, (ii) the set of units with $\hat{e}(X_i) \in [0.1, 0.9]$, based on the 0.1 rule-of-thumb, (iii) the optimal set with $\hat{e}(X_i) \in [0.1026, 0.8974]$. In Table 2 we report the sample sizes by treatment status in the $[0.1, 0.9]$ dataset.

Next we estimate the average effect and its variance for each subsample. The specific estimator we use in each case is a version of the Horvitz-Thompson estimator; see Hirano et al. (2003) for details of the implementation. First, we re-estimate the propensity score on the selected sample, using the full set of 72 covariates. Then, we estimate the average treatment effect as

$$\hat{\tau} = \sum_{i=1}^{N} \frac{W_i Y_i}{\hat{e}(X_i)} \Big/ \sum_{i=1}^{N} \frac{W_i}{\hat{e}(X_i)} - \sum_{i=1}^{N} \frac{(1-W_i)Y_i}{1-\hat{e}(X_i)} \Big/ \sum_{i=1}^{N} \frac{1-W_i}{1-\hat{e}(X_i)}.$$

We estimate the standard errors using the bootstrap, given the sample selected. We use two specific estimators. First, we simply calculate the standard deviation of the B bootstrap replications. This estimator is denoted by se(1). Second, given the ordered B bootstrap estimates, we take the difference between the $0.95\times B$ and the $0.05\times B$ bootstrap estimates and divided this difference by 2×1.645 to obtain an estimate for the standard error. This estimator is denoted by se(2). Note that these standard error estimators do not impose homoskedasticity, which was only used in the construction of the optimal set. We use 50 000 bootstrap replications and Table 3 presents the results. For both the $[0.1, 0.9]$ sample and the optimal $[0.1026, 0.8974]$ sample, the variance drops to approximately 64% of the original variance. Thus, by dropping 18% of the sample we obtain a sizeable reduction in the variance of 36%. These results further strengthen the substantive conclusions in Connors et al. (1996) that right heart catheterization has negative effects on survival.

Acknowledgements

We are grateful for helpful comments by Richard Blundell, Gary Chamberlain, Jinyong Hahn, Gary King, Michael Lechner, Robert Moffitt, Geert Ridder, Don Rubin,

participants in many seminars, and the editor, an associate editor, and an anonymous referee. Financial support by the U.S. National Science Foundation is gratefully acknowledged.

Appendix

Proofs

Define

$$\tau_{\mathrm{S},\omega}(\mathbb{A}) = \sum_{i:X_i\in\mathbb{A}} \tau(X_i)\omega(X_i) \Big/ \sum_{i:X_i\in\mathbb{A}} \omega(X_i),$$

for functions $\omega(\cdot) : \mathbb{X} \mapsto [0,\infty)$. For estimands of this type consider the minimum asymptotic variance criterion that includes that considered in Theorem 1 as a special case:

$$V_{\mathrm{S},\omega}(\mathbb{A}) = \frac{E\left[\omega(X)^2 1_{X\in\mathbb{A}}\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\}\right]}{\left[E\left\{\omega(X)1_{X\in\mathbb{A}}\right\}\right]^2}. \tag{A1}$$

We are interested in the choice of set $\mathbb{A}$ that minimizes (A1) among the set of all closed subsets of $\mathbb{X}$. The following theorem provides the characterization. Let $f(\cdot)$ be the probability density function of the covariate X.

Theorem A1. *Suppose $f_{\mathrm{l}} \leq f(x) \leq f_{\mathrm{u}}$, and $\sigma^2(x) \leq \sigma^2_{\mathrm{u}}$ for $w = 0, 1$ and all $x \in \mathbb{X}$, and suppose $\omega : \mathbb{X} \mapsto [0,\infty)$ is continuously differentiable. Then the set $\mathbb{A}^*$ that minimizes (A1) is equal to $\mathbb{X}$ if*

$$\sup_{x\in\mathbb{X}} \omega(x)\left\{\frac{\sigma_1^2(x)}{e(x)} + \frac{\sigma_0^2(x)}{1-e(x)}\right\} \leq 2\frac{E\left[\omega^2(X)\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\}\right]}{E\left\{\omega(X)\right\}},$$

and, otherwise,

$$\mathbb{A}^* = \left\{x \in \mathbb{X} \,\middle|\, \omega(x)\left\{\frac{\sigma_1^2(x)}{e(x)} + \frac{\sigma_0^2(x)}{1-e(x)}\right\} \leq \gamma\right\},$$

where γ is a positive solution to

$$\gamma = 2\frac{E\left[\omega^2(X)\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\} \middle| \omega(X)\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\} < \gamma\right]}{E\left[\omega(X) \middle| \omega(X)\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\} < \gamma\right]}.$$

Sketch of proof: Define $k(x) = \sigma_1^2(x)/e(x) + \sigma_0^2(x)/\{1-e(x)\}$, $\tilde{f}_X(x) = f_X(x)\omega(x)/\int_z f_X(z)\omega(z)dz$, and $\tilde{\omega}(x) = \omega(x)/\int_z f_X(z)\omega(z)dz$, so that $k(x)$ is bounded away from zero and infinity, and continuously differentiable on $\mathbb{X}$. Let $\tilde{X}$ be a random vector with probability density function $\tilde{f}_X(x)$ on $\mathbb{X}$, and let $\tilde{q}(\mathbb{A}) = \text{pr}(\tilde{X} \in \mathbb{A})$. Then

$$E\left\{\tilde{\omega}(X)1_{X\in\mathbb{A}}\right\} = \text{pr}\left(\tilde{X} \in \mathbb{A}\right) = \tilde{q}(\mathbb{A}),$$

and, similarly,

$$E\left[\tilde{\omega}(X)^2 1_{X\in\mathbb{A}}\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\}\right] = E\left\{\tilde{\omega}(\tilde{X})1_{\tilde{X}\in\mathbb{A}}k(\tilde{X})\right\}.$$

Since multiplying $\omega(x)$ by a constant does not change the value of the objective function in (A1), we have

$$\begin{aligned} V_{\text{S},\omega}(\mathbb{A}) = V_{\text{S},\tilde{\omega}}(\mathbb{A}) &= \frac{1}{\left[E\left\{\tilde{\omega}(X)1_{X\in\mathbb{A}}\right\}\right]^2} E\left[\tilde{\omega}(X)^2 1_{X\in\mathbb{A}}\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\}\right] \\ &= \frac{1}{\tilde{q}(\mathbb{A})} E\left\{\tilde{\omega}(\tilde{X})k(\tilde{X})\Big| 1_{\tilde{X}\in\mathbb{A}}\right\}. \end{aligned} \tag{A2}$$

Thus the question now concerns the set $\mathbb{A}$ that minimizes (A2).

The remainder of the proof of Theorem A1 consists of two stages. First, suppose there is a closed set $\mathbb{A}$ such that $x \in \text{int}(\mathbb{A})$, $z \notin \mathbb{A}$, and $\tilde{\omega}(z)k(z) < \tilde{\omega}(x)k(x)$. Then we will construct a closed set $\tilde{\mathbb{A}}$ such that $V_{\text{S},\tilde{\omega}}(\tilde{\mathbb{A}}) < V_{\text{S},\tilde{\omega}}(\mathbb{A})$. This implies that the optimal set has the form

$$\mathbb{A}^* = \{x \in \mathbb{X} | \tilde{\omega}(x)k(x) \leq \delta\},$$

for some δ. The second step consists of deriving the optimal value for δ.

For the first step define a ball around x with volume ν,

$$\mathbb{B}_\nu(x) = \left\{ z \in \mathbb{X} |\, \|z - x\| \leq \nu^{1/L} 2^{-1/L} \pi^{-1/2} \Gamma(L/2)^{1/L} \right\}.$$

Now we construct the set

$$\mathbb{A}_\nu = \left\{\mathbb{A} \cap \mathbb{B}_{\nu/\tilde{f}_X(x)}(x)\right\} \cup \mathbb{B}_{\nu/\tilde{f}_X(z)}(z).$$

For small enough ν,

$$\begin{aligned} V_{\mathrm{S},\tilde{\omega}}(\mathbb{A}_\nu) - V_{\mathrm{S},\tilde{\omega}}(\mathbb{A}) \\ = \frac{\nu}{q(\mathbb{A})^2}\Big[E\left\{\tilde{\omega}(\tilde{X})k(\tilde{X})\Big|\,\tilde{X}\in\mathbb{B}_{\nu/\tilde{f}_X(z)}(z)\right\} \\ - E\left\{\tilde{\omega}(\tilde{X})k(\tilde{X})\Big|\,\tilde{X}\in\mathbb{B}_{\nu/\tilde{f}_X(x)}(x)\right\}\Big] + o(\nu). \end{aligned}$$

Since $E\{\tilde{\omega}(\tilde{X})k(\tilde{X})|\tilde{X}\in\mathbb{B}_{\nu/\tilde{f}_X(z)}(z)\} - E\{\tilde{\omega}(\tilde{X})k(\tilde{X})|\tilde{X}\in\mathbb{B}_{\nu/\tilde{f}_X(x)}(x)\} < 0$ if $\nu \leq \nu_0$, the difference $V_{\mathrm{S},\tilde{\omega}}(\mathbb{A}_\nu) - V_{\mathrm{S},\tilde{\omega}}(\mathbb{A})$ is negative for small enough ν, which finishes the first part of the proof.

The question now is to determine the optimal value for δ, given that the optimal set has the form

$$\mathbb{A}_\delta = \{x\in\mathbb{X}|\tilde{\omega}(x)k(x)\leq\delta\}.$$

Define the random variable $Y = \tilde{\omega}(\tilde{X})k(\tilde{X})$, with probability density function $f_Y(y)$. Then

$$V_{\mathrm{S},\tilde{\omega}}(\mathbb{A}_\delta) = \frac{\int_0^\delta y f_Y(y)dy}{\left\{\int_0^\delta f_Y(y)dy\right\}^2}.$$

Either $V_{\mathrm{S},\tilde{\omega}}(\mathbb{A}_\delta)$ is minimized at $\delta = \sup_{x\in\mathbb{X}}k(x)$, or there is an interior minimum where the first order conditions are satisfied. The latter implies that

$$\delta = 2E\left\{\tilde{\omega}(\tilde{X})k(\tilde{X})\Big|\,\tilde{\omega}(\tilde{X})k(\tilde{X})<\delta\right\},$$

and thus

$$\gamma = 2E\left\{\omega(\tilde{X})k(\tilde{X})\Big|\,\omega(\tilde{X})k(\tilde{X})<\gamma\right\},$$

for $\gamma = \delta\int\omega(x)f_X(x)dx$. This in turn implies that

$$\gamma = 2\frac{E\left\{\omega^2(X)k(X)\,|\omega(X)k(X)<\gamma\right\}}{E\left\{\omega(X)\,|\omega(X)k(X)<\gamma\right\}}.$$

If we substitute back $k(x) = \sigma_1^2(x)/e(x) + \sigma_0^2(x)/\{1-e(x)\}$ this implies

$$\gamma = 2\frac{E\left[\omega^2(X)\left\{\frac{\sigma_1^2(X)}{e(X)}+\frac{\sigma_0^2(X)}{1-e(X)}\right\}\Big|\omega(X)\left\{\frac{\sigma_1^2(X)}{e(X)}+\frac{\sigma_0^2(X)}{1-e(X)}\right\}<\gamma\right]}{E\left[\omega(X)\Big|\omega(X)\left\{\frac{\sigma_1^2(X)}{e(X)}+\frac{\sigma_0^2(X)}{1-e(X)}\right\}<\gamma\right]},$$

as desired. □

Proof of Theorem 1. Substituting $\omega(x) = 1$ into Theorem A1 implies that the optimal set $\mathbb{A}^*$ is equal to $\mathbb{X}$ if

$$\sup_{x\in\mathbb{X}} \frac{\sigma_1^2(x)}{e(x)} + \frac{\sigma_0^2(x)}{1-e(x)} \leq 2E\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right\},$$

and, otherwise,

$$\mathbb{A}^* = \left\{x \in \mathbb{X} \,\middle|\, \frac{\sigma_1^2(x)}{e(x)} + \frac{\sigma_0^2(x)}{1-e(x)} \leq \gamma\right\},$$

where γ is a positive solution to

$$\gamma = 2E\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)} \,\middle|\, \frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)} < \gamma\right\}.$$

Then define $\alpha = 1/2 - (1/4 - 1/\gamma)^{1/2}$ so that $\gamma = (\alpha(1-\alpha))^{-1}$ and

$$\mathbb{A}^* = \left\{x \in \mathbb{X} \,\middle|\, \frac{\sigma_1^2(x)}{e(x)} + \frac{\sigma_0^2(x)}{1-e(x)} \leq \frac{1}{\alpha(1-\alpha)}\right\},$$

where α is a positive solution to

$$\frac{1}{\alpha(1-\alpha)} = 2E\left\{\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)} \,\middle|\, \frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)} < \frac{1}{\alpha(1-\alpha)}\right\}. \qquad \square$$

References

Abadie, A. & Imbens, G. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica.* **74**, 235-67.

Connors, A. F., Speroff T., Dawson, N. V., Thomas, C., Harrell, F. E., Wagner, D., Desbiens, N., Goldman, L., Wu, A. W., Califf, R. M., Fulkerson, W. J., Vidaillet, H., Broste, S., Bellamy, P., Lynn, J. & Knaus, W. A. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. *J. Am. Med. Assoc.* **276**, 889-97.

Grzybowski, M., Clements, E. A., Parsons, L., Welch, R., Tintinalli, A. T. & Ross, M. A. (2003). Mortality benefit of immediate revascularization of acute ST-segment elevation myocardial infarction in patients with contraindications to thrombolytic therapy: A propensity analysis. *J. Am. Med. Assoc.* **290**, 1891-98.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*. **66**, 315-31.

Heckman, J., Ichimura, H. & Todd, P. (1998). Matching as an econometric evaluation estimator. *Rev. Econ. Stud.* **65**, 261-94.

Hirano, K. & Imbens, G. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Hlth. Serv. Out. Res. Meth.* **2**, 259-78.

Hirano, K., Imbens, G. & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. **71**, 1161-89.

Horvitz, D. & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Assoc.* **46**, 663-85.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86**, 1-29.

Murphy, D. J. & Cluff, L. E. (1990). SUPPORT: Study to understand prognoses and preferences for outcomes and risks of treatments. *J. Clin. Epidemiol.* **43**, S1-S123.

Robins, J.M. & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Am. Statist. Assoc.* **90**, 122-29.

Rosenbaum, P. (1989). Optimal matching in observational studies. *J. Am. Statist. Assoc.* **84**, 1024-32.

Rosenbaum, P. (2001). *Observational Studies*, 2nd ed. New York: Springer.

Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*. **70**, 41-55.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688-701.

Rubin, D. (1997). Estimating causal effects from large data sets using propensity scores. *Ann. Intern. Med.* **127**, 757-63.

Vincent, J. L., Baron, J., Reinhart, K., Gattinoni, L., Thijs, L., Webb, A., Meier-Hellmann, A., Nollet, G. & Peres-Bota, D. (2002). Anemia and blood transfusion in critically ill patients. *J. Am. Med. Assoc.* **288**, 1499-1507.

Table 1: Variance ratios for Beta distributions

		γ			
β		0.5	1.0	2.0	4.0
0.5	$V_S(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$	13.38	11.68	13.71	12.83
	$V_{S,0.01}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$	1.70	1.64	1.71	1.58
	$V_{S,0.10}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$	1.00	1.00	1.00	1.04
1.0	$V_S(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$		2.68	2.65	3.36
	$V_{S,0.01}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$		1.39	1.39	1.47
	$V_{S,0.10}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$		1.00	1.00	1.01
2.0	$V_S(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$			1.11	1.16
	$V_{S,0.01}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$			1.09	1.12
	$V_{S,0.10}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$			1.00	1.00
4.0	$V_S(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$				1.02
	$V_{S,0.01}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$				1.02
	$V_{S,0.10}(\gamma,\beta)/V_{S,\alpha(\gamma,\beta)}(\gamma,\beta)$				1.00

Table 2: Subsample sizes for Right Heart Catheterization data

	$\hat{e}(X_i) < 0.1$	$0.1 \leq \hat{e}(X_i) \leq 0.9$	$0.9 < \hat{e}(X_i)$	all
controls	870	2671	10	3551
treated	40	2057	87	2184
All	910	4728	97	5735

Table 3: Estimates for Average Treatment Effects in Right Heart Catheterization study

	Estimate	se(1)	se(2)
Full Sample	-0.0593	0.0166	0.0167
$\hat{e}(X_i) \in [0.1, 0.9]$	-0.0590	0.0143	0.0143
$\hat{e}(X_i) \in [0.1026, 0.8974]$	-0.0601	0.0143	0.0144

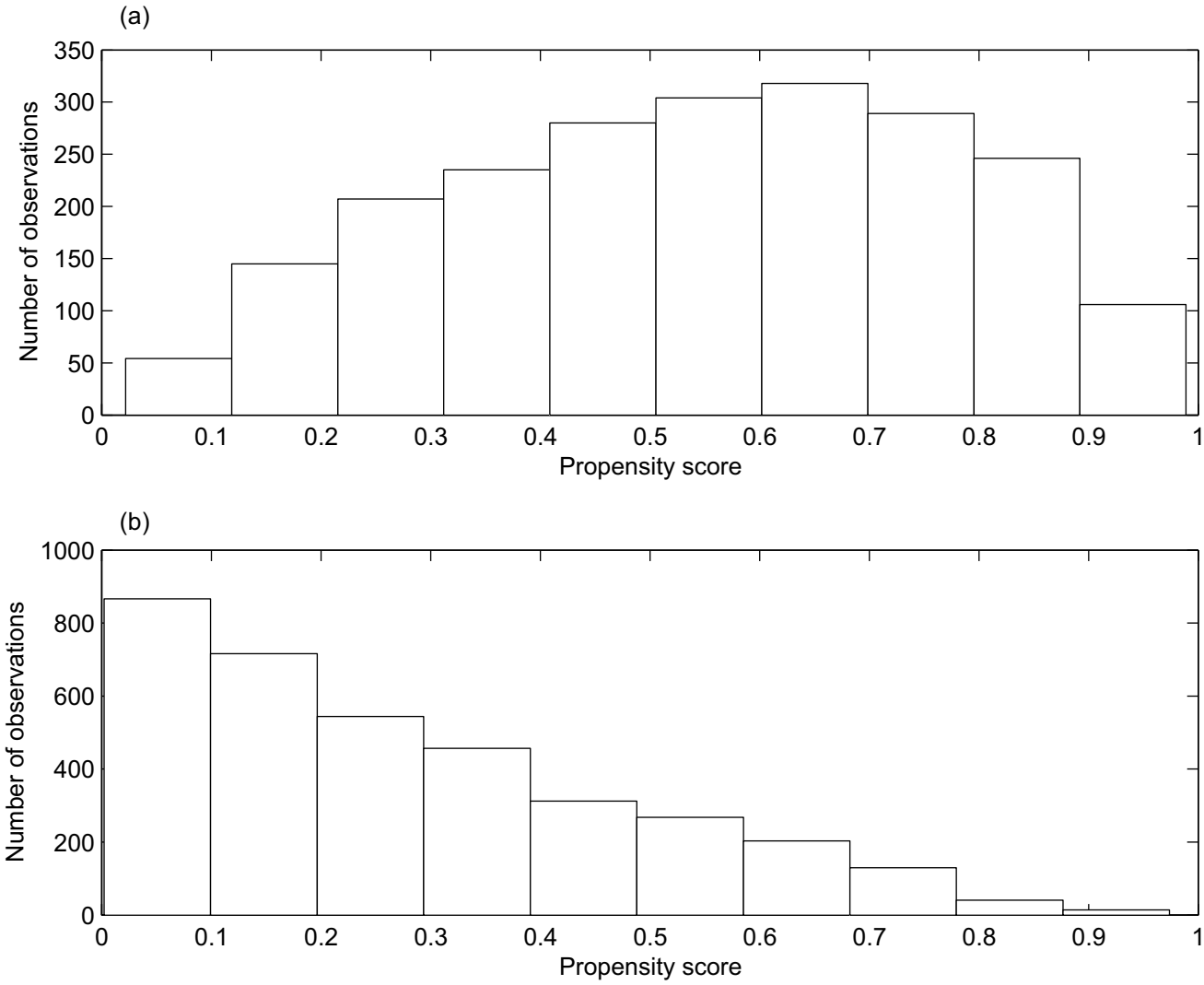

Fig. 1. Right heart catheterization study. Propensity score using all covariates for (a) treated and (b) control patients