Statistics and Probability Letters 110 (2016) 181–184

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bias reduction with Variable Percent Bias Reducing matching

Yannis G. Yatracos

Faculty of Communication and Media Studies, Cyprus U. of Technology, Cyprus

ARTICLE INFO

Article history:
Received 24 September 2015
Received in revised form 24 December 2015
Accepted 24 December 2015
Available online 4 January 2016

Keywords:
Bias reduction
Matching

ABSTRACT

The bias in linear combinations, $\mathbf{a}'\mathbf{X}$, of treatment and control $\mathbf{X}$-covariates obtained with EPBR matching can be reduced when all $\mathbf{a}$-coordinates have same sign, using Variable Percent Bias Reducing (VPBR) matching for selected $\mathbf{X}$-covariates' subgroups. For specified subsets of $\mathbf{a}$-values, VPBR matching reduces the absolute bias.

1. Introduction

In observational studies for causal effects it is important to determine a control group with $\mathbf{X}$-covariates "matching" those of the treatment group, such that the covariates' means in the groups have minimum bias; $\mathbf{X} \in R^d$. Without exact matching the bias will increase for some linear functions, $\mathbf{a}'\mathbf{X}$, unless the matching is equal percent bias reducing (EPBR), i.e. the percentage of bias reduction with respect to random matching is fixed, $\gamma(<1)$, *for each covariate* (see, e.g., Rubin, 1973; Rosenbaum and Rubin, 1983); $\mathbf{a} \in R^d$. However, when EPBR property holds, (i) γ increases with $\mathbf{X}$'s dimension, d, (ii) no special attention can be given to a covariate with large bias we wish to reduce more since γ is fixed, (iii) the possibility of reducing the absolute bias is not addressed, and (iv) $\mathbf{X}$-covariates have to be "similar" in some sense, e.g., exchangeable (Rubin and Thomas, 1996; Rubin and Stuart, 2006).

In this work problems (i) and (ii) are solved when $\mathbf{a}$'s coordinates in $\mathbf{a}'\mathbf{X}$ have the same sign, by introducing variable percent bias reducing (VPBR) matching, i.e., either EPBR matchings for subgroups of the covariates with different bias reducing coefficients, or completely different matchings, one for each covariate. Problem (iii) is also addressed when the bias reducing coefficients of the VPBR-subgroups are smaller than γ because the *absolute* bias of linear functions, $\mathbf{a}'\mathbf{X}$, is reduced for $\mathbf{a}$'s in specific R^d-subsets. Thus, VPBR matching generalizes the *one-for-all* covariates EPBR matching, complements it when leading to bias reduction, improves it when the $\mathbf{a}$'s coordinates have the same sign and motivates comparisons of absolute biases.

Moment conditions allowing for EPBR matching are presented in Yatracos (2012a) and relax, at least partially, restriction (i) that rarely holds with real data. The recently proposed Monotonic Imbalance Bounding (MIB) matching (Iacus et al., 2011) does not generalize EPBR matching and has flaws (Yatracos, 2012b).

2. EPBR matching

Let Z be the assignment-to-treatment variable, $Z = 1$ for treatment and $Z = 0$ for control. Let $\mathbf{B}$ be the bias in $\mathbf{X}(\in R^d)$ with random sampling in the control group,

$$\mathbf{B} = (B_1, \ldots, B_d) = E(\mathbf{X}|Z = 1) - E(\mathbf{X}|Z = 0), \tag{1}$$

E-mail address: yannis.yatracos@cut.ac.cy.

http://dx.doi.org/10.1016/j.spl.2015.12.027
0167-7152/

and let $\mathbf{B}_m$ be the bias with a matched control group,

$$\mathbf{B}_m = (B_{m,1}, \ldots, B_{m,d}) = E(\mathbf{X}|Z = 1) - E_m(\mathbf{X}|Z = 0); \tag{2}$$

E_m denotes expectation with respect to the distribution in matched samples. The EPBR property (Rubin, 1973),

$$\mathbf{B}_m = \gamma \mathbf{B}, \quad 0 < \gamma < 1, \tag{3}$$

implies that for any $\mathbf{a} \in R^d$

$$\frac{\mathbf{a}'\mathbf{B}_m}{\mathbf{a}'\mathbf{B}} = \gamma = \left|\frac{\mathbf{a}'\mathbf{B}_m}{\mathbf{a}'\mathbf{B}}\right|. \tag{4}$$

From (3), it follows that the ith coordinates of $\mathbf{B}$ and $\mathbf{B}_m$ have the same sign, $i = 1, \ldots, d$. Coefficient γ is the bias reducing coefficient of EPBR matching and the smaller its value is, the better the matching's quality is. In applications, it is expected to use the best available EPBR matching.

Definition 2.1. The acceptable *EPBR* matching m for all the $\mathbf{X}$-covariates has the smallest attainable γ-value in (3) denoted w.l.o.g. by γ.

Lemma 2.1. *The γ-value of the acceptable EPBR matching does not decrease when the number of covariates d is increased to $d + k$.*

Remark 2.1. From Lemma 2.1 it follows that when d is large EPBR property becomes less interesting. A potential remedy is used herein, i.e. EPBR matching for subgroups of the $\mathbf{X}$-coordinates.

From (3), without loss of generality assume (otherwise replace X_j by $-X_j$)

$$B_j > 0, \quad j = 1, \ldots, d. \tag{5}$$

Practitioners are also interested in the reduction in absolute bias, i.e. instead of (4) having

$$\left|\frac{\mathbf{a}'\mathbf{B}_m}{\mathbf{a}'\mathbf{B}}\right| < \gamma, \tag{6}$$

at least for a particular subset of $\mathbf{a}$'s in R^d. It is shown herein that this can be achieved in some situations.

3. VPBR matching

Lemma 2.1 suggests that different EPBR matchings for covariates' subsets may have smaller bias reducing coefficients than the γ-value of the acceptable EPBR m matching for all the covariates. The goal of such variable percent bias reducing (VPBR) matching is to achieve for the covariates bias $\mathbf{B}_v$ instead of $\mathbf{B}_m$ such that for every $\mathbf{a}$ *in a subset of* R^d, either the bias $\mathbf{a}'\mathbf{B}_v$ has the same sign with $\mathbf{a}'\mathbf{B}_m$ (as in EPBR matching) but is closer to 0 or, more generally,

$$\left|\frac{\mathbf{a}'\mathbf{B}_v}{\mathbf{a}'\mathbf{B}}\right| < \gamma. \tag{7}$$

Definition 3.1. A variable percent bias reducing (VPBR) matching consists of acceptable EPBR matchings for subgroups of the $\mathbf{X}$-covariates, with at least one bias reducing coefficient smaller than γ of the m matching.

The next proposition confirms that VPBR matching can improve the acceptable EPBR matching m for particular $\mathbf{a}$-subsets.

Proposition 3.1. *Consider the acceptable EPBR m matching for all the covariates with coefficient γ, assume that (5) holds and let $Y = \mathbf{a}'\mathbf{X}$, $\mathbf{a} \in R^d$.*

(a) *When all the coordinates of $\mathbf{a}$ have the same sign, consider acceptable EPBR matchings for the first p and the last $(d - p)$ $\mathbf{X}$-covariates with bias $\mathbf{B}_v = (\mathbf{B}_{v_1}, \mathbf{B}_{v_2})$ and bias reducing coefficients in $\mathbf{B}_{v_1}$ and $\mathbf{B}_{v_2}$, respectively, γ_1 and γ_2, $\gamma_1 \neq \gamma_2$. For this VPBR matching,*

(a_1)

$$\frac{\mathbf{a}'\mathbf{B}_v}{\mathbf{a}'\mathbf{B}} < \gamma, \quad \textit{and}$$

(a_2) *if, in addition,*

$$\gamma_{\max} = \max\{\gamma_1, \gamma_2\} < \gamma, \tag{8}$$

then

$$\frac{\mathbf{a}'\mathbf{B}_v}{\mathbf{a}'\mathbf{B}} < \gamma_{\max}.$$