Ann Inst Stat Math (2013) 65:69–87
DOI 10.1007/s10463-012-0358-9

Equal percent bias reduction and variance proportionate modifying properties with mean–covariance preserving matching

Yannis G. Yatracos

Received: 15 November 2010 / Revised: 4 January 2012 / Published online: 28 April 2012

Abstract Mean-preserving and covariance preserving matchings are introduced that can be obtained with conditional, randomized matching on sub-populations of a large control group. Under moment conditions it is shown that these matchings are, respectively, equal percent bias reducing (EPBR) and variance proportionate modifying (PM) for linear functions of the covariates and their standardizations. The results provide additional insight into and theory for EPBR and PM properties and confirm empirical and simulation findings that matchings can have the EPBR and PM properties also when the covariates are not exchangeable, or the treatment means are not equal.

Keywords Discriminant matching · Equal percent bias reducing · Mean–covariance preserving matching · Variance proportionate modifying matching

1 Introduction

In an observational (non-randomized) study with the objective evaluation of a treatment's effect, the use of random samples from the treated and control populations may cause estimation bias. Matched sampling on covariates is a popular technique for controlling the bias. It all started with "discriminant matching" (Cochran and Rubin 1973; Rubin 1970, 1973a,b) that evolved into propensity matching used in causal inference (Rosenbaum and Rubin 1983, 1984, 1985). Related theory with applications and extensions for matching and causal inference for two or more treatments are presented, among others, in Rubin and Thomas (1992, 1996), Joffe and Rosenbaum (1999), Imbens (2000), Imai and Van Dyk (2004) and Rubin and Stuart (2006). Stuart

Y. G. Yatracos (✉)
School of Management and Economics, Cyprus University of Technology,
P. O. Box 50329, 3603 Lemesos, Cyprus
e-mail: yannis.yatracos@cut.ac.cy

(2010, p. 15) provides a review on matching methods for causal inference suggesting future research directions one of which is covariate balancing for multiple treatments. Yatracos (2011) presents causal inference for multiple treatments with *s*-matching obtained via sufficiency.

Without exact matching, the bias will increase for some linear functions $Y = \boldsymbol{\alpha}'\mathbf{X}$ of the covariates $\mathbf{X}$ unless the matching method is equal percent bias reducing (EPBR), i.e. the bias in each coordinate of $\mathbf{X}$ is reduced by the same percentage (Rubin 1976a,b); $\mathbf{X} \in R^p$, $\boldsymbol{\alpha} \in R^p$. This makes EPBR a desirable property in matched sampling.

Matched sampling on covariates is ρ^2-proportionate modifying of the variance (PM-1) of the matched sample means' difference $\bar{Y}_{mt} - \bar{Y}_{mc}$ of the Y's, when it allows to express for *all* $Y = \boldsymbol{\alpha}'\mathbf{X}$, with $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$, the variances' ratio $\frac{\mathrm{Var}(\bar{Y}_{mt}-\bar{Y}_{mc})}{\mathrm{Var}(\bar{Y}_{rt}-\bar{Y}_{rc})}$ as weighted sum of the same variances' ratio but with $\mathbf{X}$ projected, respectively, along the best linear discriminant and its uncorrelated covariate; the weights are ρ^2 and $1-\rho^2$, $\bar{Y}_{rt} - \bar{Y}_{rc}$ is the random sample means' difference and t and c denote, respectively, treatment and control. Arguments for the desirability of PM-1 property appear in Rubin and Thomas (1996, Section 2.1, p. 251) where it is mentioned that when $\bar{Y}_{mt} = \bar{Y}_{rt} = \bar{Y}_t$ and for a specific setting that implies also EPBR matching, "The entire differential effect of the matching on different Y variables is determined by ρ^2, the effect of matching being the same for all discriminant's uncorrelated covariates." In addition, analytic approximations are obtained for $\frac{\mathrm{Var}(\bar{Y}_t-\bar{Y}_{mc})}{\mathrm{Var}(\bar{Y}_t-\bar{Y}_{rc})}$ using equations (2.7) and (2.8) in Rubin and Thomas (1996, p. 252). Analogous results hold for ρ^2-proportionate modifying matching of the expectations of the sample variances (PM-2) and make PM-2 property desirable; see Rubin and Thomas (1996, Section 2.1, p. 252).

Rubin and Thomas (1992) (hereafter R&T) showed that when the distribution of $\mathbf{X}$ is proportional ellipsoidally symmetric (PES), affine invariant matching methods are EPBR, PM-1 and PM-2. Rubin and Stuart (2006) (hereafter R&S) extended these results when $\mathbf{X}$ follows a "discriminant mixture of proportional ellipsoidally symmetric" (DPEMS) distribution. Crucial for the results is that PES and DPEMS distributions allow for the reduction of the covariates models in a canonical form with one common linear transformation and the intention to use affine invariant matching methods (R&T, p. 1081, l. −1 to −6). However, it has been noticed that

(a) when $\mathbf{X}$ follows a PES distribution and the number of covariates is large, exchangeability restricts the covariances to be non-negative, and
(b) when the propensity score or the linear discriminant are used for matching in examples and simulations with various other distributions, EPBR, PM-1 and PM-2 properties still hold (R&T, R&S).

EPBR and PM are moment properties and moment conditions should be sufficient for these to hold. In Sect. 2, simulations indicate that EPBR and PM properties hold, respectively, for randomly obtained exponential covariates and their standardizations (with respect to the means and variances of the control population), when the matched control covariates are obtained using a uniform distribution in a neighborhood of each of the observed exponential treatment covariates. In Sect. 3, mean-preserving (MP), covariance preserving (CP) and mean and covariance preserving (MCP) matchings are defined and matching Lemma 1 is provided that is used to obtain MP and CP matchings. In Sect. 4, it is shown that with MP-matching as in the simulation example, mild moment conditions on the treatment populations are necessary and sufficient for the the EPBR property to hold for linear functions of $\mathbf{X}$. In Sect. 5, it is shown that moment conditions on the populations and CP-matching are sufficient for the PM-properties to hold for standardized covariates $\mathbf{X}^*$ and in some cases also for $\mathbf{X}$. These results provide additional insight into the EPBR and PM properties and explain why the EPBR property holds often. With MCP-matching, that is both MP and CP, EPBR and PM properties hold for $\mathbf{X}^*$, and the results for $\mathbf{X}$ in R&T and R&S are obtained as special case. The theoretical results suggest sampling from sub-populations of the control populations to obtain MP and CP matchings via matching Lemma 1. A discussion follows in Sect. 6 on the Lemma's assumptions. The proofs are in the Appendix.

The goal of matched sampling is pairing of treatment and control units which are similar with respect to covariates, with no concern for the difference between these covariates' distribution functions. In machine learning, the distribution functions of the training and test data differ arbitrarily, causing the covariate shift problem (see, for example, Shimodaira 2000; Sugiyama et al. 2008). In classification problems, for example, the covariate shift problem affects the minimum value of the risk used to determine a classifier, as well as the classifier's choice (Bickel et al. 2007). The introduction of distributional matching methods, with goal similarity of the distributions of control and treatment populations, will be helpful in reducing the effect of covariate shift.

2 The motivating examples

The definitions of EPBR and PM properties follow, preceding the examples that indicate moment conditions are sufficient for these to hold. For matched and random samples of U's denote, respectively, their means $\bar{U}_{ms}$ and $\bar{U}_{rs},\ s = t, c$.

Definition 1 With the notation in the introduction, let $Y = \boldsymbol{\alpha}'\mathbf{X},\ \boldsymbol{\alpha} \in R^p$.

(a) The matching is EPBR (Rubin 1976a) if

$$\frac{E(\bar{Y}_{mt} - \bar{Y}_{mc})}{E(\bar{Y}_{rt} - \bar{Y}_{rc})}$$

is independent of $\boldsymbol{\alpha} \in R^p$.

(b) The matching is ρ^2-proportionate modifying of the variance of the difference in matched sample means (PM-1, R&T) if

$$\frac{\mathrm{Var}(\bar{Y}_{mt} - \bar{Y}_{mc})}{\mathrm{Var}(\bar{Y}_{rt} - \bar{Y}_{rc})} = \rho^2\tilde{V}_1 + (1-\rho^2)\tilde{V}_2,$$

with only ρ depending on $\boldsymbol{\alpha},\ 0 \le \rho^2 \le 1$. $\tilde{V}_1$ and $\tilde{V}_2$ are numbers taking the same two values for all Y.

In R&T, Y_{mt} has correlation ρ with the best, standardized linear discriminant D of $\mathbf{X},\ W$ is a standardized linear combination of $\mathbf{X}$ uncorrelated with D and

$$\tilde{V}_1 = \frac{\mathrm{Var}(\bar{D}_{mt} - \bar{D}_{mc})}{\mathrm{Var}(\bar{D}_{rt} - \bar{D}_{rc})}, \quad \tilde{V}_2 = \frac{\mathrm{Var}(\bar{W}_{mt} - \bar{W}_{mc})}{\mathrm{Var}(\bar{W}_{rt} - \bar{W}_{rc})}.$$

(c) The matching is ρ^2-proportionate modifying of the expectations of the sample variances (PM-2, R&T) if

$$\frac{Ev_{ms}(Y)}{Ev_{rs}(Y)} = \rho^2 \tilde{V}_{1,s} + (1-\rho^2)\tilde{V}_{2,s}, \quad s = t,\ c,$$

with only ρ depending on $\boldsymbol{\alpha}$, $0 \le \rho^2 \le 1$; $v_{mt},\ v_{mc},\ v_{rt}$ and v_{rc} are, respectively, the sample variances of n_t and n_c matched and randomly chosen treated and control units using $(n_t - 1)$ and $(n_c - 1)$ in the denominators. $\tilde{V}_{1,t},\ \tilde{V}_{2,t},\ \tilde{V}_{1,c},\ \tilde{V}_{2,c}$ are numbers taking the same four values for all Y.
In R&T,

$$\tilde{V}_{1,s} = \frac{Ev_{ms}(D)}{Ev_{rs}(D)}, \quad \tilde{V}_{2,s} = \frac{Ev_{ms}(W)}{Ev_{rs}(W)}, \ s = t, c.$$

2.1 EPBR simulation example

Althauser and Rubin (1970) use a subset V at ϵ-distance from the observed treatment covariates to obtain matched control covariates. Rosenbaum and Rubin (1985) construct V using the propensity score and obtain via Mahalanobis distance matching control covariates with the EPBR property. We also use in this example a neighborhood V of the observed treatment's covariates $\mathbf{X}_{mt} = \mathbf{x}_{mt}$, and obtain the matching control covariates $\mathbf{X}_{mc}$ via any distribution P on V with mean $\mathbf{x}_{mt} + z(\mathbf{x}_{mt} - \bar{\mathbf{x}}_c)$; $z \in R$, $\bar{\mathbf{x}}_c$ is a sample average from the control population used in random sampling without matching. It is assumed that $\mathbf{X}_{mc}$ takes values in the control population.

Let $Y = \boldsymbol{\alpha}'\mathbf{X}$, and let $\bar{Y}_{rt,\boldsymbol{\alpha}},\ \bar{Y}_{rc,\boldsymbol{\alpha}}$ be Y-averages obtained, respectively, from the random samples of treatment and control covariates. Let $\bar{Y}_{mt,\boldsymbol{\alpha}},\ \bar{Y}_{mc,\boldsymbol{\alpha}}$ be Y-averages obtained from the samples of matched covariates. The EPBR property holds when for *any* $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$, the ratios difference

$$\frac{E(\bar{Y}_{mt,\boldsymbol{\alpha}_1} - \bar{Y}_{mc,\boldsymbol{\alpha}_1})}{E(\bar{Y}_{rt,\boldsymbol{\alpha}_1} - \bar{Y}_{rc,\boldsymbol{\alpha}_1})} - \frac{E(\bar{Y}_{mt,\boldsymbol{\alpha}_2} - \bar{Y}_{mc,\boldsymbol{\alpha}_2})}{E(\bar{Y}_{rt,\boldsymbol{\alpha}_2} - \bar{Y}_{rc,\boldsymbol{\alpha}_2})} = 0. \tag{1}$$

In R&T and R&S it is shown that (1) holds for PES and DPEMS distributions.

A sample with size 200 of 2-dimensional, independent, random, treatment and control covariates are obtained, respectively, from exponential distributions with parameters $\theta_1 = (5.211, 3.215)$ and $\theta_2 = (4.019, 4.295)$; the parameters are absolute values of observed, independent normal variables with mean 5 and variance 4. For the matched samples, 200 treatment covariates $(X_{mt,1}, X_{mt,2})$ are obtained from an exponential distribution with parameter θ_1 and *conditionally* on $X_{mt,i} = x_{mt,i}$, the matching control variable $X_{mc,i}$ is the average of 100 independent random variables obtained from a uniform distribution on the interval

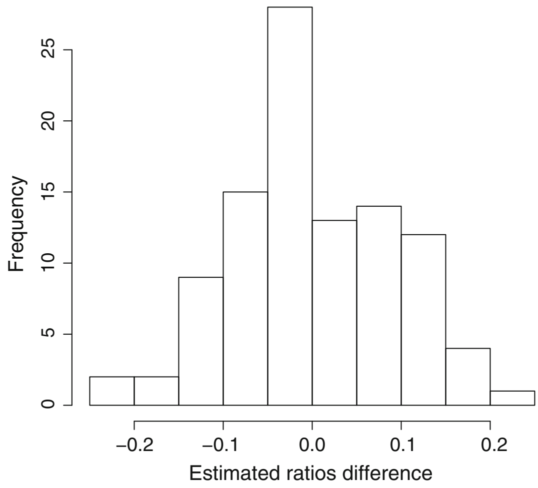


Fig. 1 Simulations supporting the EPBR property

$$(x_{mt,i} + z(x_{mt,i} - \bar{x}_{c,i}) - 1, x_{mt,i} + z(x_{mt,i} - \bar{x}_{c,i}) + 1), \quad i = 1, 2; \tag{2}$$

$z = -0.548231$ is obtained from a standard normal distribution, $\bar{x}_{c,i}$ is the average of a random sample of size 500 from the control population.

Values $\boldsymbol{\alpha}_1 = (7.673, 8.781)$, $\boldsymbol{\alpha}_2 = (8.486, 8.952)$ are obtained independently from a normal distribution with mean 10 and standard deviation 4. We obtain 300 samples of the covariates, calculate $\bar{Y}_{mt,\boldsymbol{\alpha}_i}$, $\bar{Y}_{mc,\boldsymbol{\alpha}_i}$, $\bar{Y}_{rt,\boldsymbol{\alpha}_i}$, $\bar{Y}_{rc,\boldsymbol{\alpha}_i}$, $i = 1, 2$, for each sample and their averages are used to estimate the expectations in (1). The procedure is repeated 100 times, to obtain in Fig. 1 the histogram of estimated values of the ratios difference in the left-hand side of (1) that have mean -0.0013 and variance 0.0081. The results support the EPBR property (1) of the matching.

2.2 Variance PM simulation example

Treatment and control samples with size 200 of 2-dimensional, independent exponential random variables are obtained with parameters, respectively, (1.740, 1.740) and (1, 1). The samples' coordinates are standardized in the same way to obtain covariates $\mathbf{X}_t^*$, $\mathbf{X}_c^*$ with $\mathbf{X}_c^*$ having coordinates with mean zero and variance one; t denotes treatment and c denotes control. The matched treatment covariates $\mathbf{X}_{mt}^*$ and $\mathbf{X}_t^*$ are obtained from the same population. Conditionally on $X_{mt,i}^* = x_{mt,i}^*$, the matching control variable $X_{mc,i}^*$ is the average of 20, independent random variables obtained from a uniform distribution on the interval $(x_{mt,i}^* + z x_{mt,i}^* - \sqrt{3}, x_{mt,i}^* + z x_{mt,i}^* + \sqrt{3})$, $i = 1, 2$; $z = 1.629$ is obtained from a standard normal distribution.

For $\mathbf{X}^* = \mathbf{X}_{mt}^*$ we can write when $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$,

$$Y = \boldsymbol{\alpha}'\mathbf{X}^* = \rho_{mt}\Delta_{mt}^* + \sqrt{1 - \rho_{mt}^2}\, W_{mt}^*, \tag{3}$$

where

$$\Delta^*_{mt} = \boldsymbol{\mu}^{*'}_{mt}\mathbf{X}^*/||\boldsymbol{\mu}^*_{mt}||, \quad \rho_{mt} = \boldsymbol{\alpha}'\boldsymbol{\mu}^*_{mt}/||\boldsymbol{\mu}^*_{mt}||, \quad W^*_{mt} = \boldsymbol{\gamma}'_{mt}\mathbf{X}^*;$$

Δ^*_{mt} is a linear discriminant, $\boldsymbol{\mu}^*_{mt}$ is the mean of the standardized treatment population, $\boldsymbol{\gamma}_{mt}$ is a norm 1 vector orthogonal to $\boldsymbol{\mu}^*_{mt}$ on the plane determined by $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}^*_{mt}$, $||\mathbf{x}||$ is the usual Euclidean norm of $\mathbf{x}$. A similar decomposition to (3) holds for $\mathbf{X}^* = \mathbf{X}^*_{mc}$ and for the random covariates $\mathbf{X}^* = \mathbf{X}^*_{rt}, \mathbf{X}^*_{rc}$.

In this set-up, Definition 1 (b), (c) is satisfied and the matching is:

(a) ρ^2_{mt}-proportionate modifying of the variance of the difference in matched sample means (PM-1) if

$$\frac{\mathrm{Var}(\bar{Y}_{mt} - \bar{Y}_{mc})}{\mathrm{Var}(\bar{Y}_{rt} - \bar{Y}_{rc})} - \rho^2_{mt}\frac{\mathrm{Var}(\bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc})}{\mathrm{Var}(\bar{\Delta}^*_{rt} - \bar{\Delta}^*_{rc})} - (1-\rho^2_{mt})\frac{\mathrm{Var}(\bar{W}^*_{mt} - \bar{W}^*_{mc})}{\mathrm{Var}(\bar{W}^*_{rt} - \bar{W}^*_{rc})} = 0. \quad (4)$$

(b) ρ^2_{mt}-proportionate modifying of the expectations of the sample variances (PM-2) if

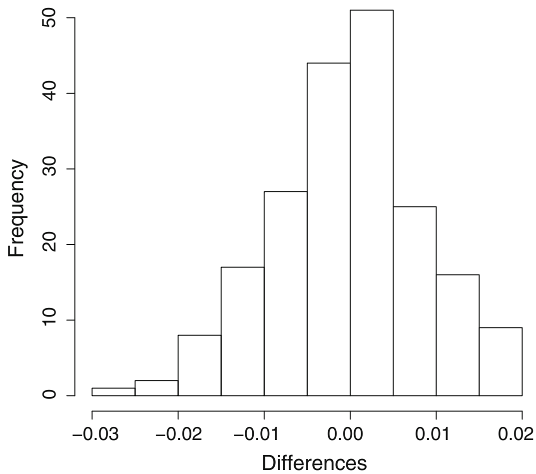
$$\frac{Ev_{mt}(Y)}{Ev_{rt}(Y)} - \rho^2_{mt}\frac{Ev_{mt}(\Delta^*)}{Ev_{rt}(\Delta^*)} - (1-\rho^2_{mt})\frac{Ev_{mt}(W^*)}{Ev_{rt}(W^*)} = 0, \quad (5)$$

and

$$\frac{Ev_{mc}(Y)}{Ev_{rc}(Y)} - \rho^2_{mt}\frac{Ev_{mc}(\Delta^*)}{Ev_{rc}(\Delta^*)} - (1-\rho^2_{mt})\frac{Ev_{mc}(W^*)}{Ev_{rc}(W^*)} = 0, \quad (6)$$

where v_{mt}, v_{mc}, v_{rt} and v_{rc} are, respectively, the sample variances of n_t and n_c matched and randomly chosen treated and control units using $(n_t - 1)$ and $(n_c - 1)$ in the denominators.

Fig. 2 The PM-1 property



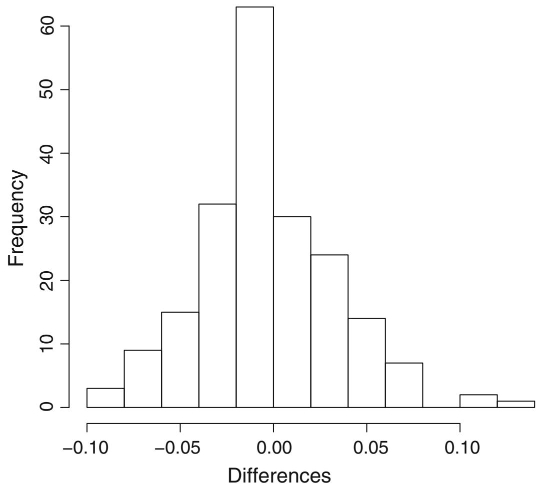


Fig. 3 The PM-2 treatment property

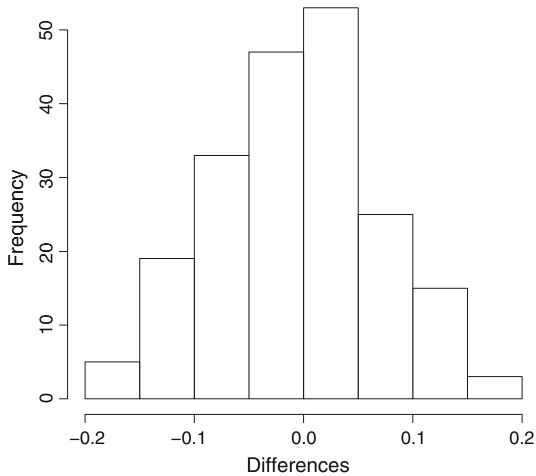


Fig. 4 The PM-2 control property

Values $\boldsymbol{\alpha} = (7.962, 9.381)$ are obtained independently from a normal distribution with mean 10 and standard deviation 4. We obtain 300 samples of covariates and the differences $\bar{Y}_{mt} - \bar{Y}_{mc},\ \bar{Y}_{rt} - \bar{Y}_{rc},\ \bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc},\ \bar{\Delta}^*_{rt} - \bar{\Delta}^*_{rc},\ \bar{W}^*_{mt} - \bar{W}^*_{mc},\ \bar{W}^*_{rt} - \bar{W}^*_{rc}$ are used to estimate the PM-1 difference (4). Estimates for the PM-2 differences (5) and (6) are similarly obtained. The procedure is repeated 200 times to obtain in Figs. 2, 3 and 4 histograms of the estimated values of PM-1 and PM-2 differences. The PM-1 mean difference is -0.00035 and the variance is 0.000078, the PM-2 mean treatment difference is -0.00319 and the variance is 0.00134, the PM-2 mean control difference is -0.00623 and the variance is 0.00578. The results support the PM properties (4)–(6) of the matching.

3 MP, CP and MCP matchings and the matching lemma

Let $\mathbf{X}_t$, $\mathbf{X}_c$ be p-dimensional covariates obtained randomly from the treatment and control populations $\mathcal{P}_t$ and $\mathcal{P}_c$, respectively, with mean and covariance matrices $\boldsymbol{\mu}_t$, Σ_t, $\boldsymbol{\mu}_c$, Σ_c, respectively; t denotes treatment, c denotes control. Let $\mathbf{X}_{mt}$, $\mathbf{X}_{mc}$ be p-dimensional covariate vectors from populations $\mathcal{MP}_t$ and $\mathcal{MP}_c$ used for matched sampling, with mean and covariance matrices, respectively, $\boldsymbol{\mu}_{mt}$, Σ_{mt}, $\boldsymbol{\mu}_{mc}$, Σ_{mc}. Often, $\mathcal{P}_t = \mathcal{MP}_t$, $\mathcal{P}_c$ and $\mathcal{MP}_c$ differ, and $\mathbf{X}_{mt}$ is obtained randomly from $\mathcal{P}_t$. In the examples in Sect. 2, $\mathcal{P}_t = \mathcal{MP}_t$.

Definition 2 (a) Assume without loss of generality $\boldsymbol{\mu}_c = \mathbf{0}$. Mean preserving (MP) matching on covariates $\mathbf{X}$ is any matching method for which

$$E(\mathbf{X}_{mt}) \propto E(\mathbf{X}_{mc}), \tag{7}$$

with subscripts mt and mc referring to the selected treatment unit and the matched control unit, respectively.

(b) Let $\mathbf{1}$ be the p-dimensional all-ones vector. Covariance preserving (CP) matching on $\mathbf{X}$ is any matching method for which

$$\mathrm{Var}(\mathbf{X}_{ms}) \propto \mathbf{I} + k_s\mathbf{1}\mathbf{1}', \quad k_s \in R,\ s = t, c, \tag{8}$$

$$\mathrm{Var}(\mathbf{X}_{mt} - \mathbf{X}_{mc}) \propto \mathbf{I} + k\mathbf{1}\mathbf{1}', \quad k \in R. \tag{9}$$

(c) Mean–covariance preserving (MCP) matching on covariates $\mathbf{X}$ is any matching method for which (7), (8) and (9) hold.

Remark 1 Relations (8) and (9) mean that in each covariance matrix all the variances are equal, all the covariances are equal, and this is preserved from covariates $\mathbf{X}_{mt}$ of the selected treatment unit to the covariates $\mathbf{X}_{mc}$ of the matched control unit and their difference $\mathbf{X}_{mt} - \mathbf{X}_{mc}$.

A matching Lemma follows which determines the mean and covariance structure of $\mathbf{X}_{mc}$ from that of $\mathbf{X}_{mt}$ and can be used for MP and CP matchings in practice. Use the notation

$$\Sigma_{mt} = (\sigma_{mt,ij}),\ \Sigma_{mc} = (\sigma_{mc,ij}), \mathrm{Var}(\mathbf{X}_{mt} - \mathbf{X}_{mc}) = (\sigma_{mt,mc,ij}), \quad 1 \le i, j \le p.$$

Lemma 1 *Select $\mathbf{X}_{mt}$ in $\mathcal{MP}_t$ and assume $\boldsymbol{\mu}_c = 0$. Conditionally on $X_{mt,i} = x_{mt,i}$, $X_{mc,i}$ is obtained at random from a distribution with mean $\delta x_{mt,i}$, and variance $\beta\sigma_{mt,ii}$, and support in $\mathcal{MP}_c$; $\delta \in R$, $\beta \ge 0$, $i = 1, \ldots, p$. Then,*

(a) *The matching on $\mathbf{X}_{mt}$ is MP-matching since*

$$\boldsymbol{\mu}_{mc} = \delta\boldsymbol{\mu}_{mt}, \tag{10}$$

(b) *It holds*

$$\sigma_{mc,ii} = (\delta^2 + \beta)\sigma_{mt,ii}, \qquad \sigma_{mt,mc,ii} = [\beta + (1-\delta)^2]\sigma_{mt,ii}, \quad i = 1, \ldots, p. \tag{11}$$

(c) *If the* $\mathbf{X}_{mt}$*-covariates are independent, then*

$$\sigma_{mc,ij} = 0 = \sigma_{mt,mc,ij}, \quad i \neq j. \tag{12}$$

(d) *If the* $\mathbf{X}_{mt}$*-covariates are uncorrelated and* $\sigma_{mc,ii} \approx 0,\ i = 1, \ldots, p$, *then*

$$\sigma_{mc,ij} \approx 0 = \sigma_{mt,mc,ij}, \quad i \neq j.$$

(e) *In addition to the assumptions for* (a), (b)*, and either* (c) *or* (d)*, assume that* $\sigma_{mt,ii} = \sigma^2_{mt},\ i = 1, \ldots, p$. *Then, the so-obtained matching is MCP-matching and* $k_t = k_c = k = 0$ *in* (8), (9).

Corollary 1 *Under the assumptions in Lemma 1, conditional nearest neighbor matching with one possible choice of matching value* $\mathbf{X}_{mc} = \delta \mathbf{x}_{mt}$ *is MCP-matching.*

Remark 2 When $\mu_{c,i} = 0$, the matching in the EPBR simulation example using the uniform distribution is MP matching since $EX_{mc,i} = (1+z)EX_{mt,i},\ i = 1, 2$. When the treatment and control populations follow PES or DPEMS distributions in canonical form, affinely invariant matching methods are MCP-matching (Theorem 3.1 in R&T and R&S). The matching in the variance PM simulation example is MCP matching.

4 MP-matching and the EPBR property

Let $\boldsymbol{\alpha} \in R^p,\ \boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$, and for $\mathbf{X}$ in either $\mathcal{P}_t$ or $\mathcal{P}_c$ let

$$Y = \boldsymbol{\alpha}'\mathbf{X}. \tag{13}$$

EPBR is a property that depends on expected values only; thus we decompose Y using a decomposition of $\boldsymbol{\alpha}$ along $\boldsymbol{\mu}_t$. Let $\boldsymbol{\gamma}_t \in R^p,\ \boldsymbol{\gamma}_t'\boldsymbol{\gamma}_t = 1$, be a vector orthogonal to $\boldsymbol{\mu}_t$ on the plane determined by $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}_t$,

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}'\boldsymbol{\mu}_t/||\boldsymbol{\mu}_t||)\boldsymbol{\mu}_t/||\boldsymbol{\mu}_t|| + (\boldsymbol{\alpha}'\boldsymbol{\gamma}_t)\boldsymbol{\gamma}_t,$$

to obtain the decomposition of Y,

$$Y = \rho_t \Delta + \sqrt{1-\rho_t^2}\,W \tag{14}$$

$$\rho_t = \boldsymbol{\alpha}'\boldsymbol{\mu}_t/||\boldsymbol{\mu}_t||, \quad \Delta = \boldsymbol{\mu}_t'\mathbf{X}/||\boldsymbol{\mu}_t||, \quad W = \boldsymbol{\gamma}_t'\mathbf{X}. \tag{15}$$

For $\mathbf{X}$ in either $\mathcal{MP}_t$ or $\mathcal{MP}_c$, (14) and (15) hold with $\rho_{mt}, \boldsymbol{\gamma}_{mt}$ or $\rho_{mc}, \boldsymbol{\gamma}_{mc}$ instead of $\rho_t, \boldsymbol{\gamma}_t$.

In this and subsequent sections, $\bar{U}_{rs}$ (resp. $\bar{U}_{ms}$) is the average of obtained U-values from $\mathcal{P}_s$ (resp. $\mathcal{MP}_s$), $s = t, c$.

Proposition 1 *Assume without loss of generality* $\boldsymbol{\mu}_c = \mathbf{0}$.

(a) *If the matching is MP and, in addition, it holds* $\boldsymbol{\mu}_{mt} = \zeta\boldsymbol{\mu}_t,\ \zeta \in R$, *then the matching is also EPBR,*

$$\frac{E(\bar{Y}_{mt} - \bar{Y}_{mc})}{E(\bar{Y}_{rt} - \bar{Y}_{rc})} = sign(\zeta)\frac{E(\bar{\Delta}_{mt} - \bar{\Delta}_{mc})}{E(\bar{\Delta}_{rt} - \bar{\Delta}_{rc})}, \tag{16}$$

where the subscripts rt and rc refer, respectively, to a randomly chosen sample of n_t treated and n_c control units. The percent reduction in bias is the same for any Y (i.e. any linear combination of X), since $\frac{E(\bar{\Delta}_{mt}-\bar{\Delta}_{mc})}{E(\bar{\Delta}_{rt}-\bar{\Delta}_{rc})}$ takes the same value for all Y. The result is independent of the covariance structure.

(b) *If the matching is MP and, in addition, it is also EPBR, then* $\boldsymbol{\mu}_{mt} = \zeta\boldsymbol{\mu}_t,\ \zeta \in R$.

Remark 3 The mild moment condition $\boldsymbol{\mu}_{mt} = \zeta\boldsymbol{\mu}_t,\ \zeta \in R$, holds when $\mathcal{P}_t = \mathcal{MP}_t$ with $\zeta = 1$. Under MP-matching, Proposition 1 provides a necessary and sufficient condition for the EPBR property to hold.

5 CP-matching and the PM properties

In R&T and R&S conditions ($\mathcal{C}1$), on $\mathcal{P}_t$ and $\mathcal{P}_c$, and ($\mathcal{C}2$), on $\mathcal{MP}_t$ and $\mathcal{MP}_c$, are sufficient for the EPBR and PM properties to hold for the **X**-covariates and affine invariant matching methods;

($\mathcal{C}1$) the control and the treatment populations are in canonical form,

$$\boldsymbol{\mu}_t = \eta\mathbf{1},\ \Sigma_t = \sigma^2\mathbf{I},\ \boldsymbol{\mu}_c = \mathbf{0},\ \Sigma_c = \mathbf{I}, \tag{17}$$

with **1** the all-ones vector (called also unit vector in R&T, p. 1081), **0** the zero vector, η positive scalar constant,

($\mathcal{C}2$) under ($\mathcal{C}1$), matching vectors $\mathbf{X}_{mt},\ \mathbf{X}_{mc}$ are obtained, respectively, from the treatment and the control groups with

(i) means proportional to **1**, and
(ii) the covariance matrix of the sample means difference $\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc}$ proportional to $\mathbf{I} + k\mathbf{1}\mathbf{1}'$, $k \in R$.

Under ($\mathcal{C}1$), **X** is projected along **1** that is proportional to $\boldsymbol{\mu}_t$ because of (17) and the standardized linear discriminant

$$\Delta^* = \mathbf{1}'\mathbf{X}/\sqrt{p}, \qquad \mathbf{1}' = (1, \ldots, 1), \tag{18}$$

is used.

There are cases where the **X**-coordinates have not the same mean and we would like to know when the PM-properties hold. A new set-up is presented herein to deal with this situation. $(\mathcal{C}1)$ is not necessary for the results to hold, a discriminant other than (18) is used and CP-property will hold for transformed covariates $\mathbf{X}^*$ and in particular cases for $\mathbf{X}$.

Assume that

$(\mathcal{A}1)$ the $\mathbf{X}$-coordinates are uncorrelated in both populations, and
$(\mathcal{A}2)$ for the variances $\sigma^2_{t,i}$, $\sigma^2_{c,i}$, of X_i in $\mathcal{P}_t$ and $\mathcal{P}_c$, respectively,

$$\frac{\sigma^2_{t,i}}{\sigma^2_{c,i}} = \sigma^{*2}, \quad i = 1, \ldots, p. \tag{19}$$

$(\mathcal{A}1)$ and $(\mathcal{A}2)$ hold when $(\mathcal{C}1)$ holds. $(\mathcal{A}2)$, like the condition of exchangeability in R&T and R&S, makes the covariates in the treatment and control populations “similar” and holds when $\Sigma_t \propto \Sigma_c$. $(\mathcal{A}2)$ may hold when a subpopulation of $\mathcal{P}_c$ is used.

Standardize the covariates in $\mathcal{P}_t, \mathcal{P}_c$ using the means and variances in $\mathcal{P}_c$,

$$X_i^* = \frac{X_i - \mu_{c_i}}{\sigma_{c_i}}, \qquad i = 1, \ldots, p, \tag{20}$$

and observe that for these transformed covariates

$$E_c(X_i^*) = 0, \quad \mathrm{Var}_c(X_i^*) = 1$$
$$E_t(X_i^*) = \frac{\mu_{t_i} - \mu_{c_i}}{\sigma_{c_i}} = \mu_i^*, \quad \mathrm{Var}_t(X_i^*) = \frac{\sigma^2_{t_i}}{\sigma^2_{c_i}} = \sigma^{*2}, \; i = 1, \ldots, p,$$

with E_t, Var_t, E_c, Var_c denoting, respectively, the means and variances in the treatment and control populations. The means $\boldsymbol{\mu}_t^*$, $\boldsymbol{\mu}_c^*$ and covariance matrices $\Sigma_t^* = (\sigma^*_{t,ij})$, $\Sigma_c^* = (\sigma^*_{c,ij})$ of $\mathbf{X}^{*\prime} = (X_1^*, \ldots, X_p^*)$, respectively, in the treatment and control populations are

$$\boldsymbol{\mu}_t^{*'} = (\mu_1^*, \ldots, \mu_p^*), \quad \boldsymbol{\mu}_c^* = \mathbf{0}, \tag{21}$$
$$\Sigma_t^* = \sigma^{*2}\mathbf{I}, \qquad \Sigma_c^* = \mathbf{I}. \tag{22}$$

Note that (21) holds by construction but (22) holds from $(\mathcal{A}1)$ and $(\mathcal{A}2)$. $(\mathcal{A}2)$ is the assumption $\sigma_{mt,ii} = \sigma^2_{mt}$, $i = 1, \ldots, p$, needed in Corollary 1 for MCP matching of the $\mathbf{X}^*$ covariates.

Remark 4 When $\boldsymbol{\mu}_c = \mathbf{0}$, $\Sigma_t = \sigma^2\mathbf{I}$ and $\Sigma_c = \mathbf{I}$, $(\mathcal{A}1)$ and $(\mathcal{A}2)$ hold and $\mathbf{X}^* = \mathbf{X}$.

The analogous of the standardized linear discriminant (18) is

$$\Delta^* = \boldsymbol{\mu}_t^{*'}\mathbf{X}^*/||\boldsymbol{\mu}_t^*||. \tag{23}$$

For $Y = \boldsymbol{\alpha}^{'}\mathbf{X}^*$, the analogous of (14) is

$$Y = \rho_t \Delta^* + \sqrt{1-\rho_t^2}\, W^*, \tag{24}$$

$$\rho_t = \boldsymbol{\alpha}'\boldsymbol{\mu}_t^*/||\boldsymbol{\mu}_t^*||, \quad W^* = \boldsymbol{\gamma}_t'\mathbf{X}^*. \tag{25}$$

A similar decomposition holds when $\mathbf{X}^*$ is obtained from either $\mathcal{MP}_{mt}$ or $\mathcal{MP}_{mc}$ with $\rho_{mt},\ \boldsymbol{\gamma}_{mt}$ instead of $\rho_t,\ \boldsymbol{\gamma}_t$.

Remark 5 Under ($\mathcal{A}1$) and ($\mathcal{A}2$), one can use on $\mathbf{X}^*$-covariates the matching described in Lemma 1 with $\sigma^*_{mt,ij} = 0$ to obtain MCP-matching with $k_t = k_c = k = 0$.

The following proposition decomposes the effect on Y of CP-matching on $\mathbf{X}^*$, into the effects of the matching on Δ^* and on W^*.

Proposition 2 *Under* ($\mathcal{A}1$) *and* ($\mathcal{A}2$), *for the transformed covariates* (20)

(a) *CP-matching with $k = 0$ is ρ^2_{mt}-proportionate modifying of the variance of the difference in matched sample means* (*PM-1*),

$$\frac{\mathrm{Var}(\bar{Y}_{mt} - \bar{Y}_{mc})}{\mathrm{Var}(\bar{Y}_{rt} - \bar{Y}_{rc})} = \rho^2_{mt}\frac{\mathrm{Var}(\bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc})}{\mathrm{Var}(\bar{\Delta}^*_{rt} - \bar{\Delta}^*_{rc})} + (1-\rho^2_{mt})\frac{\mathrm{Var}(\bar{W}^*_{mt} - \bar{W}^*_{mc})}{\mathrm{Var}(\bar{W}^*_{rt} - \bar{W}^*_{rc})} \tag{26}$$

where the ratios

$$\frac{\mathrm{Var}(\bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc})}{\mathrm{Var}(\bar{\Delta}^*_{rt} - \bar{\Delta}^*_{rc})}, \qquad \frac{\mathrm{Var}(\bar{W}^*_{mt} - \bar{W}^*_{mc})}{\mathrm{Var}(\bar{W}^*_{rt} - \bar{W}^*_{rc})}$$

take the same two values for all Y.

(b) *CP-matching with $k_t = k_c = 0$ is ρ^2_{mt}-proportionate modifying of the expectations of the sample variances* (*PM-2*),

$$\frac{Ev_{mt}(Y)}{Ev_{rt}(Y)} = \rho^2_{mt}\frac{Ev_{mt}(\Delta^*)}{Ev_{rt}(\Delta^*)} + (1-\rho^2_{mt})\frac{Ev_{mt}(W^*)}{Ev_{rt}(W^*)} \tag{27}$$

and

$$\frac{Ev_{mc}(Y)}{Ev_{rc}(Y)} = \rho^2_{mt}\frac{Ev_{mc}(\Delta^*)}{Ev_{rc}(\Delta^*)} + (1-\rho^2_{mt})\frac{Ev_{mc}(W^*)}{Ev_{rc}(W^*)} \tag{28}$$

where $v_{mt},\ v_{mc},\ v_{rt}$ and v_{rc} are, respectively, the sample variances of n_t and n_c matched and randomly chosen treated and control units using $(n_t - 1)$ and $(n_c - 1)$ in the denominators. The ratios

$$\frac{Ev_{mt}(\Delta^*)}{Ev_{rt}(\Delta^*)}, \quad \frac{Ev_{mt}(W^*)}{Ev_{rt}(W^*)}, \quad \frac{Ev_{mc}(\Delta^*)}{Ev_{rc}(\Delta^*)}, \quad \frac{Ev_{mc}(W^*)}{Ev_{rc}(W^*)}$$

take the same four values for all Y.

Corollary 2 *When the* $\mathbf{X}$*-covariates in the control population have the same variance, i.e.* $\sigma_{c_i}^2 = \sigma_c^2,\ i = 1, \ldots, p$, *Propositions* 1 *and* 2 *hold for linear functions of* $\mathbf{X}$.

The next result implies results in R&T and R&S.

Corollary 3 *When the distributions of* $\mathbf{X}$ *are in canonical form* (17)*, MCP matching on* $\mathbf{X}$ *with* $\boldsymbol{\mu}_{mt} = \zeta \boldsymbol{\mu}_t,\ \zeta \in R$, *is for linear functions of* $\mathbf{X}$

(i) *EPBR, and*
(ii) *proportionate modifying of the variance of the difference in matched sample means and of the expectations of the sample variances.*

6 Discussion

When treatment samples are obtained from one population, i.e. $\mathcal{P}_t = \mathcal{MP}_t$, the EPBR and PM properties hold, respectively, with MP and CP matchings on sub-populations of the control population according to Lemma 1. Most conditions in the Lemma are moment conditions that hold when $\mathcal{M}P_c$ is continuous, has a smooth density and we can sample from uniform distributions in neighborhoods of the observed treatment values. The EPBR property holds since we can easily have $EX_{mc,i} = \delta EX_{mt,i},\ i = 1, \ldots, p$. For the CP property to hold, the $\mathbf{X}^*$ covariates have to be either independent or uncorrelated with (a) $\mathrm{Var}(X^*_{mc,i}|X^*_{mt,i} = x^*_{mt,i}) = \beta\sigma^{*2}_{mt}$, b) $\sigma^*_{mc,ii} \approx 0$ and (c) $\sigma^*_{mt,ii} = \sigma^{*2}_{mt}$ (i.e. ($\mathcal{A}2$) holds), $i = 1, \ldots, p$. Condition (a) is satisfied when sampling from a uniform distribution on an interval with length proportional to σ_{mt}. Condition (b) holds when, as in the variance PM simulations example, the matching value is average of p-values with p large. The most stringent condition (c) may hold when using only a subset of $\mathcal{P}_c$ chosen such that the variances ratios (19) all coincide.

For a large, discreet control population $\mathcal{M}P_c$, assume that when $X_{mt,i} = x_{mt,i}$ its nearest neighbors $x_{c,i,L}, x_{c,i,U}$ in the control group satisfy $x_{c,i,L} \le x_{mt,i} < x_{c,i,U},\ 1 \le i \le p$. For the EPBR property to hold use probabilities $q_i = q_i(x_{mt,i}) = \frac{x_{c,i,U} - x_{mt,i}}{x_{c,i,U} - x_{c,i,L}}$ such that $q_i x_{c,i,L} + (1 - q_i)x_{c,i,U} = x_{mt,i}$ and draw an observation $\tilde{X}_{mc,i}$ from $\{x_{c,i,L}, x_{c,i,U}\}$ with $P[\tilde{X}_{mc,i} = x_{c,i,L}] = q_i$ such that $E\tilde{X}_{c,i} = EX_{mt,i},\ i = 1, \ldots, p$. For CP matching, use neighbors $\{x^*_{c,i,L}, x^*_{c,i,U}\}$ such that the conditional variance $(x^*_{c,i,U} - x^*_{mt,i})(x^*_{mt,i} - x^*_{c,i,L})$ is the same for $i = 1, \ldots, p$, in order to satisfy a). This is feasible if $\mathcal{M}P_c$ is large. More than two neighbors can be used in this matching.

7 Appendix

Proof of the matching Lemma 1 (a) For $i = 1, \ldots, p$, it holds

$$E(X_{mc,i}) = E[E(X_{mc,i}|X_{mt,i})] = E(\delta X_{mt,i}) = \delta\mu_{mt,i}.$$

(b) For $i = 1, \ldots, p$, it holds

$$\begin{aligned}
&\mathrm{Var}(X_{mc,i}) = \mathrm{Var}[E(X_{mc,i}|X_{mt,i})] + E[\mathrm{Var}(X_{mc,i}|X_{mt,i})] = (\delta^2 + \beta)\sigma_{mt,ii},\\
&E(X_{mt,i}X_{mc,i}) = E[E(X_{mt,i}X_{mc,i}|X_{mt,i})] = \delta E(X^2_{mt,i}),\\
&\mathrm{Cov}(X_{mt,i}, X_{mc,i}) = E(X_{mt,i}X_{mc,i}) - \delta\mu^2_{mt,i} = \delta\sigma_{mt,ii},
\end{aligned}$$

to obtain

$$\begin{aligned}
\mathrm{Var}(X_{mt,i} - X_{mc,i}) &= \mathrm{Var}(X_{mt,i}) + \mathrm{Var}(X_{mc,i}) - 2\mathrm{Cov}(X_{mt,i}, X_{mc,i})\\
&= [\beta + (1-\delta)^2]\sigma_{mt,ii}.
\end{aligned}$$

(c) For $1 \leq i,\ j \leq p,\ i \neq j,\ X_{mc,i},\ X_{mc,j}$ are conditionally independent given $X_{mt,i},\ X_{mt,j}$, thus

$$\begin{aligned}
&E(X_{mc,i}X_{mc,j}) = E[E(X_{mc,i}X_{mc,j}|X_{mt,i},\ X_{mt,j})] = \delta^2 E(X_{mt,i}X_{mt,j}),\\
&\mathrm{Cov}(X_{mc,i}, X_{mc,j}) = E(X_{mc,i}X_{mc,j}) - \delta^2\mu_{mt,i}\mu_{mt,j} = \delta^2\sigma_{mt,ij},\\
&E(X_{mt,i}X_{mc,j}) = E[E(X_{mt,i}X_{mc,j}|X_{mt,i}, X_{mt,j})] = \delta E(X_{mt,i}X_{mt,j}),\\
&\mathrm{Cov}(X_{mt,i}, X_{mc,j}) = E(X_{mt,i}X_{mc,j}) - \delta\mu_{mt,i}\mu_{mt,j} = \delta\sigma_{mt,ij},
\end{aligned}$$

to obtain

$$\begin{aligned}
&\mathrm{Cov}(X_{mt,i} - X_{mc,i}, X_{mt,j} - X_{mc,j})\\
&\quad = \mathrm{Cov}(X_{mt,i}, X_{mt,j}) - \mathrm{Cov}(X_{mt,i}, X_{mc,j}) - \mathrm{Cov}(X_{mc,i}, X_{mt,j})\\
&\quad\quad + \mathrm{Cov}(X_{mc,i}, X_{mc,j})\\
&\quad = \sigma_{mt,ij} - 2\delta\sigma_{mt,ij} + \delta^2\sigma_{mt,ij} = (1-\delta)^2\sigma_{mt,ij}.
\end{aligned}$$

(d) The result $\sigma_{mc,ij} \approx 0$ follows from the inequality $|\mathrm{Cov}(U, V)| \leq \sqrt{\mathrm{Var}(U)}\sqrt{\mathrm{Var}(V)}$. The result $\sigma_{mt,mc,ij} = 0$ that improved my result $\sigma_{mt,mc,ij} \approx 0,\ i \neq j$, follows because, as the referee cleverly observed, $X_{mt,1}, \ldots, X_{mt,p}$ are uncorrelated and the calculations of $E(X_{mt,i}X_{mc,j})$ and of $\mathrm{Cov}(X_{mt,i}, X_{mc,j})$, $i \neq j$, in c) do not use the assumed conditional independence.

(e) Follows from the definition of MCP matching. □

Proof of Corollary 1 Since $E\mathbf{X}_{mc} = \delta E\mathbf{X}_{mt}$ and

$$\mathrm{Var}(X_{mc,i}) = \mathrm{Var}[E(X_{mc,i}|X_{mt,i})] = \delta^2\mathrm{Var}(X_{mt,i}) = \delta^2\sigma^2_{mt},\ i = 1, \ldots, p,$$

the result follows. □

Proof of Proposition 1 (a) For Δ and W in (14), it holds from (15)

$$\begin{aligned}
E_t(\Delta) = \boldsymbol{\mu}_t'\boldsymbol{\mu}_t/||\boldsymbol{\mu}_t|| = ||\boldsymbol{\mu}_t||, \qquad E_t(W) = \boldsymbol{\gamma}'\boldsymbol{\mu}_t = 0,\\
E_c(\Delta) = \boldsymbol{\mu}_t'\boldsymbol{\mu}_c/||\boldsymbol{\mu}_t|| = 0, \qquad E_c(W) = \boldsymbol{\gamma}'\boldsymbol{\mu}_c = 0,
\end{aligned}$$

and for random samples from $\mathcal{P}_t$ and $\mathcal{P}_c$

$$E(\bar{\Delta}_{rt}) = ||\boldsymbol{\mu}_t||, \qquad E(\bar{W}_{rt}) = 0, \qquad E(\bar{Y}_{rt}) = \rho_t ||\boldsymbol{\mu}_t||, \tag{29}$$

$$E(\bar{\Delta}_{rc}) = E(\bar{W}_{rc}) = E(\bar{Y}_{rc}) = 0. \tag{30}$$

If the matching is MP and, in addition, $\boldsymbol{\mu}_{mt} = \zeta \boldsymbol{\mu}_t,\ \zeta \in R$, Eqs. (14) and (15) with $\rho_{mt} = sign(\zeta)\rho_t$, $\boldsymbol{\gamma}_{mt} = \boldsymbol{\gamma}_t$, imply

$$E(\bar{Y}_{mt} - \bar{Y}_{mc}) = sign(\zeta)\rho_t E(\bar{\Delta}_{mt} - \bar{\Delta}_{mc}) + \sqrt{1-\rho_t^2}\, E(\bar{W}_{mt} - \bar{W}_{mc}).$$

By (15),

$$E(\bar{W}_{mt} - \bar{W}_{mc}) = \boldsymbol{\gamma}'_t E(\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc}), \qquad \boldsymbol{\gamma}'_t \boldsymbol{\mu}_{mt} = \boldsymbol{\gamma}'_t \zeta \boldsymbol{\mu}_t = 0,$$

and by the MP-matching, $E(\bar{\mathbf{X}}_{mt} - \bar{\mathbf{X}}_{mc}) \propto \boldsymbol{\mu}_{mt}$, thus

$$E(\bar{Y}_{mt} - \bar{Y}_{mc}) = sign(\zeta)\rho_t E(\bar{\Delta}_{mt} - \bar{\Delta}_{mc}).$$

Relation (16) follows from (14), (29) and (30) by noting that

$$E(\bar{Y}_{rt} - \bar{Y}_{rc}) = \rho_t E(\bar{\Delta}_{rt} - \bar{\Delta}_{rc}).$$

(b) From the EPBR property, there is $b \in R$ such that for any $\boldsymbol{\alpha} \in R^p$,

$$\frac{E(\bar{Y}_{mt} - \bar{Y}_{mc})}{E(\bar{Y}_{rt} - \bar{Y}_{rc})} = b = \frac{\rho_{mt} E(\bar{\Delta}_{mt} - \bar{\Delta}_{mc})}{\rho_t E(\bar{\Delta}_{rt} - \bar{\Delta}_{rc})},$$

with the last equality due to the MP-matching assumption, or

$$\rho_{mt} = b^* \rho_t, \qquad b^* \in R, \qquad \text{i.e.} \qquad \boldsymbol{\alpha}'\boldsymbol{\mu}_{mt}/||\boldsymbol{\mu}_{mt}|| = b^* \boldsymbol{\alpha}'\boldsymbol{\mu}_t/||\boldsymbol{\mu}_t||,$$

and this holds for any vector $\boldsymbol{\alpha}$ in the usual orthonormal basis in R^p implying that

$$\mu_{mt,i} = b^* \frac{||\boldsymbol{\mu}_{mt}||}{||\boldsymbol{\mu}_t||} \mu_{t,i}, \qquad i = 1, \ldots, p.$$

□

When ($\mathcal{A}1$) and ($\mathcal{A}2$) hold, for the transformed covariates $\mathbf{X}^*$ under treatment and control it holds

$$\text{Var}_t(\Delta^*) = \boldsymbol{\mu}_t^{*'} \Sigma_t^* \boldsymbol{\mu}_t^* / ||\boldsymbol{\mu}_t^*||^2 = \sigma^{*2}, \qquad \text{Var}_c(\Delta^*) = \boldsymbol{\mu}_t^{*'} \Sigma_c^* \boldsymbol{\mu}_t^* / ||\boldsymbol{\mu}_t^*||^2 = 1. \tag{31}$$

For the variances of W^* in the treatment and control populations it holds

$$\text{Var}_t(W^*) = \boldsymbol{\gamma}'_t \Sigma_t^* \boldsymbol{\gamma}_t = \sigma^{*2}, \qquad \text{Var}_c(W^*) = \boldsymbol{\gamma}'_t \Sigma_c^* \boldsymbol{\gamma}_t = 1, \tag{32}$$

and the covariance of Δ^* and W^*

$$\mathrm{Cov}_s(\Delta^*, W^*) = \mathrm{Cov}_s(\boldsymbol{\mu}_t^{*'}\mathbf{X}^*/||\boldsymbol{\mu}_t^*||, \boldsymbol{\gamma}_t'\mathbf{X}^*) = \frac{\boldsymbol{\mu}_t^{*'}}{||\boldsymbol{\mu}_t^*||}\Sigma_s^*\boldsymbol{\gamma}_t = 0, \qquad s = t, c. \tag{33}$$

Relations (32), (33) hold for any W^* for which

$$W^* = \boldsymbol{\gamma}_t'\mathbf{X}^*, \qquad \boldsymbol{\gamma}_t'\boldsymbol{\mu}_t^* = 0, \qquad \boldsymbol{\gamma}_t'\boldsymbol{\gamma}_t = 1. \tag{34}$$

From (24), (31) and (32), for random samples of size n_t in the treated population and of size n_c in the control population it holds

$$\mathrm{Var}(\bar{\Delta}_{rt}^*) = \mathrm{Var}(\bar{W}_{rt}^*) = \mathrm{Var}(\bar{Y}_{rt}) = \frac{\sigma^{*2}}{n_t}, \tag{35}$$

$$\mathrm{Var}(\bar{\Delta}_{rc}^*) = \mathrm{Var}(\bar{W}_{rc}^*) = \mathrm{Var}(\bar{Y}_{rc}) = \frac{1}{n_c}. \tag{36}$$

Remark 6 Decomposition (24) and (25) are used in the sequel for the selected and the matched samples, with $\boldsymbol{\mu}_t^*$ replaced by $\boldsymbol{\mu}_{mt}^*$, the mean of the selected treatment sample, and with $\boldsymbol{\gamma}_{mt}$, ρ_{mt} instead of $\boldsymbol{\gamma}_t$, ρ_t.

Proposition 3 *Matched treated and control samples on* $\mathbf{X}^*$ *are obtained, respectively, with sizes* n_t, n_c *and means* $\bar{\mathbf{X}}_{mt}^*$, $\bar{\mathbf{X}}_{mc}^*$.

(a) *If the matching is MP-matching,*

$$E(\bar{\mathbf{X}}_{mt}^*) \propto E(\bar{\mathbf{X}}_{mc}^*). \tag{37}$$

(b) *If the matching is CP-matching with* $k_t = k_c = k = 0$,

$$\mathrm{Var}(\bar{\mathbf{X}}_{mt}^*) \propto \mathrm{Var}(\bar{\mathbf{X}}_{mc}^*) \propto \mathrm{Var}(\bar{\mathbf{X}}_{mt}^* - \bar{\mathbf{X}}_{mc}^*) \propto \mathbf{I}, \tag{38}$$

$$E(v_{mt}(\mathbf{X}^*)) \propto E(v_{mc}(\mathbf{X}^*)) \propto \mathbf{I}, \tag{39}$$

where $v_{mt}(\mathbf{X}^*)$ *and* $v_{mc}(\mathbf{X}^*)$ *are the unbiased sample covariance matrices of* $\mathbf{X}^*$ *in the matched treated and control samples.*

Proof It follows from the definitions of MP-matching and CP-matching. □

The following results are useful:

Corollary 4 *Under the hypotheses of Proposition* 3 b)*, the quantities Var*$(\bar{W}_{mt}^* - \bar{W}_{mc}^*)$, $E(v_{mt}(W^*))$ *and* $E(v_{mc}(W^*))$ *take the same values for all* W^* *satisfying* (34)*. When in addition* (22) *holds, the analogous three results apply for statistics in random samples indexed by rt and rc. Since* Δ^* *is the discriminant, defined without regard to the choice of* Y, *the analogous quantities for* Δ^* *are also the same for all* Y.

Proof Since

$$W^* = \boldsymbol{\gamma}'_{mt}\mathbf{X}^*,\ \ \bar{W}^*_{mt} - \bar{W}^*_{mc} = \boldsymbol{\gamma}'_{mt}(\bar{\mathbf{X}}^*_{mt} - \bar{\mathbf{X}}^*_{mc}),$$

it holds that

$$\mathrm{Var}(\bar{W}^*_{mt} - \bar{W}^*_{mc}) = \boldsymbol{\gamma}'_{mt}\mathrm{Var}(\bar{\mathbf{X}}^*_{mt} - \bar{\mathbf{X}}^*_{mc})\boldsymbol{\gamma}_{mt}$$

which from (34) and (38) takes the same value for each $\boldsymbol{\gamma}_{mt}$. Since

$$v_{ms}(W^*) = v_{ms}(\boldsymbol{\gamma}'_{mt}X^*), \qquad s = t, c,$$

from (39) the results for $E(v_{mt}(W^*))$ and $E(v_{mc}(W^*))$ hold. The analogous results for random samples follow from (32), (35) and (36). The remaining results follow easily. □

Proof of Proposition 2 (a) For the matched samples, (24) implies that

$$\mathrm{Var}(\bar{Y}_{mt} - \bar{Y}_{mc}) = \rho^2_{mt}\mathrm{Var}(\bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc}) + (1 - \rho^2_{mt})\mathrm{Var}(\bar{W}^*_{mt} - \bar{W}^*_{mc}),$$

since from (23) and (34)

$$\bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc} = \frac{\boldsymbol{\mu}^{*'}_{mt}}{||\boldsymbol{\mu}^*_{mt}||}(\bar{\mathbf{X}}^*_{mt} - \bar{\mathbf{X}}^*_{mc}), \qquad \bar{W}^*_{mt} - \bar{W}^*_{mc} = \boldsymbol{\gamma}'_{mt}(\bar{\mathbf{X}}^*_{mt} - \bar{\mathbf{X}}^*_{mc}),$$

and from (38)

$$\mathrm{Cov}(\bar{\Delta}^*_{mt} - \bar{\Delta}^*_{mc}, \bar{W}^*_{mt} - \bar{W}^*_{mc}) \propto \boldsymbol{\mu}^{*'}_{mt}\mathrm{Var}(\bar{\mathbf{X}}^*_{mt} - \bar{\mathbf{X}}^*_{mc})\boldsymbol{\gamma}_{mt} = 0. \tag{40}$$

Decomposition (26) of the variances ratio follows from (35) and (36) since for random samples from the treated and control populations

$$\mathrm{Var}(\bar{Y}_{rt} - \bar{Y}_{rc}) = \mathrm{Var}(\bar{\Delta}^*_{rt} - \bar{\Delta}^*_{rc}) = \mathrm{Var}(\bar{W}^*_{rt} - \bar{W}^*_{rc}) = \frac{\sigma^{*2}}{n_t} + \frac{1}{n_c}.$$

The results for the ratios (27) follow from Corollary 4.

(b) For the matched treated sample

$$Ev_{mt}(Y) = \rho^2_{mt}Ev_{mt}(\Delta^*) + (1 - \rho^2_{mt})Ev_{mt}(W^*)$$

because from (39) the expected matched treated sample covariance of Δ^* and W^* is proportional to

$$\frac{1}{||\boldsymbol{\mu}^*_{mt}||}E(\boldsymbol{\mu}^{*'}_{mt}v_{mt}(\mathbf{X}^*)\boldsymbol{\gamma}_{mt}) \propto \boldsymbol{\mu}^{*'}_{mt}\mathbf{I}\boldsymbol{\gamma}_{mt} = 0. \tag{41}$$

Noting that

$$Ev_{rt}(Y) = Ev_{rt}(\Delta^*) = Ev_{rt}(W^*)$$

(27) follows. Analogous derivations for the matched control sample establishes (28). The proof is completed using Corollary 4. □

Proof of Corollary 2 Observe that $\boldsymbol{\alpha}'(\mathbf{X}^*_{mt} - \mathbf{X}^*_{mc}) = \frac{1}{\sigma_c}\boldsymbol{\alpha}'(\mathbf{X}_{mt} - \mathbf{X}_{mc})$ and that σ_c cancels because all the results concern ratios. □

Proof of Corollary 3 From Remark 4, ($\mathcal{A}1$) and ($\mathcal{A}2$) are satisfied and $\mathbf{X}^* = \mathbf{X}$.

(i) follows from Proposition 1 (a).
(ii) follows from the proof of Proposition 2 since $\boldsymbol{\mu}_t = \eta\mathbf{1}$ and the equivalent to relations (40) and (41) are, respectively,

$$\mathbf{1}'(\mathbf{I} + k\mathbf{1}\mathbf{1}')\boldsymbol{\gamma} = 0,$$

and

$$\mathbf{1}'(\mathbf{I} + k_s\mathbf{1}\mathbf{1}')\boldsymbol{\gamma} = 0, \;\; s = t, c.$$

□

Acknowledgments Many thanks are due to all the members of the Statistics Department, Harvard University, for their warm hospitality during my sabbatical visit from N.U.S. in 2005–2006. In particular, I express my whole-hearted thanks to Professor Donald B. Rubin, my host, for introducing me to matching methods in his lectures, for many inspiring conversations related also to this work, and for the unforgettable and hospitable experience. Many special thanks are also due to an anonymous referee for the very careful reading of this work and the useful suggestions that helped to improve its readability and content, and in particular for providing the equality in matching Lemma 1 d). Part of this work was done while the author was affiliated with the Department of Statistics and Applied Probability, National University of Singapore. This research was partially supported by the National University of Singapore and a Cyprus U. of Technology start-up grant.

References

Althauser, R. P., Rubin, D. B. (1970). The computerized construction of a matched sample. *American Journal of Sociology, 76*, 325–346.

Bickel, S., Brückner, M., Scheffer, T. (2007). Discriminative learning for differing training and test distributions. In *Proceedings of the 24th international conference on machine learning.*

Cochran, W. G., Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyã A, 35*, 417–446.

Imai, K., Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association, 99*, 854–866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*, 706–710.

Joffe, M. M., Rosenbaum, P. R. (1999). Propensity scores. *American Journal of Epidemiology, 150*, 327–333.

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

Rosenbaum, P. R., Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 516–524.
Rosenbaum, P. R., Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*, 33–38.
Rubin, D. B. (1970). The use of matched sampling and regression adjustment in observational studies. Ph.D. Thesis, Statistics Department, Harvard University.
Rubin, D. B. (1973a). Matching to remove bias in observational studies. *Biometrics, 29*, 159–183. Correction (1974) *30*, 728.
Rubin, D. B. (1973b). The use of matching and regression adjustments to remove bias in observational studies. *Biometrics, 29*, 185–203.
Rubin, D. B. (1976a). Multivariate matching methods that are equal percent bias reducing, I: Some examples. *Biometrics, 32*, 109–120.
Rubin, D. B. (1976b). Multivariate matching methods that are equal percent bias reducing, II: Maximum on bias reduction. *Biometrics, 32*, 121–132.
Rubin, D. B., Stuart, E. A. (2006). Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions. *Annals of Statistics, 34*, 1814–1826.
Rubin, D. B., Thomas, N. (1992). Affinely invariant matching methods with ellipsoidal distributions. *Annals of Statistics, 20*, 1079–1093.
Rubin, D. B., Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics, 52*, 249–264.
Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference, 90*, 227–244.
Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1–21.
Sugiyama, M., Suzuki, T., Nakajima, S., Kashima, H., Von Bünau, P., Kawanabe, M. (2008). Direct importance estimation for covariate shift adaptation. *Annals of the Institute of Statistical Mathematics, 60*, 699–746.
Yatracos, Y. G. (2011). Causal inference for multiple treatments via sufficiency and ratios of generalized propensity scores. *Submitted for publication.* http://ktisis.cut.ac.cy/handle/10488/5127.