Covariate Balance Weighting Methods in Estimating Treatment Effects: An Empirical Comparison*†‡

Mingfeng Zhan[a], Zongwu Cai[b], Ying Fang[a,c], and Ming Lin[a,c]

[a]Wang Yanan Institute for Studies in Economics and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China.

[b]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

[c]Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.

December 28, 2020

HIGHLIGHTS

- The pros and cons of four methodologies for covariate balance weighting are explored.
- Simulation shows that four covariate balance weighting method dominates MLE.
- There is no a clear domination among four covariate balance weighting methods.

Abstract: We conduct a series of simulations to compare the finite sample performance of the average treatment effect estimators based on four recently proposed methodologies — the covariate balancing propensity score method, the stable balance weighting approach, the calibration balance weighting procedure, and the integrated propensity score method. Simulation results show that the performance of the four covariate balance weighting methods are generally better than that for the conventional method, maximum likelihood estimation method without covariate balance, and among the four covariate balance weighting methods, it is difficult to tell which covariate balance weighting method can dominate the others.

Keywords: Covariate balance; Propensity score; Treatment effects.

JEL Classification: C3, C5.

*We thank Professor José R. Zubizarreta for his providing of the *sbw* package and Professor Pedro H.C. Sant'Anna and Dr. Xu Qi for their kind help in using the *IPS* package. Also, the authors acknowledge the financial supports, in part, from the National Science Fund of China (NSFC) for Distinguished Scholars (71625001), the NSFC key projects with grant numbers 71631004 and 72033008, and Science Foundation of Ministry of Education of China (19YJA910003).

†*Corresponding author:* Y. Fang (E-mail: yifst1@xmu.edu.cn)

‡E-mail addresses: zhanmf@stu.xmu.edu.cn (M. Zhan), caiz@ku.edu (Z. Cai), yifst1@xmu.edu.cn (Y. Fang), linming50@xmu.edu.cn (M. Lin).

1 Introduction

During the recent years, different covariate balance weighting methods for estimating treatment effect have been proposed by researchers from different perspectives. A natural question one may ask is that what is the difference among these covariate balance weighting methods and is there any particular method that can outperform the others? In this paper, we explore the pros and cons of four methodologies proposed recently: the covariate balancing propensity score method (CBPS) by Imai and Ratkovic (2014), the stable balance weighting (SBW) approach by Zubizarreta (2015), the calibration balance weighting (CBW) procedure by Chan et al. (2016), and the integrated propensity score (IPS) method by Sant'Anna et al. (2020). Among them, the CBPS and IPS methods achieve covariate balance via modeling the propensity score while the SBW and CBW methods circumvent the propensity score model and directly obtain the covariate balance weights through optimization designs. To answer the second part of the previous question, similar to the paper by Wan, Xie and Hsiao (2018), we conduct a series of simulations to study the performance of these four covariate balance weighting methods as well as the maximum likelihood estimation (MLE) weighting method without covariate balance in estimating the average treatment effect (ATE). Through simulation studies, we find that generally speaking, the covariate balance weighting methods can outperform the MLE weighting method and the performance of the covariate balance weighting methods in estimating ATE seems to be comparable to some extent.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the four covariate balance weighting methods to explore their pros and cons. We conduct a series of simulations and summarize the main findings in Section 3. Section 4 concludes the paper.

2 Covariate Balance Weighting Methods

2.1 Background and MLE

Let T be a binary treatment variable, X be p-dimensional vector of observed pre-treatment covariates, and Y be the observed outcome variable given by $Y = TY(1) + (1 - T)Y(0)$, where $Y(1)$ and $Y(0)$ are potential outcomes under treated and untreated status, respectively.

Suppose we have a random sample $\{T_i, Y_i, X_i\}_{i=1}^n$ from (T, Y, X). The propensity score is defined as $\pi(X) = P(T = 1|X)$ and the ATE is defined as $\Delta = E[Y(1)-Y(0)]$. It is common to assume that the strong ignorability assumption:

$$(Y(0), Y(1)) \perp\!\!\!\perp T \mid X \quad \text{and} \quad \epsilon < \pi(X) < 1 - \epsilon,$$

for some $\epsilon > 0$ holds, where $\perp\!\!\!\perp$ denotes the statistical independence, under which and by the law of iterated expectations, the ATE can be expressed as

$$\Delta = E\left[\frac{TY}{\pi(X)} - \frac{(1-T)Y}{1-\pi(X)}\right]. \tag{1}$$

If the propensity score is known, then using the sample mean of (1), one can obtain a consistent estimator of the ATE

$$\hat{\Delta} = \frac{1}{n}\sum_{i=1}^n \left(\frac{T_i Y_i}{\pi(X_i)} - \frac{(1-T_i)Y_i}{1-\pi(X_i)}\right). \tag{2}$$

In practice, it is preferred to using the stabilized version of (2) as

$$\hat{\Delta}_s = \frac{\frac{1}{n}\sum_{i=1}^n \left(\frac{T_i}{\hat{\pi}(X_i)} Y_i\right)}{\frac{1}{n}\sum_{i=1}^n \frac{T_i}{\hat{\pi}(X_i)}} - \frac{\frac{1}{n}\sum_{i=1}^n \left(\frac{(1-T_i)}{1-\hat{\pi}(X_i)} Y_i\right)}{\frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\hat{\pi}(X_i)}}. \tag{3}$$

Generally speaking, the propensity score is unknown in observational study and needs to be estimated. A popular method is to parameterize the propensity score by a particular parametric model such as logistic regression model

$$\pi_\beta(X) = \frac{\exp(\beta^\top X)}{1+\exp(\beta^\top X)}, \tag{4}$$

where $\beta \in \Theta \subseteq \mathbb{R}^p$ is a p-dimensional vector of parameters. Then, by the MLE method, one can obtain an estimator of the unknown parameter β, denoted as $\hat{\beta}_{mle}$. Plugging $\hat{\beta}_{mle}$ into (4), one can obtain the estimated propensity score function $\hat{\pi}_{mle}$. By (3), the ATE estimator based on the MLE method is given by

$$\hat{\Delta}_{mle} = \frac{\frac{1}{n}\sum_{i=1}^n \left(\frac{T_i}{\hat{\pi}_{mle}(X_i)} Y_i\right)}{\frac{1}{n}\sum_{i=1}^n \frac{T_i}{\hat{\pi}_{mle}(X_i)}} - \frac{\frac{1}{n}\sum_{i=1}^n \left(\frac{(1-T_i)}{1-\hat{\pi}_{mle}(X_i)} Y_i\right)}{\frac{1}{n}\sum_{i=1}^n \frac{1-T_i}{1-\hat{\pi}_{mle}(X_i)}}.$$

Clearly, one can see that the MLE method quite sensitive to the misspecification of the propensity score as addressed in Kang and Schafer (2007). To address this issue, Imai and Ratkovic (2014) introduced the following covariate balancing propensity score method.

2.2 Covariate Balancing Propensity Score

By the definition of propensity score and the law of iterated expectations, for any continuous function $f(\cdot): R^p \to R^m$, it holds that

$$E\left[\frac{Tf(X)}{\pi(X)}\right] = E\left[\frac{(1-T)f(X)}{1-\pi(X)}\right] = E[f(X)]. \tag{5}$$

Indeed, the idea in (5) can balance finite order moments of the covariates by taking $f(X) = X^k$ for $k \geq 1$. To be specific, the CBPS method first assumes the logistic model as in (4) for propensity score, then plugging the propensity score function into the framework of generalized method of moments to obtain the estimator of β, denoted as $\hat{\beta}_{cbps}$. Substituting $\hat{\beta}_{cbps}$ into (4), one can obtain the estimated propensity score function $\hat{\pi}_{cbps}$. Finally, the ATE estimator based on the CBPS method is given by

$$\hat{\Delta}_{cbps} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i}{\hat{\pi}_{cbps}(X_i)}Y_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\hat{\pi}_{cbps}(X_i)}} - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1-T_i)}{1-\hat{\pi}_{cbps}(X_i)}Y_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\frac{1-T_i}{1-\hat{\pi}_{cbps}(X_i)}}.$$

2.3 Integrated Propensity Score

Clearly, the CBPS method is only limited to finite order moments. In fact, the covariate balance indicates that the propensity score can balance the entire distribution of the covariates among different groups. To this end, Sant'Anna et al. (2020) proposed the integrated propensity score method. Similar to the CBPS method, the IPS approach assumes the propensity score as the logistic model and then tries to balance the entire distribution of the covariates among different groups. The basic idea is to replace $f(\cdot)$ in (5) with parametric functions such as the indicator functions $I(X \leq u), u \in (-\infty, \infty)^p$ or the projection functions $I(\alpha^{\mathrm{T}} X \leq u), \alpha \in \{\alpha \in \mathbb{R}^p : ||\alpha|| = 1\}, u \in (-\infty, \infty)$, and then integrate over u to obtain a loss function with respect to β. By solving the minimization problem of the sample analogue of the loss function, one can obtain the estimator of β, denoted as $\hat{\beta}_{ips}$. Plugging $\hat{\beta}_{ips}$ into (4), one can obtain the estimated propensity score function $\hat{\pi}_{ips}$. Finally, the ATE estimator based on the IPS method is given by

$$\hat{\Delta}_{ips} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i}{\hat{\pi}_{ips}(X_i)}Y_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\frac{T_i}{\hat{\pi}_{ips}(X_i)}} - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(1-T_i)}{1-\hat{\pi}_{ips}(X_i)}Y_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\frac{1-T_i}{1-\hat{\pi}_{ips}(X_i)}}.$$

2.4 Stable Balance Weighting

Recently, Zubizarreta (2015) proposed a weighting procedure, which directly constrains covariate imbalances and minimizes the the variance of the weights, termed as stable balance weighting method. Unlike the CBPS and IPS method, the SBW method is model-free. Specifically, the stable balance weights are constructed by restricting the distance of the sample moments of the covariates between the treated group and combined group or between the control and combined groups. The optimization goal is to minimize the variance of the required weights. By solving the constrained optimization problem, one obtains the weights with minimum variance and balancing covariates as well. Denote the weights that balance covariates between the treated and combined groups as w_1 and the weights that balance covariates between the control and combined group as w_0. Then, the weighted estimator of the ATE based on the SBW method is given by

$$\hat{\Delta}_{sbw} = \sum_{i=1}^{n} w_{1i} Y_i - \sum_{i=1}^{n} w_{0i} Y_i.$$

2.5 Calibration Balance Weighting

Another covariate balance weighting procedure that circumvents the modeling of propensity score is the so-called calibration balance weighting method proposed by Chan et al. (2016). The CBW method includes a wide class of calibration weights, such as exponential tilting, empirical likelihood and continuous updating estimator of GMM, and attains an exact three-way balance of the moments of observed covariates among the treated, control and the combined groups. Specifically, the calibration balance weights are constructed through two restriction equations. The first one achieves the covariate balance between the treated and combined groups and the other one achieves the covariate balance between the control and combined groups. The optimization goal is to minimize the calibration distance between the required weights and the uniform design weights. Denote the weights that balance covariates between the treated and combined groups as σ_1 and the weights that balance covariates between the control and combined groups as σ_0. Then, the weighted estimator of the ATE based on the CBW method is given by

$$\hat{\Delta}_{cbw} = \sum_{i=1}^{n} \sigma_{1i} Y_i - \sum_{i=1}^{n} \sigma_{0i} Y_i.$$

3 Monte Carlo Simulations

Let $X_i = (X_{i1}, X_{i2}, X_{i3})^\top$ be independently distributed as $N(0, I_3)$, where I_3 is 3×3 identity matrix, $\varepsilon_i(1)$ and $\varepsilon_i(0)$ be independently distributed as $N(0, 1)$ and independent with X_i and let $\varphi(x)$ be the density distribution of the standard normal. The data generating processes of the potential outcomes and the propensity score are set up as follows.

Design 1:

$$Y_i(1) = 100 + 5X_{i1} + 7X_{i2} + 9X_{i3} + \varepsilon_i(1), \quad \text{and} \quad Y_i(0) = X_{i1} + X_{i2} + X_{i3} + \varepsilon_i(0), \tag{6}$$

where $\pi(X_i)$ is the logist regression as in (4) with $\beta = (0.5, 0.3, -0.7)$.

Design 2: $Y_i(1)$ and $Y_i(0)$ are generated by (6) with $\pi(X_i)$ set as below

$$\pi(X_i) = \frac{\exp[1.7 \exp(X_{i1}^2) + 2.5 X_{i2} X_{i3}]}{1 + \exp[1.7 \exp(X_{i1}^2) + 2.5 X_{i2} X_{i3}]}. \tag{7}$$

Design 3: $Y_i(1)$ and $Y_i(0)$ are generated by (6) with $\pi(X_i) = \varphi(X_{i1} X_{i2} + 2X_{i3})$.

Design 4: With $\pi(X_i)$ set as same as in Design 1,

$$\begin{aligned} Y_i(1) &= 100 + 5X_{i1} + 7X_{i2} + 9X_{i3} + 11X_{i1}X_{i2} + X_{i3}^2 + \varepsilon_i(1), \\ Y_i(0) &= X_{i1} + X_{i2} + X_{i3} + X_{i1}X_{i2} + X_{i3}^2 + \varepsilon_i(0), \end{aligned} \tag{8}$$

Design 5: $Y_i(1)$ and $Y_i(0)$ are generated by (8) with $\pi(X_i)$ set as as same as in (7).

In all simulation designs, the treatment variable $T_i \sim B(1, \pi(X_i))$, where $B(1, \pi(X_i))$ is the Bernoulli distribution with probability $\pi(X_i)$, the observed outcome variable is $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ and the final data set we observed is $\{(T_i, Y_i, X_i) : i = 1, \cdots, n\}$. It can be easily calculated that the true ATE is equal to 100 ($\Delta = 100$) in all simulation designs. For simplicity, the parameter function family used in the IPS method is the family of projection functions and the weights used in the CBW method correspond to the implied weights of the exponential tilting. For each design, we conduct 1000 Monte Carlo simulations with sample size $n = 200$, 400 and 800. Next, we calculate the median of the 1000 absolute deviation errors (ADE) ($|\hat{\Delta} - \Delta|$) and its standard deviation (SD) in parentheses and the results are displayed in Table 1.

Table 1: Mdedian and SD of 1000 ADE values for different weighting estimators

		MLE	CBPS	IPS	SBW	CBW
Design 1	$n = 200$	0.642 (0.599)	0.617 (0.533)	0.626 (0.541)	0.532 (0.477)	0.538 (0.473)
	$n = 400$	0.465 (0.435)	0.441 (0.388)	0.455 (0.396)	0.385 (0.331)	0.393 (0.333)
	$n = 800$	0.294 (0.273)	0.278 (0.254)	0.296 (0.263)	0.258 (0.233)	0.243 (0.220)
Design 2	$n = 200$	0.901 (1.052)	0.529 (0.470)	0.613 (0.545)	0.559 (0.514)	0.537 (0.464)
	$n = 400$	0.577 (0.617)	0.379 (0.327)	0.401 (0.366)	0.436 (0.363)	0.397 (0.346)
	$n = 800$	0.400 (0.383)	0.240 (0.219)	0.319 (0.264)	0.269 (0.278)	0.253 (0.228)
Design 3	$n = 200$	0.615 (0.589)	0.512 (0.472)	0.650 (0.584)	0.547 (0.493)	0.512 (0.479)
	$n = 400$	0.423 (0.361)	0.355 (0.325)	0.477 (0.433)	0.398 (0.360)	0.399 (0.336)
	$n = 800$	0.284 (0.251)	0.228 (0.208)	0.332 (0.291)	0.263 (0.240)	0.257 (0.221)
Design 4	$n = 200$	1.115 (1.148)	1.093 (1.021)	1.174 (0.987)	1.059 (0.902)	1.039 (0.869)
	$n = 400$	0.866 (0.868)	0.826 (0.730)	0.792 (0.797)	0.777 (0.709)	0.703 (0.669)
	$n = 800$	0.575 (0.566)	0.541 (0.520)	0.577 (0.578)	0.635 (0.531)	0.523 (0.485)
Design 5	$n = 200$	1.590 (1.771)	1.302 (1.143)	1.304 (1.146)	1.379 (1.186)	1.331 (1.164)
	$n = 400$	1.077 (1.035)	0.913 (0.815)	0.982 (0.826)	0.958 (0.844)	0.916 (0.836)
	$n = 800$	0.830 (0.699)	0.724 (0.627)	0.760 (0.619)	0.769 (0.641)	0.729 (0.635)

In terms of the median of the ADE, we can draw the following conclusions. First, the performance of the estimators based on the covariate balance weighting methods is better than that based on the MLE method without covariate balance expect the estimators based on the SBW method in Design 4 ($n = 800$) and the estimators based on the IPS method in Design 3, Design 1 ($n = 800$) and Design 4 ($n = 200$). Second, among the four covariate balance weighting methods, it is difficult to tell which covariate balance weighting method can outperform the others and they seem to be comparable to some extent. Third, based on Designs 1 and 4, it can be seen that the performance of the weighting methods is affected by the data generating process of the potential outcomes and this is consistent with Proposition 4.1 in Zubizarreta (2015) which provides a theoretical insight for this fact. Finally, based on Design 3, we find that if the propensity score seriously deviates from the logistic model, balancing the whole covariate distribution may result in the worse performance. In terms of the SD, the estimators based on the covariate balance weighting approaches can outperform that based on the MLE method without covariate balance expect the estimators based on the IPS method in Design 3 ($n = 400$ and $n = 800$) and Design 4 ($n = 800$), suggesting that the covariate balance weighting methods are generally more stable than the MLE weighting method. However, among the four covariate weighting methods, it is still not easy to tell which covariate balance weighting method can outperform the others while they seem to be comparable to some extent.

4 Conclusion

The CBPS method and IPS methods exploit the covariate balance of propensity score while the SBW and CBW procedures aim at obtaining the covariate balance weights directly. Through various simulation studies, we find that the performance of the covariate balance weighting methods is generally better than that of the MLE weighting method without covariate balance in terms of the median and standard deviation of the ADE. Another finding is that the performance of the four covariate balance weighting methods varies with the data generating processes of the potential outcomes and the propensity score and none of them can dominate the others in all simulation designs. Finally, although the performance of the estimators based on different covariate balance weighting methods has their own advantages and disadvantages, generally speaking, their performance does not differ too much from each other. This is because although these methods are proposed from different perspectives, they are all based on the idea of balancing the covariates among different groups as much as possible.

References

Chan, K.C.G., S.C.P. Yam and Z. Zhang (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *Journal of the Royal Statistical Society, Series B*, 78(3), 673.

Imai, K. and M. Ratkovic (2014). Covariate balancing propensity score. *Journal of the*

Royal Statistical Society, Series B, 76(1), 243-263.

Kang, J.D. and J.L. Schafer (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4), 523-539.

Sant'Anna, P.H., X. Song and Q. Xu (2020). Covariate distribution balance via propensity scores. arXiv preprint arXiv:1810.01370v4.

Wan, S.-K, Y. Xie and C. Hsiao (2018). Panel data approach vs synthetic control method. *Economics Letters*, 164(3), 121-123.

Zubizarreta, J.R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511), 910-922.