Economics Bulletin

Volume 40, Issue 1

A note on Covariate Balancing Propensity Score and Instrument-like variables

Adeola Oyenubi
School of Economics and Fiance, University of the Witwatersrand

Abstract

We use the term instrument-like variables to describe a variable that is highly correlated with treatment (or programme participation) and weakly correlated with outcome. This kind of variable cannot be used in instrumental variable estimation because they are not instruments and the literature also show that they should not also be used in Propensity Score Matching (PSM) because of their high correlation with treatment. The literature is therefore silent on the estimation approach that performs better when it is necessary to control for such an instrument-like variable. In this paper, we consider the estimation of treatment effect in the presence of an instrument-like variable and an unobserved confounder. The result shows that a particular variant of propensity score estimation namely Covariate Balancing Propensity Score (CBPS) performs better than alternatives in the presence of instrument-like variable and an unmeasured confounder. Our simulation result suggests that by trading off treatment prediction for balance CBPS reduces the influence of instrument-like variables on the propensity scores. This leads to lower bias and mean square error for the estimate that is based on the CBPS model.

Citation: Adeola Oyenubi, (2020) "A note on Covariate Balancing Propensity Score and Instrument-like variables", *Economics Bulletin*, Volume 40, Issue 1, pages 202-209
Contact: Adeola Oyenubi - adeola.oyenubi@wits.ac.za.
Submitted: August 24, 2019. **Published:** February 05, 2020.

1. Introduction

Rosenbaum and Rubin (1983) show that under the Conditional Independence Assumption (CIA) and the Common Support Assumption (CSA), matching on (a one dimensional) propensity score suffices to adjust for differences in covariates. There are a number of simulation studies that have investigated the performance of Propensity Score Matching (PSM) estimators in terms of bias and mean square error (MSE) (for example, see Busso et al, (2014), Zhao (2004) and Frolich (2004). One estimation concern that has been raised by these studies (and others) is how to choose variables to be included in propensity score specification (see Cuong (2013) for example). The literature strongly recommends that only variables that affect both programme participation and outcome should be used in propensity score estimation (Cuong, 2013; Kang & Schafer, 2007; Lechner, 2008; Caliendo and Kopeinig, 2008). However, this advice seems to ignore the fact that correlation with participation and outcome can be thought of as a spectrum. By this, we mean that there are variables that are highly correlated with participation but only weakly correlated with outcome (we refer to this kind of variable as being instrument-like hereafter) while the converse is also possible.

The literature provides direction on how to handle the extreme cases, i.e. variables that are correlated with participation or outcome alone (Bhattacharya and Vogt, 2007; Augurzky and Schmidt, 2001; Cuong, 2013). In the case of the former (variables that affect participation alone) this variable can be thought of as an instrument because it only affects the outcome through participation. Bhattacharya and Vogt (2007) show that absent strong ignorability[1], using PSM yields greater inconsistency than the naïve estimator when an instrument is used in propensity score estimation. Cuong (2013) show that in the case of the latter (variables that affect outcome alone) there are efficiency gains when such variables are used in the estimation of propensity scores. In the case of variables that are highly correlated with participation but weakly correlated with outcome (instrument-like variables), Zhao (2004) show that when strong ignorability is satisfied, such variables increase bias but reduces the standard error of matching estimators.

In this note, we focus on the case where ignorability is not satisfied (similar to Bhattacharya and Vogt (2007)) and there is a variable that is highly correlated with participation but only weakly correlated with the outcome. Under such conditions, the variable does not qualify as an instrument and given the advice in the literature (Bhattacharya and Vogt, 2007), it should not be included as a covariate in propensity score estimation. However, this variable may be necessary to make the conditional independence assumption more plausible. The key question in this paper is, what is the best estimator under these set of conditions?

Our simulation result shows that Covariate Balancing Propensity Score (CBPS) performs better than other estimators (i.e. instrumental variable and PSM based on logit model) that may be appealing under the stated set of conditions. The rest of the paper is organized as follows: section 2 provides a brief review of CBPS. Section 3 discusses our Data Generating Process (DGP) for the simulation study. Section 4 presents the results and section 5 concludes.

2. Brief review

There has been a lot of developments in the matching literature (for example, see Diamond and Sekhon (2013), Imai and Ratkovic (2014) and Iacus et. al (2012)). The one we focus on in the light of the problem just discussed is called the Covariate Balancing Propensity Score (CBPS). Introduced by Imai and Ratkovic (2014), CBPS exploits the dual characteristics of propensity

[1] i.e. when there is an unmeasured confounder

scores as a covariate balancing score and the conditional probability of treatment assignment. CBPS extends logistic regression to simultaneously optimize covariate balance and treatment prediction (Wyss et. al, 2014). This is achieved by trading off some accuracy in treatment assignment precision in order to satisfy the balancing condition[2].

Imai and Ratkovic (2014) show that CBPS performs better in terms of bias and MSE when compared to the conventional logit estimation of propensity score (under strong ignorability). They also show that the method is robust to propensity score model misspecification. In this note, we show that when there is an unmeasured confounder and an instrument-like variable, CBPS outperforms the conventional logit model and the Instrumental Variable (IV) approach in terms of bias and Mean Square Error (MSE). In providing a plausible explanation for our finding, we argue that trading-off precision in predicting the probability of participation in favour of satisfying the balancing condition becomes important in the context of variables that are highly correlated with participation but weakly correlated with outcome. Specifically, this trade-off eases imbalance by dampening the influence of the instrument-like variable on the propensity score distribution.

A logistic propensity score model uses Maximum Likelihood Estimation (MLE), which is designed to find parameter estimates by maximizing the assumed likelihood function. When the model is misspecified, parameter estimates that maximize the fit of the data, or minimize prediction error may not correspond to the parameter estimate that minimizes imbalance (Wyss et. al, 2014). CBPS replaces MLE with a Generalized Method of Moments (GMM) estimation to simultaneously optimize prediction of treatment assignment and covariate balance (Imai and Ratkovic, 2014, Wyss et. al, 2014)[3]. The implication of this is that the optimal parameters (betas) under the CBPS and the conventional logit model will be different (see Oyenubi (2019) for example). Specifically, the influence of variables in the logit model is unrestricted while under CBPS this influence is mediated by the balancing condition (one can think of this as constrained versus unconstrained optimization where the constraint is the balancing condition). When there is a variable that is highly correlated with participation and weakly correlated with outcome, the influence of such variables (which are likely to increase imbalance by creating support problems) is dampened by balance consideration under CBPS. Better balance will then lead to less bias and a more precise estimate of treatment effect.

Lechner (2008 pg 224) noted that from the small sample perspective[4], it is preferable to omit some good predictors of treatment which have only a small influence on the outcome. The problem with such variables is that just like instruments they are likely to create a common support problem by reducing the randomness required for identification by the CIA. In other words, in an attempt to make the CIA more plausible, common support may be compromised. Note that this compromise may come in the form of thin support problem which increases biases and variances of estimators (e.g. Crump et al, 2009; Khan and Tamer, 2010). However, unlike instruments, because instrument-like variables are correlated with the outcome (howbeit weakly) they cannot be used for instrumental variable estimation. We show that absent ignorability, placing a little less emphasis on predicting the probability of treatment, seem to help CBPS handles instrument-like variables better by dampening their influence on the propensity scores.

[2] One can think of the balancing condition as a constraint in the optimization that estimate probability of treatment assignment.
[3] See the paper Imai and Ratkovic (2014) for detailed description of the CBPS method and the generalized methods of moments estimation.
[4] But not a asymptotic one

Bhattacharya and Vogt (2007) consider a case where ignorability is not satisfied and they show that using an instrument to estimate propensity score lead to inconsistency in the effect estimate. Heckman and Navarro-Lozano (2004) consider a similar situation and show that using a variable that is highly correlated with treatment, but weakly correlated with the outcome also increases inconsistency of the effect estimate. In this paper, we show that when ignorability is not satisfied and there is an instrument-like variable, CBPS improves the effect estimate from matching relative to the conventional logistic approach (to estimating propensity scores) and the instrumental variable approach.

3. Simulation Study

The simulation study follows the work of Bhattacharya and Vogt (2007). Let z be an instrument-like variable, ϵ_1 and ϵ_2 are the error terms in the outcome and treatment equation, d is an indicator for treatment, and y is the outcome variable. We assume the following data generating process for the Monte Carlo experiments:

$$z \sim beta(0.05, 0.05) * 8$$ [5]

$$(\epsilon_1, \epsilon_2) \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$$

$$d = -4 + z\beta_0 + x_1\beta_1 + x_2\beta_2 - \epsilon_1$$

$$y = \theta d + \gamma z + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \epsilon_2$$

$x_i \sim N(0,1)$, $i = 1$ *to* 3. x_1 and x_2 are correlated with both treatment and outcome while x_3 is correlated with treatment alone. The parameter ρ describes the correlation between ϵ_1 and ϵ_1, we consider $\rho = 0, 0.2, 0.4, 0.6, 0.8$. When $\rho = 0$, strong ignorability holds, as ρ increases, the strength of the unobserved confounder increases. To vary the correlation of the instrument-like variable with the outcome we consider $\gamma = 0$ *and* 0.2.

The remaining parameters are set as follows $\theta = 0.1$[6], $\beta_0 = 1, \beta_1 = 0.1, \beta_2 = 0.1, \beta_3 = 4$. Each simulation contains 400 observations with about 80% in the treatment group, this is done to mimic the situation in which there is a scarcity of good control units. The simulation is repeated 500 times, in each case, we match using propensity scores obtained from CBPS and logit model, we also use the IV estimator (using z as an instrument) to calculate the treatment effect.

One can think of the beta estimate from a logit or CBPS model as weights assigned to covariates in order to reduce the dimension of the covariates. Variables that are highly correlated with participation will, therefore, have high betas. To get a sense of the influence of the instrument-like covariate on the propensity scores, we calculate $\frac{z}{\sum_i \beta_i + z}$ i.e. the proportion of the total weight due to the instrument-like variable or the influence of z on the propensity scores.

What we expect is that since CBPS trades-off likelihood maximization for balance, this trade-off will imply lower influence for variables that are highly correlated with treatment (in general). The lower influence will be off-set by improved balance. So that relative to the

[5] For the simulation in Bhattacharya and Vogt (2007) z follows an exponential distribution. We chose to make it a bimodal distribution which inform the choice of beta distribution with the listed parameters. This induces a difference in means of this variable across treatment arms which makes the variable a good predictor of treatment. While we have generated this characteristic artificially, it is not strange to find variable with significant difference in means across treatment arms in observational studies.

[6] Note that this means that the treatment effect is constant over the support.

conventional logit model we expect the influence of z to be lower in the CBPS model. This lower influence and better balance should lead to better behaviour for the estimator relative to the logit model and the IV approach (note that an instrument-like variable should fail under the IV approach since it is not a proper instrument). In the extreme case where $\gamma = 0$, z is a proper instrument and we expect the IV approach to dominate other estimators.

4. Results

Table 1 presents the simulation results. There are two panels, panel 1 shows the result when $\gamma = 0$ and panel 2 shows the result when $\gamma = 0.2$. The last three rows in each panel show the correlation between treatment and the instrument-like variable ($\boldsymbol{corr(z,d)}$), the correlation between the outcome and the instrument-like variable ($\boldsymbol{corr(z,y)}$) and the percentage of treatment observations (ratio(d/n)) respectively. It is clear that in both panels that z is highly correlated with d but weakly correlated with y. Note that in panel 1 the correlation coefficient $\boldsymbol{corr(z,y)}$ is very close to zero this is because we set $\gamma = 0$ in panel 1 which means z is a proper instrument (by design). The next three rows (rows 6,7 and 8 from the top) show the MSE for the different estimators (i.e. CBPS MSE, logit MSE and IV MSE). Rows 3, 4 and 5 show the bias of the different estimators (CBPS bias, logit bias, IV bias) while the first two rows show the percentage of total beta attributable to z.

Table 1: Simulation Results

			$\boldsymbol{\gamma = 0}$		
$\boldsymbol{\rho}$	**0**	**0.2**	**0.4**	**0.6**	**0.8**
		Panel 1			
CBPS beta %	0.5431	0.5255	0.5370	0.5304	0.5306
logit beta %	0.6315	0.6254	0.6326	0.6197	0.6254
CBPS Bias	0.6106	0.6356	0.7101	0.7381	0.8410
logit Bias	0.7344	0.7278	0.7676	0.7795	0.9047
IV Bias	0.1357	0.1324	0.1268	0.1296	0.1344
CBPS MSE	0.6364	0.6903	0.8674	0.8578	1.1165
logit MSE	0.8693	0.8344	0.9956	1.0394	1.2947
IV MSE	0.0283	0.0275	0.0254	0.0278	0.0297
$\boldsymbol{corr(z,d)}$	0.8622	0.8650	0.8640	0.8649	0.8646
$\boldsymbol{corr(z,y)}$	0.0095	0.0060	0.0126	0.0153	0.0099
ratio (d/n)	0.8311	0.8337	0.8357	0.8323	0.8268
			$\boldsymbol{\gamma = 0.2}$		
		Panel 2			
CBPS beta %	0.5431	0.5255	0.5370	0.5304	0.5306
logit beta %	0.6315	0.6254	0.6326	0.6197	0.6254
CBPS Bias	0.6104	0.6353	0.7060	0.7323	0.8347
logit Bias	0.7350	0.7281	0.7680	0.7796	0.9060

IV Bias	1.7707	1.7824	1.7860	1.7895	1.7782
CBPS MSE	0.6320	0.6881	0.8591	0.8467	1.0993
logit MSE	0.8703	0.8352	0.9961	1.0389	1.2987
IV MSE	3.1655	3.2078	3.2196	3.2328	3.1965
$corr(z, d)$	0.8622	0.8650	0.8640	0.8649	0.8646
$corr(z, y)$	0.1861	0.1823	0.1885	0.1912	0.1860
ratio (d/n)	0.8311	0.8337	0.8357	0.8323	0.8268

Panel 1 shows that when z is a proper instrument the IV approach yields the lowest bias and MSE (which is expected). Furthermore, as the strength of the unmeasured confounder increases the performance of the CBPS and conventional logit model deteriorates. This is also expected since these estimators assume ignorability. However, we note that the CBPS approach performs better than the conventional logit model across panel 1 (in terms of bias and MSE).

The central result of this paper is presented in panel 2. For panel 2, z is not a proper instrument. Therefore, both the bias and MSE for the IV estimate increase considerably (relative to panel 1), while the result for the other two estimators remains largely unchanged. Furthermore, similar to panel 1 the CBPS continues to outperform the logit model. This means that in the presence of an unmeasured confounder and an instrument-like variable the matching estimate from the CBPS model outperforms the other estimators.

Rows 1 and 2 of both panels provide a plausible explanation for why the CBPS performs better than the conventional logit model in the presence of an instrument-like variable and an unmeasured confounder (the reason why it outperforms the IV approach is obvious). Note that the influence of the beta estimate for z is on average 10% higher under the conventional logit model compared to the CBPS approach. Also note that the beta due to z is the largest influence on the predicted probabilities accounting for more than 50% of the total sum of betas. This means that the influence of z on the support of the propensity score density is reduced in the CBPS model, and this will enhance pre-matching and post-matching balance. All things being equal as β_0 increases, there will be more selection due to z[7], this will increase the variance of d and stretch out the propensity score distribution (across treatment arms). A similar argument can be found in Zhao (2004; pg 97)[8].

The point here is that by dampening the effect of the instrument-like variables CBPS makes the (pre-matching) propensity score densities "closer" to each other. This enhances overall balance and improves the performance of the matching estimator. CBPS can, therefore, enhance balance by reducing the influence of an instrument-like variable on the propensity scores. The fact that CBPS improves the performance of matching estimators has been highlighted by Imai and Ratkovic (2014), what our result shows is that in the presence of an

[7] Note that more selection in general means increased disparity between the covariates (across treatment arms) or the propensity score density that summarizes the covariates. In results not presented here but available on request from the author, we calculate the Kolmogorov Smirnov distance and Standardized difference in means between the propensity score densities across treatment arms. The result show that the pre-matching distances between the densities under CBPS is on average slightly less than the pre-matching distance between the densities under the logit model. This is the case even though we use the same specification and covariates. However, this is not surprising since CBPS is constrained by the balancing condition.

[8] Specifically paragraph 2 of the cited paper discusses how decreasing a parameter g (which is equivalent to β_0 in our set up) can decrease the variance of the propensity score distribution

unmeasured confounder and an instrument-like variable, CBPS performs better than competing models that may be appealing under such conditions.

Next, we investigate if there is a threshold for γ beyond which CBPS becomes the best estimator in the class of estimators considered in this study. Recall that when $\gamma = 0$ IV dominates the other estimators and when $\gamma = 0.2$ CBPS dominates the other estimators . This means that there is some threshold at which the dominance shifts from IV to CBPS. To explore this table 2 present the results presented in table 1 when $\gamma = 0.05\ and\ 0.1$. For $\gamma = 0.05$ (in table 2) the IV estimator still dominates other estimators (but the size of bias has increased relative to $\gamma = 0$). While for $\gamma = 0.1$ the CBPS performs better. Therefore for the configuration of data in our simulation, somewhere between $\gamma = 0.05$ and $\gamma = 0.1$ CBPS should be preferred to the IV estimator. We note that this range may be different under a different set of data configuration.

Table 2: Simulation Results

			$\boldsymbol{\gamma = 0.05}$		
$\boldsymbol{\rho}$	**0**	**0.2**	**0.4**	**0.6**	**0.8**
		Panel 1			
CBPS beta %	0.5431	0.5255	0.5370	0.5304	0.5306
logit beta %	0.6315	0.6254	0.6326	0.6197	0.6254
CBPS Bias	0.6105	0.6355	0.7090	0.7366	0.8393
logit Bias	0.7346	0.7279	0.7677	0.7795	0.9050
IV Bias	0.4377	0.4537	0.4555	0.4595	0.4472
CBPS MSE	0.6351	0.6896	0.8651	0.8549	1.1119
logit MSE	0.8695	0.8346	0.9957	1.0393	1.2957
IV MSE	0.2196	0.2333	0.2332	0.2389	0.2298
$\boldsymbol{corr(z,d)}$	0.8622	0.8650	0.8640	0.8649	0.8646
$\boldsymbol{corr(z,y)}$	0.0544	0.0508	0.0574	0.0601	0.0547
ratio (d/n)	0.8311	0.8337	0.8357	0.8323	0.8268
			$\boldsymbol{\gamma = 0.1}$		
		Panel 2			
CBPS beta %	0.5431	0.5255	0.5370	0.5304	0.5306
logit beta %	0.6315	0.6254	0.6326	0.6197	0.6254
CBPS Bias	0.6105	0.6354	0.7080	0.7352	0.8375
logit Bias	0.7347	0.7279	0.7678	0.7795	0.9053
IV Bias	0.8818	0.8964	0.8989	0.9028	0.8905
CBPS MSE	0.6339	0.6890	0.8630	0.8520	1.1076
logit MSE	0.8698	0.8348	0.9958	1.0391	1.2967
IV MSE	0.8062	0.8320	0.8348	0.8434	0.8244

$corr(z,d)$	0.8622	0.8650	0.8640	0.8649	0.8646
$corr(z,y)$	0.0990	0.0953	0.1018	0.1045	0.0991
ratio (d/n)	0.8311	0.8337	0.8357	0.8323	0.8268

5. Conclusion

This paper considers the estimation of treatment effect when there is an unmeasured confounder and a variable that is highly correlated with participation but does not qualify to be an instrument because it is also weakly correlated with outcome. Specifically, in the presence of such instrument-like variable, there is no clear advise in the literature as to what the preferred estimator should be. Instrumental variable and PSM (in general) are approaches that might be attractive in such situations. We show through a simulation study that a particular variant of the propensity score estimation method namely CBPS works best under the stated conditions.

Our result suggests that CBPS works better under the stated condition because it trades-off treatment prediction for balance and this is particularly important when instrument-like variables are needed to justify the conditional independence assumption. The result echoes the point made by Imai and Ratkovic (2014) i.e. CBPS will improve the balance of observed covariates even if there are unmeasured confounders.

Finally, as noted by Wyss et. al (2014) it is unlikely that a single propensity score estimation method will be optimal in every setting. More work is needed to better understand the performance of different propensity score estimators. We see this work as providing a set of conditions under which CBPS works better than competing estimators.

Bibliography

Angrist, J.D., Imbens, G.W. and Rubin, D.B., 1996. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), pp.444-455.

Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J. and Stürmer, T., 2006. Variable selection for propensity score models. *American journal of epidemiology*, *163*(12), pp.1149-1156.

Busso, M., DiNardo, J. and McCrary, J., 2014. New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics, 96(5), pp.885-897.*

Caliendo, M. and Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), pp.31-72.

Crump, R.K., Hotz, V.J., Imbens, G.W. and Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, *96*(1), pp.187-199.

Cuong, N.V., 2013. Which covariates should be controlled in propensity score matching? Evidence from a simulation study. *Statistica Neerlandica*, *67*(2), pp.169-180.

Diamond, A. and Sekhon, J.S., 2013. Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, *95*(3), pp.932-945.

Frölich, M., 2004. Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics*, 86(1), pp.77-90.

Heckman, J. and Navarro-Lozano, S., 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and statistics*, *86*(1), pp.30-57.

Hirano, K. and Imbens, G.W., 2001. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology*, *2*(3-4), pp.259-278.

Iacus, S.M., King, G. and Porro, G., 2012. Causal inference without balance checking: Coarsened exact matching. *Political analysis*, *20*(1), pp.1-24.

Imai, K. and Ratkovic, M., 2014. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), pp.243-263.

Imbens, G.W. and Angrist, J.D., 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), pp.467-475.

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

Khan, S. and Tamer, E., 2010. Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, *78*(6), pp.2021-2042.

Lechner, M., 2008. A note on the common support problem in applied evaluation studies. *Annales d'Économie et de Statistique*, pp.217-235.

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), pp.41-55.

Schmidt, C.M. and Augurzky, B., 2001. The propensity score: A means to an end. *Tech. rep., IZA Discussion Paper Series*.

Weitzen, S., Lapane, K.L., Toledano, A.Y., Hume, A.L. and Mor, V., 2004. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiology and drug safety*, *13*(12), pp.841-853.

Wyss, R., Ellis, A.R., Brookhart, M.A., Girman, C.J., Jonsson Funk, M., LoCasale, R. and Stürmer, T., 2014. The role of prediction modeling in propensity score estimation: an evaluation of logistic regression, bCART, and the covariate-balancing propensity score. *American journal of epidemiology*, *180*(6), pp.645-655.

Oyenubi, A., 2019. Optimising balance using covariate balancing propensity score: The case of South African child support grant. Development Southern Africa, pp.1-17.

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *Review of economics and statistics, 86*(1), 91-107