Covariate Balancing Propensity Score for General Treatment Regimes*

Christian Fong† Kosuke Imai‡

First Draft: July 31, 2014
This Draft: August 30, 2014

Abstract

Propensity score matching and inverse-probability weighting are popular methods for causal inference in observational studies. Under the assumption of unconfoundedness, these methods enable researchers to estimate causal effects by balancing observed covariates across different treatment values. While their extensions to general treatment regimes exist, a vast majority of applications have been confined to a binary treatment. Moreover, applied researchers often dichotomize a non-binary treatment in order to utilize propensity score methods. Balancing covariates with respect to the dichotomized treatment, however, does not imply that they are balanced regarding the original non-binary treatment variable. In this paper, we extend the covariate balancing propensity score (CBPS) methodology of Imai and Ratkovic (2014) to general treatment regimes. Specifically, we estimate the generalized propensity score such that the resulting association between a treatment and covariates is minimized. Two social science applications are used to demonstrate that the CBPS methodology significantly improves covariate balance and offer substantive insights the original analyses fail to identify. The proposed methodology is implemented through publicly available open-source software.

Key words: causal inference, covariate balance, generalized propensity score, inverse-probability weighting, treatment effect

---

*The proposed methods are implemented through open-source software `CBPS` (Fong *et al.*, 2014), which is freely available as an R package at the Comprehensive R Archive Network (CRAN http://cran.r-project.org/package=CBPS). We thank Marc Ratkovic and Dylan Small for their comments and suggestions.

†Graduate Student, Graduate School of Business, Stanford University, Stanford CA 94305. Email: christianfong@stanford.edu.

‡Professor, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 609–258–6601, Email: kimai@princeton.edu, URL: http://imai.princeton.edu

1 Introduction

Propensity score matching/subclassification and inverse-probability weighting are popular methods for causal inference in observational studies where researchers wish to infer the causal effects without randomizing treatment assignment (e.g., Rosenbaum and Rubin, 1983, 1984, 1985; Robins *et al.*, 2000; Hirano *et al.*, 2003). Under the assumption of unconfoundedness, the propensity score methods aim to balance observed covariates across different values of a treatment variable (e.g., Imbens, 2004; Ho *et al.*, 2007).

Despite the popularity of these propensity score methods, a vast majority of applications have been confined to a binary treatment. This dearth of applications for non-binary treatments cannot be explained by the unavailability of methodology. To the contrary, several researchers have extended propensity score methods to general treatment regimes. For example, inverse-probability weighting can be done with a multi-valued or even continuous treatment by using the estimated density as a basis of weights (e.g., Imbens, 2000; Robins *et al.*, 2000). Similarly, Imai and van Dyk (2004) consider subclassification on the propensity function for general treatment regimes while Hirano and Imbens (2004) proposes a regression adjustment based on the estimated generalized propensity score for a continuous treatment (see also Joffe and Rosenbaum, 1999; Lu *et al.*, 2001; Rassen *et al.*, 2013; Yang *et al.*, 2014).

All of these promising methods, however, presume the accurate estimation of an unknown (generalized) propensity score. However, this is not a trivial task. In fact, scholars have found that even in the case of binary treatment where relatively straightforward diagnostics tools are available, the empirical results can be sensitive to model misspecification (e.g., Smith and Todd, 2005; Kang and Schafer, 2007). This problem is exacerbated for non-binary treatments where checking covariate balance

is more difficult and less intuitive because the treatment variable takes more than two values, e.g., a continuum of values.

An important consequence of this complication is that applied researchers often dichotomize a non-binary treatment in order to utilize propensity score methods for binary treatments. Because many treatment variables of interest in social and medical sciences are non-binary, this practice of dichotomization can be found across a number of disciplines (e.g., Donohue III and Ho, 2007; Harder *et al.*, 2008; Boyd *et al.*, 2010; Nielsen *et al.*, 2011; De and Ratha, 2012). In Section 2, we present two motivating studies where the original authors recoded three-category and continuous treatment variables into binary variables and applied propensity score methods. Our analysis illustrates a general point that balancing covariates with respect to the dichotomized treatment variable does not imply that they are balanced regarding the original non-binary treatment variable. More importantly, the dichotomization of treatment variable results in the loss of information, which can compromise substantive insights gained from the data analysis.

To address this gap between methodological and applied research, we propose a new method to estimate the propensity score for general treatment regimes in Section 3. Specifically, we address the difficulty of checking covariate balance by directly minimizing the association between a treatment variable and covariates in order to estimate the (generalized) propensity score. This extends the covariate balancing propensity score (CBPS) methodology of Imai and Ratkovic (2014), who demonstrate the effectiveness of the methodology in the binary treatment case (see also Wyss *et al.*, 2014), to general treatment regimes.

Once researchers obtain the estimated propensity score using the CBPS methodology, they can use a variety of methods to estimate causal effects (e.g., Lu *et al.*, 2001; Hirano and Imbens, 2004; Imai and van Dyk, 2004; Rassen *et al.*, 2013; Yang *et al.*,

2014). In this paper, we focus on the inverse-probability weighting (Imbens, 2000; Robins *et al.*, 2000), as it is directly related to the covariate balance measure used in the CBPS estimation. The proposed methods are implemented through publicly available open-source software `CBPS` (Fong *et al.*, 2014).

Using the proposed methodology, in Section 4 we reanalyze the two motivating studies introduced without dichotomizing treatment variables. We first show that the CBPS reduces the association between the treatment variables and covariates more effectively than the standard estimation method. We then demonstrate that additional substantive insights can be obtained by analyzing the original non-binary treatment variables rather than their dichotomized versions. Finally, we offer concluding remarks in Section 5.

2 Motivating Applications

In this section, we introduce two empirical studies from political science that motivate our methodology. Both studies share the same problem in that their original analyses dichotomized non-binary treatment variables in order to apply standard propensity score methods. We show that this dichotomization yields two important issues. First, dichotomizing a non-binary treatment variable identifies a different causal quantity of interest. Second, balancing covariates with respect to dichotomized treatment variables does not achieve balance across different values of the original non-binary treatment variables. To illustrate these issues, we provide a brief theoretical overview and then discuss them in the context of the two motivating applications.

2.1 The Problems of Dichotomizing a Non-binary Treatment Variable

We first briefly review the relevant theoretical issues about dichotomizing a non-binary treatment variable. Comprehensive discussions can be found in Hernán and VanderWeele (2011) and VanderWeele and Hernán (2013). Suppose that we have a non-binary treatment T_i for unit i whose support is $\mathcal{T}$. Let $\widetilde{T}_i$ represent the dichotomized treatment variable such that $\widetilde{T}_i = 1$ ($\widetilde{T}_i = 0$) if $T_i \in \mathcal{T}_1$ ($T_i \in \mathcal{T}_0$) where $\mathcal{T}_1 \cup \mathcal{T}_0 = \mathcal{T}$ and $\mathcal{T}_1 \cap \mathcal{T}_0 = \emptyset$. Consider the strong ignorability assumption with respect to the original non-binary treatment variable,

$$T_i \;\perp\!\!\!\perp\; Y_i(t) \;\mid\; X_i \quad \text{and} \quad p(T_i = t \mid X_i) \;>\; 0 \quad \text{for all } t \;\in\; \mathcal{T} \tag{1}$$

where $Y_i(t)$ is the potential outcome given the treatment value $T_i = t$, and X_i is a vector of observed pre-treatment covariates. Note that the potential outcomes must be defined with respect to the original treatment variable in order to satisfy the stable unit treatment value assumption or SUTVA (Rubin, 1990). Furthermore, the conditional distribution of treatment $p(T_i \mid X_i)$ is called the *generalized propensity score* (Joffe and Rosenbaum, 1999; Imbens, 2000; Hirano and Imbens, 2004; Imai and van Dyk, 2004). Finally, as part of the SUTVA, we assume no interference among units.

Now, suppose that researchers compute the difference in average outcome given the covariates. Under the aforementioned assumptions, this estimator identifies the following causal quantity,

$$\begin{aligned} &\mathbb{E}(Y_i \mid \widetilde{T}_i = 1, X_i) - \mathbb{E}(Y_i \mid \widetilde{T}_i = 0, X_i) \\ &= \int_{\mathcal{T}_1} \mathbb{E}(Y_i(t) \mid X_i) p(T_i = t \mid \widetilde{T}_i = 1, X_i) dt \;-\; \int_{\mathcal{T}_0} \mathbb{E}(Y_i(t) \mid X_i) p(T_i = t \mid \widetilde{T}_i = 0, X_i) dt \end{aligned} \tag{2}$$

That is, when using a dichotomized treatment variable the causal interpretation of the usual difference-in-means estimator depends critically on the distribution of the original treatment variable given the dichotomized variable, i.e., $p(T_i \mid \widetilde{T}_i, X_i)$. This distribution is used to aggregate the causal quantity for the original treatment variable $\mathbb{E}(Y_i(t) \mid X_i)$. Because of this aggregation, the dichotomization may conceal important causal effects. In addition, external validity may be compromised if this key distribution, $p(T_i \mid \widetilde{T}_i, X_i)$, in the sample differs significantly from that in a target distribution.

Another difficulty associated with the dichotomization of a non-binary treatment variable concerns the covariate adjustment. In observational studies, the treatment assignment mechanism is unknown to the researchers. This means that the generalized propensity score must be estimated from the data. To avoid model misspecification, a common diagnostic is to check covariate balance by examining the association between the treatment variable and the covariates conditional on the estimated propensity score (see Hirano and Imbens, 2004; Imai and van Dyk, 2004). As we demonstrate below with examples, however, balancing covariates with respect to a dichotomized treatment variable may not balance the covariates regarding the original non-binary treatment variable. This means that the researchers must accurately estimate the generalized propensity score by modeling the original non-binary treatment variable. Developing such a method is the primary goal of the current paper. We now turn to two motivating applications.

2.2 The Effect of Education on Political Participation

Since as early as Berelson *et al.* (1954)'s foundational study on public opinion, political scientists have recognized the intimate relationship between education and political participation. Verba *et al.* (1995) went further, arguing that education has a special role in the acquisition of nearly every other facilitating factor of participation. Kam

and Palmer (2008) challenged this view and maintained that education serves largely as a proxy for pre-adult experiences and predispositions.

Kam and Palmer conducted a standard propensity score matching analysis. They used a dichotomous measure of educational attainment, indicating whether a person has attended college or not, and a combined index of participatory acts (such as voting, donating to a political campaign, and contacting a public official) as the main outcome variable. For this binary treatment variable, the propensity score was estimated using the logistic regression with dozens of covariates including opinions about the efficacy and fairness of government, knowledge of public affairs, participation in school and community organizations, and family background characteristics. Recently, others called their results into question, objecting that the original matching analysis fails to adequately balance the covariates (Henderson and Chatfield, 2011; Mayer, 2011). These critics employed an alternative matching method to achieve a better balance and found that attending college actually increases political participation.

While all of these analyses focus on the binary treatment variable, the original data set, the Political Socialization Study, also records whether a person graduated from a college rather than merely whether the person attended it. By dichotomizing, the researchers ignore a potentially important source of treatment heterogeneity. Those who graduate college may have significantly different educational experiences compared to those who attend college but do not graduate. As formally discussed in Section 2.1, if graduating rather than attending college has an impact, then merging the two groups might conceal the effect of education.

Using the original computer code used by the authors, we replicate the empirical results of Kam and Palmer (2008) as well as those of Henderson and Chatfield (2011). Kam and Palmer included 81 covariates in their standard propensity score

matching where the logistic regression was used to estimate the propensity score for the dichotomized treatment. In contrast, Henderson and Chatfield used the genetic matching of Diamond and Sekhon (2013) and with the same 81 covariates.

We examine the degree to which covariate imbalance remains with respect to the three-valued treatment variable even after matching adjustment based on the dichotomized treatment variable. We divide the sample into three treatment groups: those who did not attend college (no college), those who attended college but did not graduate (some college), and those who graduated from college (graduated). Note that we code graduating with an associate degree as "some college." As a measure of covariate imbalance, we use the absolute standardized difference in means between two groups. This measure is defined as follows,

$$\mathcal{I}_k(t, t') \;=\; \sqrt{\frac{N-1}{\sum_{i=1}^{N}(X_{ik} - \overline{X}_k)^2}} \left| \frac{\sum_{i \in \mathcal{M}} \mathbf{1}\{T_i = t\} X_{ik}}{\sum_{i \in \mathcal{M}} \mathbf{1}\{T_i = t\}} - \frac{\sum_{i \in \mathcal{M}} \mathbf{1}\{T_i = t'\} X_{ik}}{\sum_{i \in \mathcal{M}} \mathbf{1}\{T_i = t'\}} \right| \quad (3)$$

where $\overline{X}_k = \sum_{i=1}^{N} X_{ik}/N$ and $\mathcal{M}$ represents the set of matched observations.

The left plot of Figure 1 presents the covariate imbalance in the original data (white boxplot), the remaining imbalance after propensity score matching (grey; Kam and Palmer, 2008), and that after genetic matching (dark grey; Henderson and Chatfield, 2011). The plot shows that both matching techniques, based on the dichotomized treatment variable, leave a substantial amount of covariate imbalance with respect to the original three-category treatment variable. In fact, according to this particular measure, both matching methods worsen the covariate imbalance for every pair of treatment groups.

Examining each covariate shows that propensity score matching achieves better balance on the best predictor of college achievement (i.e., the respondent's current post-graduation plan) relative to the original data set. Nevertheless, the method also exacerbates the imbalance on many other covariates which include moderately strong predictors, such as the parents' partisan identification and the student's involvement

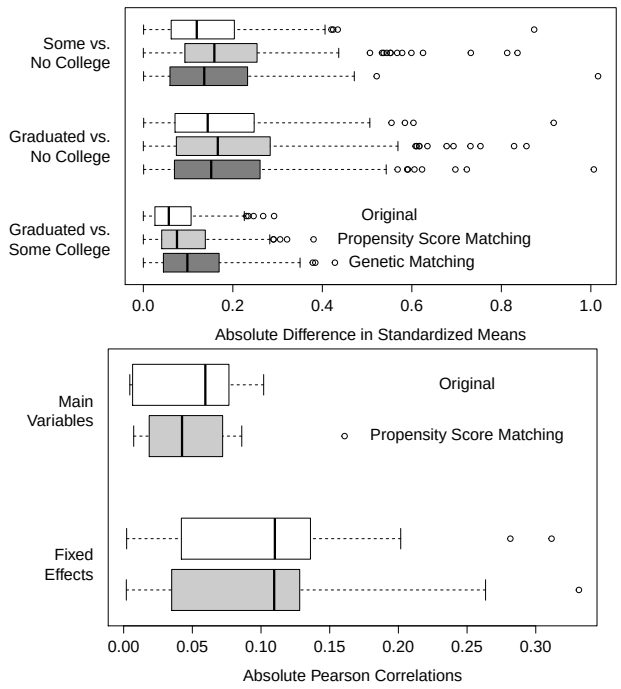

Figure 1: Covariate Imbalance in the Kam and Palmer (2008) Study and the Urban and Niebler (2014) Study after Covariate Adjustment with a Dichotomized Treatment. The plot on the left shows the covariate imbalance for the three-valued treatment before any adjustment (white boxplot), after Kam and Palmer's propensity score matching (grey), and after Henderson and Chatfield (2011)'s genetic matching (darkgrey). The plot on the right shows covariate imbalance for the continuous treatment before any adjustment (white boxplot) and after Urban and Niebler's propensity score matching (grey). Covariate imbalance with respect to the original treatment variables remains even after matching adjustment based on the dichotomized treatment variable.

in neighborhood clubs. Genetic matching, on the other hand, optimizes a different measure of balance and fails to reduce the imbalance in mean differences for many covariates, including the post-graduation plans variable.

In addition, the dichotomization of the treatment variable may fail to capture important causal relationships. If a college degree facilitates participation by giving graduates access to better careers, then there should be a positive effect for completing college but little effect for attending without finishing college. Conversely, if attending college socializes an individual to the tastes and habits of the middle or upper class, then simply attending college may have just as large of an effect as actually graduating. If college endows students with cognitive skills which are helpful

for engaging the political world, then attendance might have a positive (but, relative to graduating, smaller) effect. The dichotomization cannot distinguish between these causal predictions.

2.3 The Effect of Advertisements on Campaign Contributions

The second motivating application is the study of political advertisements by Urban and Niebler (2014). The authors explored the potential link between advertising and campaign contribution. Presidential campaigns ordinarily focus their advertising efforts on competitive states, but if political advertising drives more donations, then it may be worthwhile for candidates to also advertise in non-competitive states. Urban and Niebler exploit the fact that media markets sometimes cross state boundaries. This means that candidates may inadvertently advertise in non-competitive states when they purchase advertisements for media markets that serve competitive states. By restricting their analysis to non-competitive states, the authors attempt to isolate the effect of advertising from that of other campaigning, which do not incur these media market spillovers.

Although the original data set contains the number of advertisements aired in each zip code, Urban and Niebler dichotomized this political advertising variable by examining whether a zip code received more than 1,000 advertisements or not. According to this operationalization, 5,230 of 16,265 zip codes are classified as "treated." In contrast, the original variable ranges from 0 to 22,379 with the average number of advertisements being 1,902. Using this dichotomized treatment variable, the authors then conduct a standard propensity score matching method where the logistic regression is fitted to estimate the propensity score. The authors employ many different matching methods as robustness checks including kernel matching and nearest

neighbor matching based on propensity score. Our replication of their analysis uses one-to-one nearest neighbor propensity score matching. As formally argued in Section 2.1, such an analysis based on the dichotomized treatment may miss substantive insights that can be obtained by analyzing the original continuous treatment.

Along with the study described in Section 2.2, this empirical application also confirms the fact that balancing covariates with respect to the dichotomized treatment may not improve covariate balance regarding the original treatment. The right plot of Figure 1 presents the pairwise correlation between the original non-binary treatment variable and each covariate. The result shows that matching for the dichotomized treatment fails to balance covariates for the underlying non-binary treatment. On the whole, matching does not improve the balance of the state-level fixed effects or the main variables (log population, log income, percent over 65, percent black, percent Hispanic, percent college graduates, population density, and whether residents can reasonably commute to another state).

Elsewhere in their analysis, Urban and Niebler (2014) estimate the dose-response curve using the original non-binary treatment variable without matching. Thus, it is clear that the authors are interested in the underlying treatment variable rather than its binary version. For such an analysis, it is important for propensity score matching to be done with the original treatment variable rather than the dichotomized variable. The goal of this paper is to develop a method to reliably estimate the generalized propensity score when the treatment is not binary.

3 The Proposed Methodology

The motivating examples in Section 2 highlight the need for a methodology to estimate the propensity score for general treatment regimes. Currently, fitting a parametric model under the framework of maximum likelihood is the most commonly used

method (e.g., Joffe and Rosenbaum, 1999; Robins *et al.*, 2000; Lu *et al.*, 2001; Hirano and Imbens, 2004; Imai and van Dyk, 2004; Rassen *et al.*, 2013; Yang *et al.*, 2014). While, in theory, a semi-parametric or non-parametric method can be applied to the estimation of generalized propensity score, the high-dimensionality of covariates often makes such a modeling approach practically difficult.

In this section, we aim to improve the parametric estimation of generalized propensity score. Specifically, we extend the covariate balancing propensity score (CBPS) methodology of Imai and Ratkovic (2014) to general treatment regimes. The key feature of the proposed methodology is to estimate the generalized propensity score such that the resulting covariate balance is optimized. Since checking covariate balance is often difficult when the treatment variable takes more than two values, the CBPS should facilitate the use of generalized propensity score methods by applied researchers.

3.1 Multi-valued Treatment

We first develop the CBPS for a multi-valued treatment, which is applicable to our first motivating study described in Section 2.2. In this application, we have three different treatment values, i.e., $J = 3$ and $T_i \in \mathcal{T} = \{0, 1, 2\}$. The generalized propensity score is given by the following probabilities that sum up to unity,

$$\pi^j(X_i) \;=\; \Pr(T_i = j \mid X_i) \tag{4}$$

where $j \in \mathcal{T}$ and $\sum_{j=0}^{J-1} \pi^j(X_i) = 1$. We assume, as before, $\pi^j(X_i) > 0$ for all $j \in \mathcal{T}$. A commonly used parametric model, which we utilize here, is the multinomial logistic regression,

$$\pi^j_\beta(X_i) \;=\; \frac{\exp\left(X_i^\top \beta_j\right)}{\exp\left(\sum_{j'=0}^{J-1} X_i^\top \beta_{j'}\right)} \tag{5}$$

where the normalization constraint $\beta_0 = 0$ is imposed. The maximum likelihood estimate of $\beta = \{\beta_1, \ldots, \beta_{J-1}\}$ then is given by,

$$\hat{\beta}_{\mathrm{ML}} \;=\; \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{j=0}^{J-1} \mathbf{1}\{T_i = j\} \log \pi_\beta^j(X_i) \tag{6}$$

where $\beta = (\beta_1, \ldots, \beta_{J-1})$.

In contrast, the CBPS methodology estimates the generalized propensity score such that the following covariate balancing conditions are satisfied,

$$\mathbb{E}\left(\frac{\mathbf{1}\{T_i = 0\}X_i}{\pi_\beta^0(X_i)}\right) \;=\; \mathbb{E}\left(\frac{\mathbf{1}\{T_i = 1\}X_i}{\pi_\beta^1(X_i)}\right) \;=\; \cdots \;=\; \mathbb{E}\left(\frac{\mathbf{1}\{T_i = J-1\}X_i}{\pi_\beta^{J-1}(X_i)}\right) \;=\; \mathbb{E}(X_i) \tag{7}$$

These conditions are based on the inverse-probability weighting where each observation is weighted by the generalized propensity score so that the covariate distribution becomes equal across treatment values. In our application, we have three distinct treatment values and hence we use the following orthogonalized contrasts,

$$\frac{1}{N}\sum_{i=1}^{N} w_\beta(T_i, X_i) \;=\; \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} 2\frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} - \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} \\ \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} \end{pmatrix} X_i \tag{8}$$

These orthogonalized conditions are linearly equivalent to pair-wise comparisons of the three groups, but make the computation more efficient.

We then obtain the optimal Generalized Method of Moments (GMM; Hansen, 1982) estimator of β by minimizing the following global measure of covariate imbalance,

$$\hat{\beta}_{\mathrm{CBPS}} \;=\; \underset{\beta}{\operatorname{argmin}} \; \bar{g}_\beta(T, X)^\top \; \Sigma_\beta(T, X)^{-1} \; \bar{g}_\beta(T, X) \tag{9}$$

where $\bar{g}_\beta(T, X) = \sum_{i=1}^{N} g_\beta(T_i, X_i)/N = \sum_{i=1}^{N} w_\beta(T_i, X_i)/N$ is the vector of sample moment conditions, and $\Sigma_\beta(T, X)^{-1}$ is the inverse of the covariance matrix $\Sigma_\beta(T, X)$. For the applications in Section 4, we use a two-step estimator, in which we estimate $\Sigma_\beta(T, X)^{-1}$ at the maximum likelihood values of β and fix it at that value for the

subsequent optimization. This makes our method far faster than continuously updating $\Sigma_\beta(T, X)$ throughout the optimization, and simulations (not presented here) suggests that the two-step estimator also achieves superior performance.

It is also possible to use the score condition from the log-likelihood function in equation (6) as a set of additional moment conditions. The score condition can be written as,

$$\frac{1}{N}\sum_{i=1}^{N} s_\beta(T_i, X_i) = \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} \left(\frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)}\right)\frac{\partial}{\partial\beta_1}\pi_\beta^1(X_i) + \left(\frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} - \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)}\right)\frac{\partial}{\partial\beta_1}\pi_\beta^2(X_i) \\ \left(\frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)}\right)\frac{\partial}{\partial\beta_2}\pi_\beta^1(X_i) + \left(\frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} - \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)}\right)\frac{\partial}{\partial\beta_2}\pi_\beta^2(X_i) \end{pmatrix} \tag{10}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} \mathbf{1}\{T_i = 1\} - \pi_\beta^1(X_i) \\ \mathbf{1}\{T_i = 2\} - \pi_\beta^2(X_i) \end{pmatrix} X_i \tag{11}$$

The expression in equation (10) provides an alternative interpretation of score conditions as covariate balancing conditions where the derivatives of generalized propensity score are balanced across three groups. These score conditions can be incorporated into the above GMM framework by setting

$$g_\beta(T_i, X_i) = \begin{pmatrix} s_\beta(T_i, X_i) \\ w_\beta(T_i, X_i) \end{pmatrix} \tag{12}$$

Finally, to obtain the optimal GMM estimator, we derive the covariance of moment conditions, i.e, $\Sigma_\beta(T, X)$. This is given by the following $(4K \times 4K)$ matrix if the score conditions are included,

$$\Sigma_\beta(T, X) = \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & -1 & 1 \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & -1 & -1 \\ -1 & -1 & \frac{4}{\pi_0} + \frac{1}{\pi_1} + \frac{1}{\pi_2} & -\frac{1}{\pi_1} + \frac{1}{\pi_2} \\ 1 & -1 & -\frac{1}{\pi_1} + \frac{1}{\pi_2} & \frac{1}{\pi_1} + \frac{1}{\pi_2} \end{pmatrix} X_i X_i^\top \tag{13}$$

where we write $\pi_j = \pi_\beta^j(X_i)$ for the sake of notational simplicity.

It is straightforward to extend this to the case of more than three treatment values. We consider here the treatment variable with four different values as an illustration. The score and (orthogonalized) covariate balancing conditions are given by,

$$w_\beta(T_i, X_i) = \begin{pmatrix} \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} + \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} - \frac{\mathbf{1}\{T_i=3\}}{\pi_\beta^3(X_i)} \\ \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} - \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} + \frac{\mathbf{1}\{T_i=3\}}{\pi_\beta^3(X_i)} \\ -\frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} + \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} + \frac{\mathbf{1}\{T_i=3\}}{\pi_\beta^3(X_i)} \end{pmatrix} X_i \tag{14}$$

$$s_\beta(T_i, X_i) = \begin{pmatrix} \mathbf{1}\{T_i = 1\} - \pi_\beta^1(X_i) \\ \mathbf{1}\{T_i = 2\} - \pi_\beta^2(X_i) \\ \mathbf{1}\{T_i = 3\} - \pi_\beta^3(X_i) \end{pmatrix} X_i \tag{15}$$

These sample moment conditions lead to the following covariance matrix,

$$\Sigma_\beta(T, X)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & -\pi_1\pi_3 & -1 & -1 & 1 \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & -\pi_2\pi_3 & -1 & -1 & -1 \\ -\pi_1\pi_3 & -\pi_2\pi_3 & \pi_3(1-\pi_3) & -1 & 1 & 1 \\ 1 & -1 & -1 & \frac{1}{\pi_0}+\frac{1}{\pi_1}+\frac{1}{\pi_2}+\frac{1}{\pi_3} & \frac{1}{\pi_0}-\frac{1}{\pi_1}+\frac{1}{\pi_2}-\frac{1}{\pi_3} & \frac{-1}{\pi_0}+\frac{1}{\pi_1}+\frac{1}{\pi_2}-\frac{1}{\pi_3} \\ -1 & -1 & 1 & \frac{1}{\pi_0}-\frac{1}{\pi_1}+\frac{1}{\pi_2}-\frac{1}{\pi_3} & \frac{1}{\pi_0}+\frac{1}{\pi_1}+\frac{1}{\pi_2}+\frac{1}{\pi_3} & \frac{-1}{\pi_0}-\frac{1}{\pi_1}+\frac{1}{\pi_2}+\frac{1}{\pi_3} \\ 1 & -1 & 1 & \frac{-1}{\pi_0}+\frac{1}{\pi_1}+\frac{1}{\pi_2}-\frac{1}{\pi_3} & \frac{-1}{\pi_0}-\frac{1}{\pi_1}+\frac{1}{\pi_2}+\frac{1}{\pi_3} & \frac{1}{\pi_0}+\frac{1}{\pi_1}+\frac{1}{\pi_2}+\frac{1}{\pi_3} \end{pmatrix} X_i X_i^\top$$

where we again write $\pi_j = \pi_\beta^j(X_i)$ for the sake of notational simplicity.

3.2 Continuous Treatment

Next, we consider the covariate balancing propensity score for a continuous treatment so that it can be applied to the application described in Section 2.3 (where we treat a Box-Cox transformation of the number of advertisements plus one as a continuous variable). To do this, we first center both the treatment variable and each covariate by subtracting their respective sample means, i.e., $T_i^* = T_i - \sum_{i=1}^N T_i/N$ and $X_i^* = X_i - \sum_{i=1}^N X_i/N$, such that $\mathbb{E}(T_i^*) = 0$ and $\mathbb{E}(X_i^*) = 0$ hold.

Under this setting, we balance covariates such that weighted correlation between these two centered variables is minimized. The weight is given by $f(T_i^*)/f(T_i^* \mid X_i^*)$ where the numerator is a required stabilizing factor (Robins *et al.*, 2000). Formally, the covariate balancing condition is given by the weighted cross moment between these centered variables,

$$\mathbb{E}\left(\frac{f(T_i^*)}{f(T_i^* \mid X_i^*)} T_i^* X_i^*\right) = \int \left\{ \int \frac{f(T_i^*)}{f(T_i^* \mid X_i^*)} T_i^* dF(T_i^* \mid X_i^*) \right\} X_i^* dF(X_i^*) \quad (16)$$

$$= \mathbb{E}(T_i^*)\mathbb{E}(X_i^*) = 0. \quad (17)$$

We follow a common practice of assuming a homoskedastic linear model possibly after transforming the treatment variable as done in our application (e.g., Robins *et al.*, 2000; Hirano and Imbens, 2004; Imai and van Dyk, 2004). Then, the generalized propensity score is given by the following conditional normal density,

$$f_\theta(T_i^* \mid X_i^*) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(T_i^* - X_i^{*\top}\beta)^2 \right\} \quad (18)$$

where $\theta = (\beta, \sigma^2)$. In addition, we follow a typical parametric modeling approach described by Robins *et al.* (2000) and assume the marginal distribution to be normal with mean zero (due to centering). The transformation of the treatment variable should be chosen such that this distributional assumption is reasonable. Then, the stabilizing weight is given by,

$$\frac{f_{\sigma^2}(T_i^*)}{f_\theta(T_i^* \mid X_i^*)} = \exp\left[\frac{1}{2\sigma^2} \left\{ -2T_i^* X_i^{*\top}\beta + (X_i^{*\top}\beta)^2 \right\} \right]. \quad (19)$$

Under the GMM framework introduced in Section 3.1, we then have the following sample moment conditions if both the score and covariate balancing conditions are included,

$$g_\theta(T_i, X_i) = \begin{pmatrix} s_\theta(T_i, X_i) \\ w_\theta(T_i, X_i) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}(T_i^* - X_i^{*\top}\beta) X_i^* \\ -\frac{1}{2\sigma^2}\left\{ 1 - \frac{1}{\sigma^2}(T_i^* - X_i^{*\top}\beta)^2 \right\} \\ \exp\left[\frac{1}{2\sigma^2}\left\{ -2X_i^{*\top}\beta + (X_i^{*\top}\beta)^2 \right\} \right] T_i^* X_i^* \end{pmatrix} \quad (20)$$

In the appendix, we derive the covariance matrix of these sample moment conditions and the result is given here,

$$\Sigma_\theta(T,X) \;=\; \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \frac{1}{\sigma^2}X_i^*X_i^{*\top} & \mathbf{0} & X_i^*X_i^{*\top} \\ \mathbf{0} & \frac{1}{2\sigma^4} & -\frac{X_i^{*\top}\beta}{\sigma^2}X_i^{*\top} \\ X_i^*X_i^{*\top} & -\frac{X_i^{*\top}\beta}{\sigma^2}X_i^* & \exp\left(\frac{\left(X_i^{*\top}\beta\right)^2}{\sigma^2}\right)\left\{\sigma^2+\left(X_i^{*\top}\beta\right)^2\right\}X_i^*X_i^{*\top} \end{pmatrix} \tag{21}$$

Finally, we find the value of the parameter vector θ that minimizes the GMM objective function given in equation (9).

4 Empirical Analyses

We now turn to the empirical analyses of the two motivating examples introduced in Section 2 and apply the proposed methodology. In both cases, we will show that weighting with the CBPS substantially reduces the imbalance and enables flexible estimation of causal effects for non-binary treatments.

4.1 The Effect of Education on Political Participation

We analyze the causal effect of educational attainment using a three-level treatment variable. The three treatment categories are (1) not attending college, (2) attending college without graduating (some college), and (3) graduating from college. We apply the proposed methodology described in Section 3.1. The generalized propensity score is estimated using the multinomial logit regression model with the covariates listed in Section 2.2. We compare the inverse-probability weighting based on the CBPS estimation of this model with that based on the maximum likelihood (ML) estimation. For all calculations, we use the over-identified CBPS based on both score and covariate balancing conditions.

Figure 2 displays the resulting pairwise covariate imbalance based on the ML and CBPS estimation methods using the absolute difference in standardized weighted

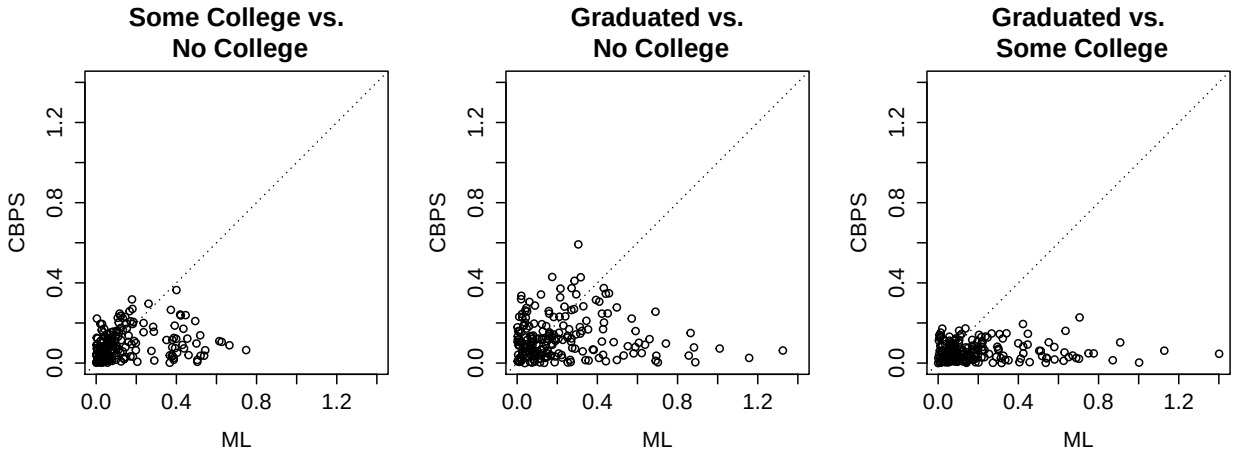

Figure 2: Pairwise Covariate Imbalance using Maximum Likelihood (ML) and Covariate Balancing Propensity Score (CBPS) Estimation Methods across Three Treatment Values. Each plot compares the absolute difference in standardized means between each pair of treatment values for the ML (x-axis) and the CBPS (y-axis). The CBPS reduces imbalance in all three comparisons.

means between each pair of treatment values, which is formally defined as follows,

$$\widetilde{\mathcal{I}}_k(t, t^*) \;=\; \sqrt{\frac{N-1}{\sum_{i=1}^N (X_{ik} - \overline{X}_k)^2}} \left| \frac{\sum_{i=1}^N \mathbf{1}\{T_i = t\} w_i X_{ik}}{\sum_{i=1}^N \mathbf{1}\{T_i = t\} w_i} - \frac{\sum_{i=1}^N \mathbf{1}\{T_i = t^*\} w_i X_{ik}}{\sum_{i=1}^N \mathbf{1}\{T_i = t^*\} w_i} \right| \tag{22}$$

where w_i is the weight and $\overline{X}_k = \sum_{i=1}^N X_{ik}/N$. This is a generalized version of the measure defined in equation (3). In each of the plots, points below (above) the 45° line indicate better balance for the CBPS (ML). The results show that the CBPS substantially reduces the pairwise imbalance between all pairs of treatment groups, and especially between those who did and did not graduate (right plot). Averaging these imbalance measures across covariates, reduces the imbalance from 0.143 to 0.089 for the no college versus some college contrast, from 0.221 to 0.126 for no college versus graduated, and from 0.200 to 0.058 for some college versus graduated.

The performance of matching and weighting methods tends to be poor when a small number of observations have extremely large weights (e.g., Kang and Schafer, 2007; Yang *et al.*, 2014). Henderson and Chatfield (2011) raises this concern in the context of the current application. Figure 3 compares the distribution of weights

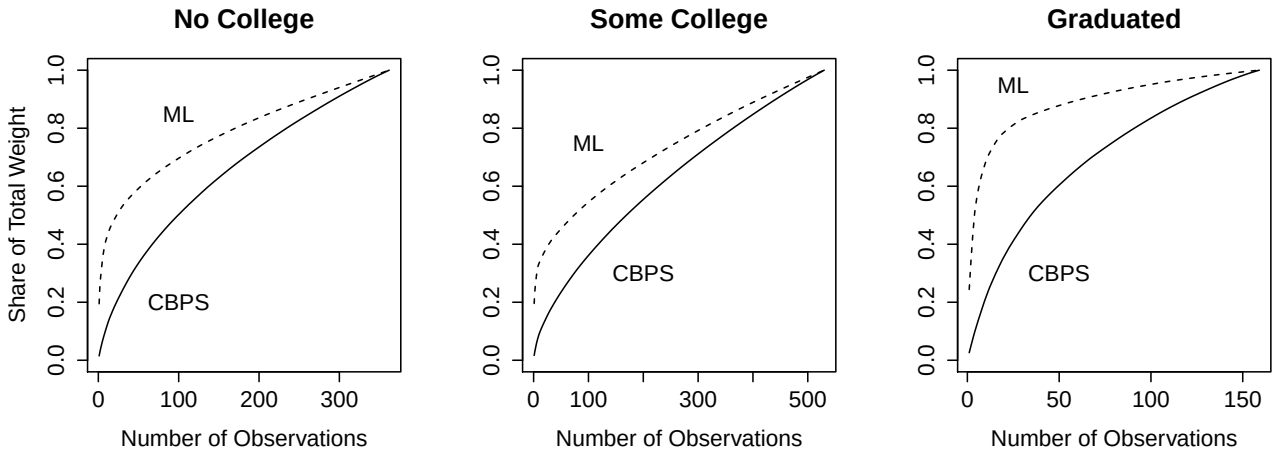

Figure 3: Distribution of Weights for the Covariate Balancing Propensity Score (CBPS) and Maximum Likelihood (ML) Estimation for Three Different Values of the Treatment. For each treatment value, we sort the observations by the magnitude of their weights and then plot the cumulative proportion of weights separately for ML (dashed lines) and CBPS (solid lines). For ML, a small proportion of the observations account for an outsized proportion of the weight. The weights for CBPS are more evenly distributed by comparison.

between the CBPS and the ML by first sorting the observations according to the size of their weights and then plotting the cumulative proportion of weights for each treatment value. The figure shows that for ML estimation, 5% of the observations account for about 40% of the total weight in each of the treatment values. The problem is most severe for the "graduated" treatment category. By comparison, the top 5% of observations for the CBPS do not account for more that 20% of the total weight in any treatment group. This suggests that a small number of outliers are unlikely to drive the results based on the CBPS.

Next, if the CBPS successfully balances covariates across all three treatment values, then it also should improve covariate balance with respect to the dichotomized treatment. We aggregate the estimated generalized propensity score into the estimated binary propensity score by summing the estimated probabilities for the "graduated college" and "some college" categories. We then weight each observation according to these aggregated binary propensity scores and compare the resulting imbalance with that of propensity score matching (Kam and Palmer, 2008) and genetic match-

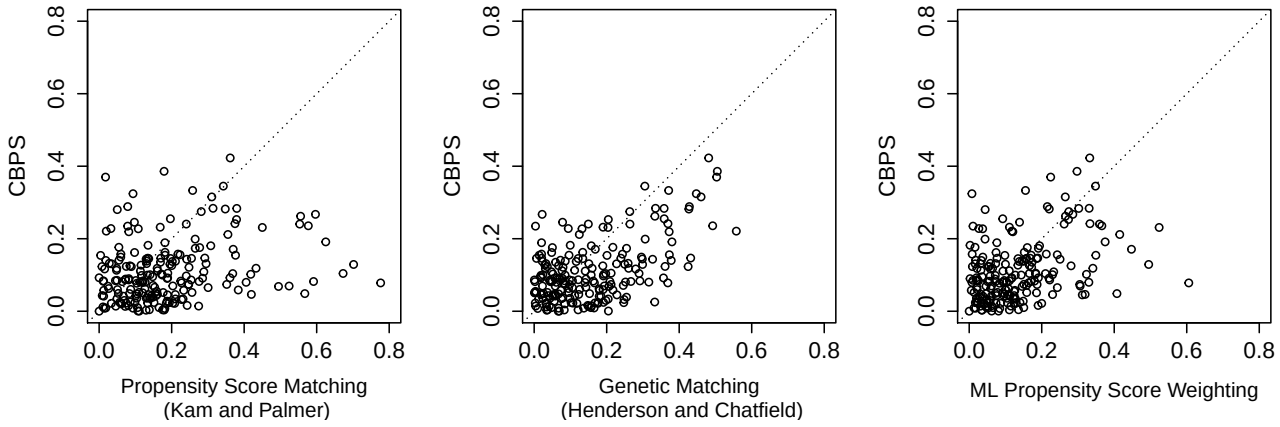

Figure 4: Comparison of Covariate Imbalance with Respect to the Dichotomized Treatment between Matching and the Multi-valued CBPS. The CBPS weights are estimated using the original treatment variable with three values and then are aggregated to the dichotomized treatment. Plots compare the absolute difference in standardized means with respect to the dichotomized treatment (college attendance vs. no college) for each covariate between the CBPS (y-axis) and various matching and weighting methods (x-axis). The propensity score and genetic matching are based on the same procedures used by Kam and Palmer (2008) and Henderson and Chatfield (2011) in their original analyses, respectively. The multi-valued CBPS weighting reduces imbalance relative to propensity score matching (left plot), genetic matching (middle plot), and the ML binary propensity score weighting (right plot).

ing (Henderson and Chatfield, 2011) as well as the binary propensity score weighting via the ML estimation of the (binary) logistic regression model. Note that when computing the covariate imbalance measure given in equation (22), for matching the weights are either one (for matched observations) or zero (for discarded observations).

Figure 4 presents the results. We observe that the aggregated multi-valued CBPS balances covariates significantly better than propensity score matching (left plot), genetic matching (middle plot), and binary propensity score weighting with ML estimation (right plot). Following Kam and Palmer, all of the weights used in this comparison estimate the average treatment effect for the treated (ATT) rather than the average treatment effect (ATE). The multi-valued CBPS weighting leaves an average covariate imbalance of 0.110, compared to 0.186 for propensity score matching, 0.161 for genetic matching, and 0.135 for propensity score weighting. While each

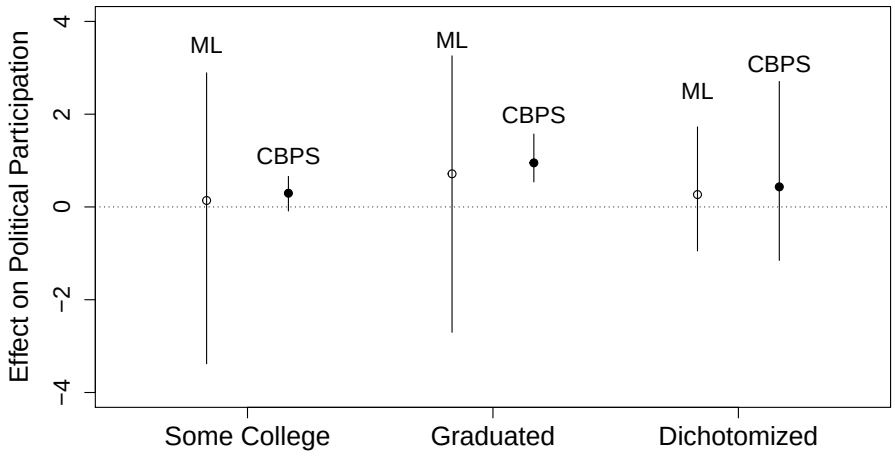

Figure 5: The Estimated Average Treatment Effects of Education on Political Participation Using the Weights based on the Covariate Balancing Propensity Score (CBPS) and the Maximum Likelihood (ML). Treatment effects are calculated with "No College" as the baseline category. 95% confidence intervals are based on 500 iterations of a non-parametric bootstrap.

method optimizes different measures of covariate balance, these results show that the multi-valued CBPS improves covariate balance well in this application.

Finally, we estimate the causal effects of education on political participation in two ways; we first use the original three-value treatment variable and then the dichotomized treatment variable. For both analyses, we use the weights based on the CBPS and ML estimation and compare the results between them. As explained in Section 2.2, the outcome variable is an index of political participation, ranging from 0 to 8. We fit the weighted linear model to estimate the average treatment effect for each treatment value where the baseline treatment category is "No college." For this linear model, we include the same set of 81 explanatory variables as the propensity score model. We report the estimates based on the model that adjusts for covariates because covariate imbalance remains even after weighting. The 95% confidence intervals are computed using 500 replications of non-parametric bootstrap, accounting for the uncertainty due to the estimation of weights as well as that of the average treatment effects.

Figure 5 compares the results of the multi-valued analysis to that of the di-

chotomized analysis. Here, we estimate the ATE rather than the ATT in the dichotomized analysis in order to make the results directly comparable to the multi-valued treatment effects. The multi-valued analysis based on the CBPS weights finds a small positive effect for attending college and a larger effect for graduating from college. However, the weights obtained from dichotomizing the CBPS fail to detect this effect (and we also do not detect an effect if we estimate the ATT, as the original authors did). The point estimate of the dichotomized effect is between some college and graduating from college, as expected, but the confidence interval for the dichotomized treatment effect is much larger. The analysis based on the ML estimation of weights fails to detect an effect for either the dichotomous or the multi-valued treatment.

These results clearly show that the analysis of a multi-valued treatment can uncover substantively interesting causal effects, which may not have been apparent in the analysis of dichotomized treatment. The CBPS improves the estimation of generalized propensity score when the treatment variable is non-binary.

4.2 The Effect of Advertising on Campaign Contributions

In their original analysis, Urban and Niebler (2014) conduct a propensity score matching analysis by dichotomizing a continuous treatment variable of advertisements (more or fewer than 1,000 advertisements over the course of campaign). As shown in Section 2.3, propensity score matching with the dichotomized treatment variable does not balance covariates well with respect to the original continuous treatment variable. Thus, we apply the proposed methodology to estimate the generalized propensity score. We then estimate the dose response function using the inverse probability weighting method.

We begin by modeling the treatment variable via the normal linear regression with the CBPS methodology described in Section 3.2. Recall that our method assumes

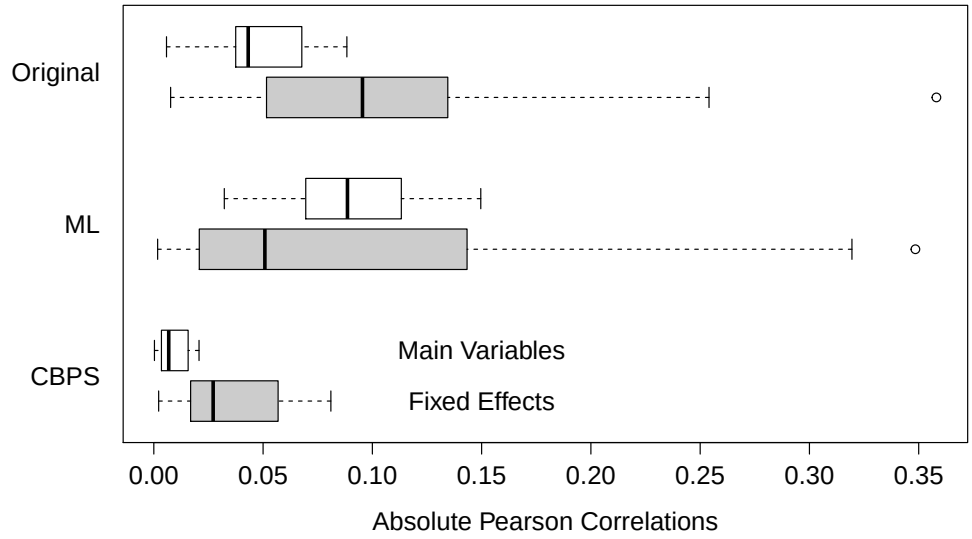

Figure 6: Comparing the Absolute Pairwise Correlations of the Original Continuous Treatment and Covariates when Using Covariate Balancing Propensity Score (CBPS) and Maximum Likelihood (ML) Estimation Methods. Along with the original unweighted correlations, correlations after weighting based on the CBPS and ML estimates of the generalized propensity score are shown. Solid circles represent the state fixed effects while open circles are other covariates. The CBPS methodology significantly improves covariate balance relative to ML estimation.

that the marginal distribution of the treatment is normal. To make this assumption credible, we take the Box-Cox transformation of the number of advertisements plus one whose quantiles are most closely correlated with the theoretical quantiles of a normal distribution. The best transformation achieves a correlation of 0.94, compared to 0.76 for the original treatment variable. In the propensity score model for this transformed treatment, we include the covariates listed in Section 2.3. We use the over-identified CBPS by incorporating score conditions as well as covariate balancing conditions. The substantive results based on the just-identified CBPS are similar to those presented here. For the sake of comparison, we also fit the model via the maximum likelihood (ML).

Figure 6 presents the absolute pairwise correlations between the continuous treatment covariates for each adjustment method along with those calculated using the unadjusted data set. The plot clearly shows that the CBPS greatly improves the covariate balance, compared to the weighting based on the ML estimation of the gen-

eralized propensity score. Quantifying imbalance as the mean of the absolute Pearson correlations between the treatment and covariates, the CBPS reduces the imbalance by 66.3% compared to the original sample whereas the ML *increases* the imbalance by 4.5%. A similar pattern holds if we focus on non-fixed effects covariates alone. The CBPS reduces their average imbalance by 93.3% while the ML only reduces imbalance by 18.4%.

Using the estimated generalized propensity score from the CBPS, we estimate the dose-response function. In the original analysis, Urban and Niebler (2014) fitted the linear regression with state and monthly fixed effects using a separate *unmatched* data set to estimate a dose response for ads on contributions, altogether separate from their dichotomized matching analysis. Given that there are a number of zip codes which report zero contribution, we fit two weighted generalized linear models where weights are based on the CBPS estimation on the same data set that Urban and Niebler (2014) used for their matching analysis. First, we fit the logistic regression to model whether there is any contribution given in each zip code. Second, we use linear regression to model the amount of contributions in each zip code given that some contributions are made in that zip code. For this second model, we use the log of contributions (incremented by one) as the outcome variable. Both models include the quadratic term of the treatment variable to address possible non-linearity. These two fitted models can then be combined to calculate the expected amount of contributions as a function of advertisements, yielding the estimated dose-response curve. Both models take the zip code's log of median household income plus one, log of population, log of advertisements plus one, the percent of the population over 65, the percentage of blacks, the percentage of Hispanics, the percent of high school graduates, the percent of college graduates, the population density, state fixed effects, and a dummy variable measuring whether people can commute from other states as covariates. We construct

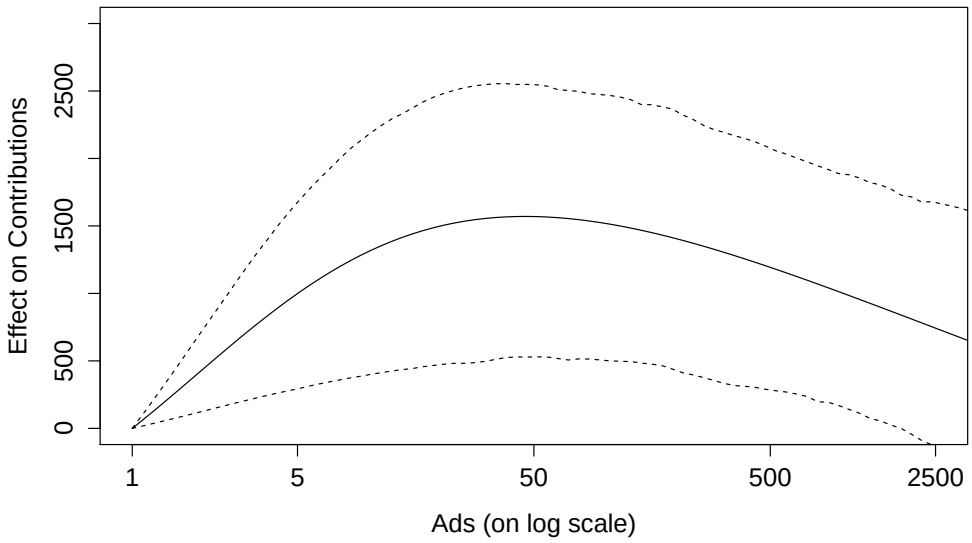

Figure 7: Estimated Dose-Response Curve between Dollar Amount of Contributions and the Number of Advertisements. The dotted lines indicate bootstrapped 95% confidence intervals, which incorporate uncertainty about the weights.

a 95% confidence interval around this estimated dose-response curve by taking 1000 non-parametric bootstraps. As in the previous example, we incorporate uncertainty about both the treatment effect and the weights.

Figure 7 suggests that a small number of advertisements leads to an increase in campaign contributions. The peak of this effect is at about 44 advertisements, and the 95% confidence interval for the effect at this level is approximately ($870, $4721). The 95% confidence interval for the effect of 1000 advertisements (Urban and Niebler's threshold) is ($767, $3543). The magnitude of this effect is considerably smaller than the effect found in the original analysis, but it is still statistically significant by conventional standards and in the same direction. Moreover, the effect is still sufficiently large such that it may be profitable to advertise in non-competitive media markets. Although we fit a quadratic regression model for the dose-response curve, further analysis could employ an even more flexible outcome model.

Our contribution to Urban and Niebler's analysis is to offer a simple and effective framework in which the propensity score methodology is directly linked to the quantity of interest: the dose response rather than an estimate of a dichotomized

treatment effect. Most importantly, we find in this context that the CBPS methodology for estimating the generalized propensity score yields superior covariate balance relative to the dominant method of maximum likelihood estimation.

5 Concluding Remarks

Despite some advances in generalizing propensity score methods to non-binary treatments, applied researchers have been using propensity score methods mostly to analyze binary treatments. Even when the original treatment is non-binary, they often dichotomize the treatment variable in order to utilize propensity score methods. One reason for this gap between statistical theory and practice is the absence of a reliable method for estimating the generalized propensity score. In this paper, we extend the covariate balancing propensity score (CBPS) of Imai and Ratkovic (2014) to general treatment regimes. We estimate the generalized propensity score such that the resulting covariate balance is optimized. We demonstrate this idea by applying the CBPS methodology to multi-valued and continuous treatments. Our empirical analyses show that the proposed methodology results in better covariate balance than the standard method and can yield substantive insights which may be difficult to obtain by analyzing the dichotomous treatment. We also find that the CBPS reduces the sensitivity to misspecification of the propensity score model for a general treatment regime.

Finally, while in this paper we focused on improving the parametric estimation of propensity score, future research should develop nonparametric estimation of propensity score for general treatment regimes. Such an extension will further improve the robustness of CBPS methodology and as a result produce more credible causal inference in observational studies.

Appendix: Derivation of $\Sigma_\theta(T, X)$ for a Continuous Treatment

Although the derivation of $\Sigma_\beta(T, X)$ for three and four valued treatments is quite straightforward, the derivation of $\Sigma_\theta(T, X)$ for continuous treatments is somewhat more involved. The log-likelihood is simply the log-likelihood of the normal.

$$\ell(\beta \mid X, T) \;=\; -\frac{1}{2}\sum_{i=1}^{n}\left\{\log(2\pi\sigma^2) + \frac{1}{\sigma^2}(T_i - X_i^{*\top}\beta)^2\right\} \tag{23}$$

This implies the following well-known score conditions,

$$s_\beta(T_i^*, X_i^*) = \frac{1}{\sigma^2}(T_i^* - X_i^{*\top}\beta)X_i^* \tag{24}$$

$$s_{\sigma^2}(T_i^*, X_i^*) = \frac{(T_i - X_i^*\beta)^2}{2\sigma^4} - \frac{1}{2\sigma^2} \tag{25}$$

The covariance of these score conditions are well-known, and so we derive the remaining elements of the covariance matrix $\Sigma_\theta(T, X)$ here,

$$\begin{aligned}
& \mathbb{E}[s_\beta(T_i, X_i)w_\theta(T_i, X_i)^\top \mid X_i] \\
= \;& \frac{1}{\sigma^2}\left\{\int_{-\infty}^{\infty}(T_i^{*2} - T_i^* X_i^{*\top}\beta)\frac{f(T_i^*)}{f(T_i^* \mid X_i^*)}f(T_i^* \mid X_i^*)dT_i^*\right\}X_i^* X_i^{*\top} \\
= \;& \frac{1}{\sigma^2}\left\{\mathbb{E}(T_i^{*2}) - \mathbb{E}(T_i^*)X_i^{*\top}\beta\right\}X_i^* X_i^{*\top} \\
= \;& X_i^* X_i^{*\top}
\end{aligned} \tag{26}$$

$$\begin{aligned}
& \mathbb{E}[s_{\sigma^2}(T_i, X_i)w_\theta(T_i, X_i)^\top \mid X_i] \\
= \;& \left[\frac{1}{2\sigma^4}\int_{-\infty}^{\infty}\{T_i^{*3} - 2T_i^{*2}X_i^{*\top}\beta + T_i^*(X_i^{*\top}\beta)^2 - T_i^*\sigma^2\}\frac{f(T_i^*)}{f(T_i^* \mid X_i^*)}f(T_i^* \mid X_i^*)dT_i^*\right]X_i^* \\
= \;& \frac{1}{2\sigma^4}\left[\mathbb{E}(T_i^{*3}) - 2\mathbb{E}(T_i^{*2})X_i^{*\top}\beta + \mathbb{E}(T_i^*)\left\{(X_i^{*\top}\beta)^2 - \sigma^2\right\}\right]X_i^* \\
= \;& -\frac{X_i^{*\top}\beta}{\sigma^2}X_i^*
\end{aligned} \tag{27}$$

$$\begin{aligned}
& \mathbb{E}[w_\theta(T_i, X_i)w_\theta(T_i, X_i)^\top \mid X_i] \\
= \;& \left(\int_{-\infty}^{\infty}\frac{T_i^{*2}f(T_i^*)^2}{f(T_i^* \mid X_i^*)^2}f(T_i^* \mid X_i^*)dT_i^*\right)X_i^* X_i^{*\top} \\
= \;& \left[\int_{-\infty}^{\infty}\frac{T_i^{*2}}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{1}{2\sigma^2}\left\{(T_i^* - X_i^{*\top}\beta)^2 - 2T_i^{*2}\right\}\right)dT_i^*\right]X_i^* X_i^{*\top} \\
= \;& \left[\int_{-\infty}^{\infty}\frac{T_i^{*2}}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left\{(T_i^* + X_i^{*\top}\beta)^2 - 2(X_i^{*\top}\beta)^2\right\}\right\}dT_i^*\right]X_i^* X_i^{*\top} \\
= \;& \left\{\sigma^2 + (X_i^{*\top}\beta)^2\right\}\exp\left\{\frac{(X_i^{*\top}\beta)^2}{\sigma^2}\right\}X_i^* X_i^{*\top}
\end{aligned} \tag{28}$$

References

Berelson, B. R., Lazarsfeld, P. F., , and McPhee, W. N. (1954). *Voting: a study of opinion formation in presidential elections*. University of Chicago Press, Chicago.

Boyd, C. L., Epstein, L., and Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science* **54**, 2, 389–411.

De, P. K. and Ratha, D. (2012). Impact of remittances on household income, asset, and human capital: evidence from sri lanka. *Migration and Development* **1**, 1, 163–179.

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

Donohue III, J. J. and Ho, D. E. (2007). The impact of damage caps on malpractice claims: randomization inferences with difference-in-differences. *Journal of Empirical Legal Studies* **4**, 1, 69–102.

Fong, C., Ratkovic, M., and Imai, K. (2014). CBPS: R package for covariate balancing propensity score. available at the Comprehensive R Archive Network (CRAN). `http://CRAN.R-project.org/package=CBPS`.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 4, 1029–1054.

Harder, V. S., Stuart, E. A., and Anthony, J. C. (2008). Adolescent cannabis problems and young adult depression: male-female stratified propensity score analyses. *American Journal of Epidemiology* **168**, 6, 592–601.

Henderson, J. and Chatfield, S. (2011). Who matches? propensity scores and bias in the causal effects of education on participation. *Journal of Politics* **73**, 03, 646–658.

Hernán, M. A. and VanderWeele, T. J. (2011). Compound treatments and transportability of causal inference. *Epidemiology* **22**, 3, 368–377.

Hirano, K., Imbens, G., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**, 4, 1307–1338.

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, chap. 7. Wiley.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* **15**, 3, 199–236.

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

Imai, K. and van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* **99**, 467, 854–866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 3, 706–710.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* **86**, 1, 4–29.

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

Kam, C. D. and Palmer, C. L. (2008). Reconsidering the effects of education on political participation. *Journal of Politics* **70**, 03, 612–631.

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

Lu, B., Zanutto, E., Hornik, R., and Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association* **96**, 456, 1245–1253.

Mayer, A. K. (2011). Does education increase political participation? *Journal of Politics* **73**, 03, 633–645.

Nielsen, R. A., Findley, M. G., Davis, Z. S., Candland, T., and Nielson, D. L. (2011). Foreign aid shocks as a cause of violent armed conflict? *American Journal of POolitical Science* **55**, 2, 219–232.

Rassen, J., Shelat, A., Franklin, J., Glynn, R., Solomon, D., and Schneeweiss, S. (2013). Matching by propensity score in cohort studies with three treatment groups. *Epidemiology* **24**, 3, 401–409.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 5, 550–560.

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

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* **79**, 387, 516–524.

Rosenbaum, P. R. and 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. (1990). Comments on "On the application of probability theory to agricultural experiments. Essay on principles. Section 9" by J. Splawa-Neyman translated from the Polish and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* **5**, 472–480.

Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* **125**, 1–2, 305–353.

Urban, C. and Niebler, S. (2014). Dollars on the sidewalk: Should US presidential candidates advertise in uncontested states? *American Journal of Political Science* **58**, 2, 322–336.

VanderWeele, T. J. and Hernán, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference* **1**, 1, 1–20.

Verba, S., Schlozman, K. L., and Brady, H. (1995). *Voice and Equality: Civic Volunteerism in American Politics*. Harvard University Press.

Wyss, R., Ellis, A. R., Brookhart, M. A., Girman, C. J., Funk, M. J., 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* Forthcoming.

Yang, S., Imbens, G. W., Zhanglin, C., Faries, D., and Kadziola, Z. (2014). Propensity score matching and subclassification with multi-level treatments. *Biometrika* 1–16.