Covariate Balancing Propensity Score for General Treatment Regimes

Kosuke Imai

Princeton University

2014 Joint Statistical Meetings

August 7, 2014

Joint work with Christian Fong

Motivation

- Central role of propensity score in causal inference
  - Adjusting for observed confounding in observational studies
  - Matching and inverse-probability weighting methods
- Extensions of propensity score to general treatment regimes
  - Weighting (e.g., Imbens, 2000; Robins et al., 2000)
  - Subclassification (e.g., Imai & van Dyk, 2004)
  - Regression (e.g., Hirano & Imbens, 2004)
- But, propensity score is mostly applied to binary treatment
  - All available methods assume correctly estimated propensity score
  - No reliable methods to estimate generalized propensity score
  - Harder to check balance across a non-binary treatment
  - Many researchers dichotomize the treatment

Contributions of the Paper

- Generalize the covariate balancing propensity score (CBPS; Imai & Ratkovic, 2014, *JRSSB*)
- Key idea: estimate the generalized propensity score such that the association between treatment and covariates is reduced
    1. Multi-valued treatment (3 and 4 categories)
    2. Continuous treatment
- Useful especially because checking covariate balance is harder for non-binary treatment
- Facilitates the use of generalized propensity score methods

The Setup

- $T_i \in \mathcal{T}$: non-binary treatment
- X_i: pre-treatment covariates
- $Y_i(t)$: potential outcomes
- Strong ignorability:

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

- $p(T_i \mid X_i)$: generalized propensity score

- $\widetilde{T}_i$: dichotomized treatment
  - $\widetilde{T}_i = 1$ if $T_i \in \mathcal{T}_1$
  - $\widetilde{T}_i = 0$ if $T_i \in \mathcal{T}_0$
  - $\mathcal{T}_0 \bigcap \mathcal{T}_1 = \emptyset$ and $\mathcal{T}_0 \bigcup \mathcal{T}_1 = \mathcal{T}$
- What is the problem of dichotomizing a non-binary treatment?

The Problems of Dichotomization

- Under strong ignorability,

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

- Aggregation via $p(T_i \mid \widetilde{T}_i, X_i)$
  1. some substantive insights get lost
  2. external validity issue

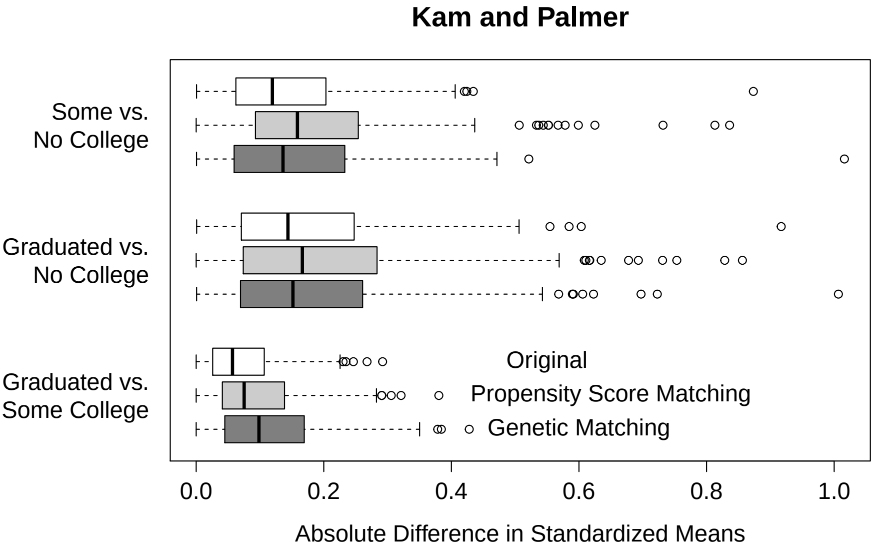
- Checking covariate balance: $\widetilde{T}_i \perp\!\!\!\perp X_i$ does not imply $T_i \perp\!\!\!\perp X_i$

Two Motivating Examples

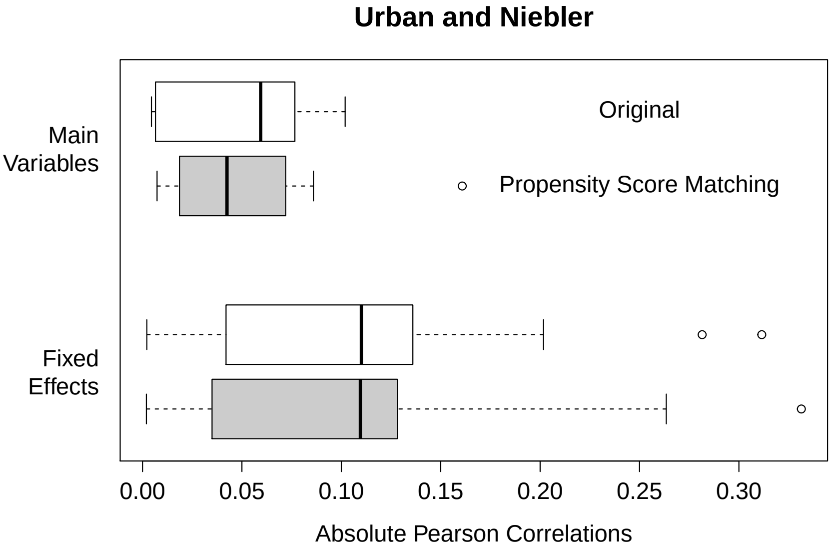
1. Effect of education on political participation
   - Education is assumed to play a key role in political participation
   - T_i: 3 education levels (graduated from college, attended college but not graduated, no college)
   - Original analysis $\rightsquigarrow$ dichotomization (some college vs. no college)
   - Propensity score matching
   - Critics employ different matching methods

2. Effect of advertisements on campaign contributions
   - Do TV advertisements increase campaign contributions?
   - T_i: Number of advertisements aired in each zip code
   - ranges from 0 to 22,379 advertisements
   - Original analysis $\rightsquigarrow$ dichotomization (over 1000 vs. less than 1000)
   - Propensity score matching followed by linear regression with an original treatment variable

Balancing Covariates for a Dichotomized Treatment



May Not Balance Covariates for the Original Treatment



Propensity Score for a Multi-valued Treatment

- Consider a multi-valued treatment: $\mathcal{T} = \{0, 1, \ldots, J-1\}$
- Standard approach: MLE with multinomial logistic regression

$$\pi^j(X_i) \;=\; \Pr(T_i = j \mid X_i) \;=\; \frac{\exp\left(X_i^\top \beta_j\right)}{1 + \exp\left(\sum_{j'=1}^{J} X_i^\top \beta_{j'}\right)}$$

where $\beta_0 = 0$ and $\sum_{j=0}^{J-1} \pi^j(X_i) = 1$

- Covariate balancing conditions with inverse-probability weighting:

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

which equals $\mathbb{E}(X_i)$

- Idea: estimate $\pi^j(X_i)$ to optimize the balancing conditions

CBPS for a Multi-valued Treatment

- Consider a 3 treatment value case as in our motivating example
- Sample balance conditions with orthogonalized contrasts:

$$\bar{g}_\beta(T, X) = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} 2\frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} - \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} \\ \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} \end{pmatrix} X_i$$

- Generalized method of moments (GMM) estimation:

$$\hat{\beta}_{\text{CBPS}} = \underset{\beta}{\text{argmin}}\ \bar{g}_\beta(T, X)\, \Sigma_\beta(T, X)^{-1}\, \bar{g}_\beta(T, X)$$

where $\Sigma_\beta(T, X)$ is the covariance of sample moments

Score Conditions as Covariate Balancing Conditions

- Balancing the first derivative across treatment values:

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

- Can be added to CBPS as over-identifying restrictions

Extension to More Treatment Values

- The same idea extends to a treatment with more values
- For example, consider a four-category treatment
- Sample moment conditions based on orthogonalized contrasts:

$$\bar{g}_\beta(T_i, X_i) = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} + \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} - \frac{\mathbf{1}\{T_i=3\}}{\pi_\beta^3(X_i)} \\ \frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} - \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} + \frac{\mathbf{1}\{T_i=3\}}{\pi_\beta^3(X_i)} \\ -\frac{\mathbf{1}\{T_i=0\}}{\pi_\beta^0(X_i)} + \frac{\mathbf{1}\{T_i=1\}}{\pi_\beta^1(X_i)} - \frac{\mathbf{1}\{T_i=2\}}{\pi_\beta^2(X_i)} + \frac{\mathbf{1}\{T_i=3\}}{\pi_\beta^3(X_i)} \end{pmatrix} X_i$$

- A similar orthogonalization strategy can be applied to marginal structural models (Imai & Ratkovic, 2014)

Propensity Score for a Continuous Treatment

- The stabilized weights:
$$\frac{f(T_i)}{f(T_i \mid X_i)}$$

- Covariate balancing condition:
$$\begin{aligned}\mathbb{E}\left(\frac{f(T_i^*)}{f(T_i^* \mid X_i^*)} T_i^* X_i^*\right) &= \int \left\{\int \frac{f(T_i^*)}{f(T_i^* \mid X_i^*)} T_i^* dF(T_i^* \mid X_i^*)\right\} X_i^* dF(X_i^*) \\ &= \mathbb{E}(T_i^*)\mathbb{E}(X_i^*) = 0.\end{aligned}$$

  where T_i^* and X_i^* are centered versions of T_i and X_i

- Again, estimate the generalized propensity score such that covariate balance is optimized

CBPS for a Continuous Treatment

- Standard approach (e.g., Robins *et al.* 2000):

$$T_i^* \mid X_i^* \overset{\text{indep.}}{\sim} \mathcal{N}(X_i^\top \beta,\ \sigma^2)$$
$$T_i^* \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\ \sigma^2)$$

  where further transformation of T_i can make these distributional assumptions more credible

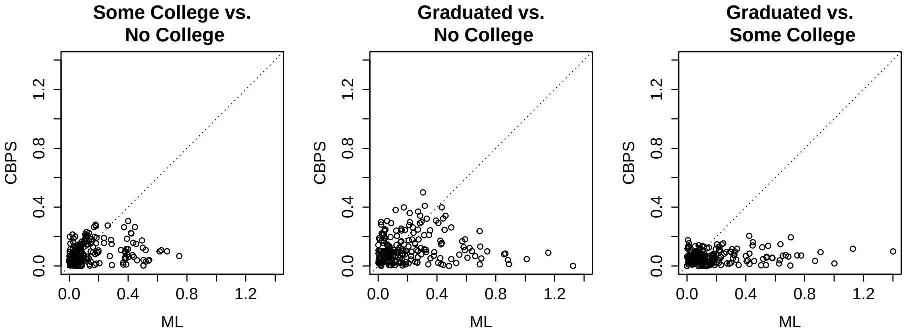
- Sample covariate balancing conditions:

$$\bar{g}_\theta(T, X) = \begin{pmatrix} \bar{s}_\theta(T, X) \\ \bar{w}_\theta(T, X) \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \frac{1}{\sigma^2}(T_i^* - X_i^{*\top}\beta) X_i^* \\ -\frac{1}{2\sigma^2}\left\{1 - \frac{1}{\sigma^2}(T_i^* - X_i^{*\top}\beta)^2\right\} \\ \exp\left[\frac{1}{2\sigma^2}\left\{-2X_i^{*\top}\beta + (X_i^{*\top}\beta)^2\right\}\right] T_i^* X_i^* \end{pmatrix}$$

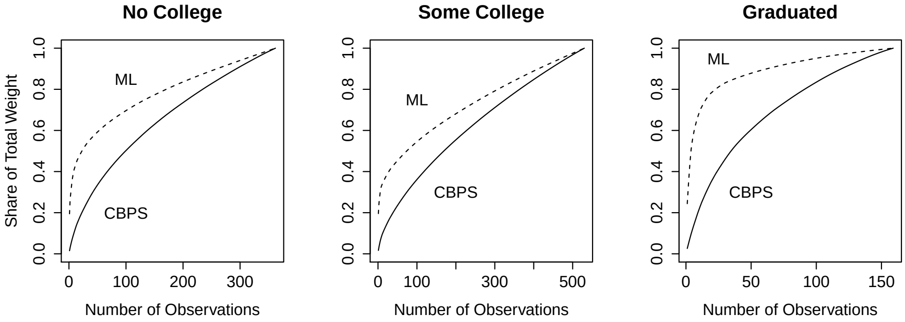
- GMM estimation: covariance matrix can be analytically calculated

Back to the Education Example: CBPS vs. ML

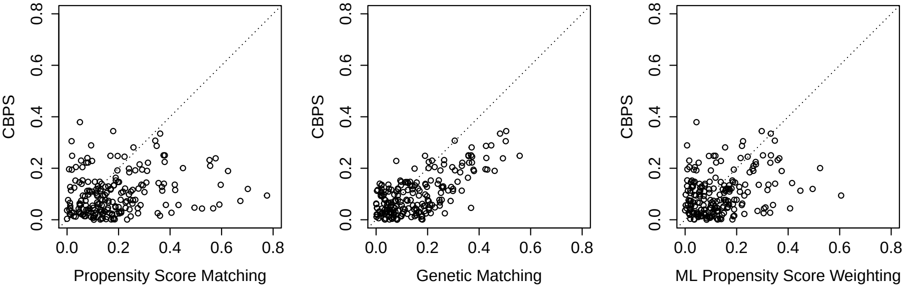
- CBPS achieves better covariate balance



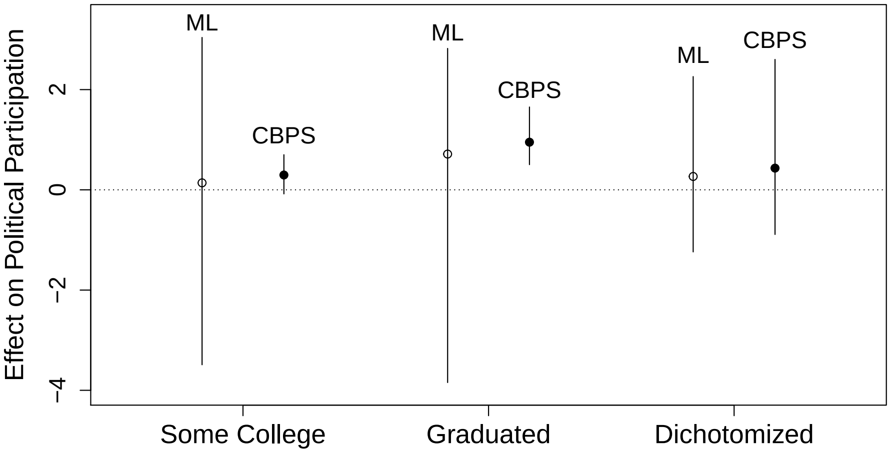
CBPS Avoids Extremely Large Weights



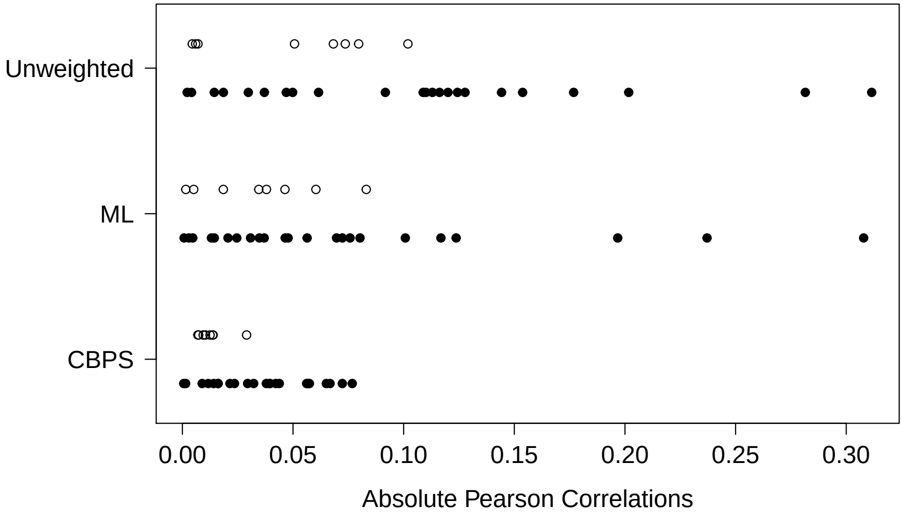
CBPS Balances Well for a Dichotomized Treatment



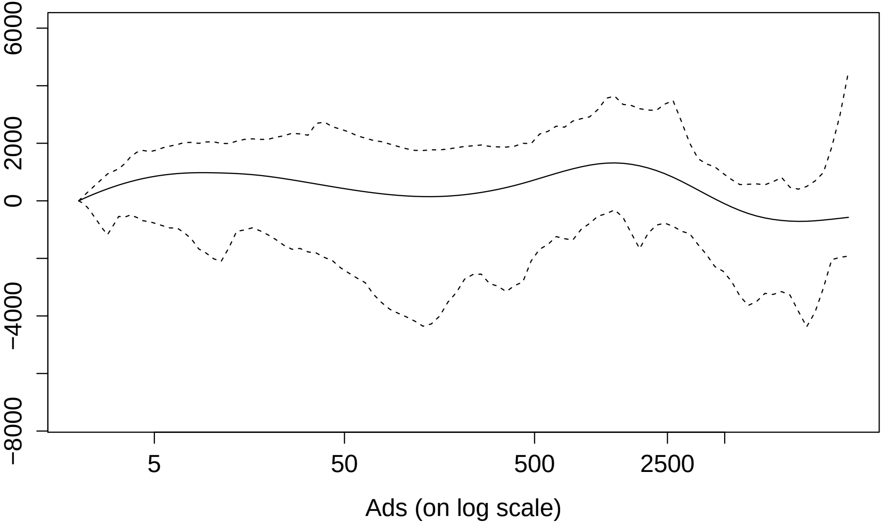
Empirical Results: Graduation Matters, Efficiency Gain



Onto the Advertisement Example



Empirical Finding: Little Effect of Advertisement



Concluding Remarks

- Numerous advances in generalizing propensity score methods to non-binary treatments
- Yet, many applied researchers don't use these methods and dichotomize non-binary treatments

- We offer a simple method to improve the estimation of propensity score for general treatment regimes
- Open-source R package: CBPS: Covariate Balancing Propensity Score available at CRAN

- Future extensions: nonparametric estimation, spatial treatments