Covariate Balancing Propensity Score for General Treatment Regimes

Kosuke Imai

Princeton University

October 14, 2014

Talk at the Department of Psychiatry, Columbia University

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 existing 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

- Results are often sensitive to misspecification of propensity score
- Solution: Estimate the generalized propensity score such that covariates are balanced
- Generalize the covariate balancing propensity score (CBPS; Imai & Ratkovic, 2014, *JRSSB*)
  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

Propensity Score for a Binary Treatment

- Notation:
  - $T_i \in \{0, 1\}$: binary treatment
  - X_i: pre-treatment covariates

- Dual characteristics of propensity score:
  1. Predicts treatment assignment:

$$\pi(X_i) \ = \ \Pr(T_i = 1 \mid X_i)$$

  2. Balances covariates (Rosenbaum and Rubin, 1983):

$$T_i \perp\!\!\!\perp X_i \ \mid \ \pi(X_i)$$

- Use of propensity score
  - Strong ignorability: $Y_i(t) \perp\!\!\!\perp T_i \mid X_i$ and $0 < \Pr(T_i = 1 \mid X_i) < 1$
  - Propensity score matching: $Y_i(t) \perp\!\!\!\perp T_i \mid \pi(X_i)$
  - Propensity score (inverse probability) weighting

Propensity Score Tautology

- Propensity score is unknown and must be estimated
  - Dimension reduction is purely theoretical: must model T_i given X_i
  - Diagnostics: covariate balance checking
- In theory: ellipsoidal covariate distributions
  $\Longrightarrow$ equal percent bias reduction
- In practice: skewed covariates and adhoc specification searches
- Propensity score methods are sensitive to model misspecification
- Propensity score tautology (Ho *et al*. 2007 *Political Analysis*):

  > *it works when it works, and when it does not work, it does not work (and when it does not work, keep working at it).*

Kang and Schafer (2007, *Statistical Science*)

- Simulation study: the deteriorating performance of propensity score weighting methods when the model is misspecified

- 4 covariates X_i^*: all are *i.i.d.* standard normal
- Outcome model: linear model
- Propensity score model: logistic model with linear predictors
- Misspecification induced by measurement error:
  - $X_{i1} = \exp(X_{i1}^*/2)$
  - $X_{i2} = X_{i2}^*/(1 + \exp(X_{1i}^*) + 10)$
  - $X_{i3} = (X_{i1}^* X_{i3}^*/25 + 0.6)^3$
  - $X_{i4} = (X_{i1}^* + X_{i4}^* + 20)^2$

Weighting Estimators Evaluated

1. Horvitz-Thompson (HT):

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

2. Inverse-probability weighting with normalized weights (IPW): HT with normalized weights (Hirano, Imbens, and Ridder)
3. Weighted least squares regression (WLS): linear regression with HT weights
4. Doubly-robust least squares regression (DR): consistently estimates the ATE if *either* the outcome or propensity score model is correct (Robins, Rotnitzky, and Zhao)

Weighting Estimators Do Fine If the Model is Correct

		Bias		RMSE	
Sample size	Estimator	GLM	True	GLM	True
(1) Both models correct					
$n = 200$	HT	0.33	1.19	12.61	23.93
	IPW	−0.13	−0.13	3.98	5.03
	WLS	−0.04	−0.04	2.58	2.58
	DR	−0.04	−0.04	2.58	2.58
$n = 1000$	HT	0.01	−0.18	4.92	10.47
	IPW	0.01	−0.05	1.75	2.22
	WLS	0.01	0.01	1.14	1.14
	DR	0.01	0.01	1.14	1.14
(2) Propensity score model correct					
$n = 200$	HT	−0.05	−0.14	14.39	24.28
	IPW	−0.13	−0.18	4.08	4.97
	WLS	0.04	0.04	2.51	2.51
	DR	0.04	0.04	2.51	2.51
$n = 1000$	HT	−0.02	0.29	4.85	10.62
	IPW	0.02	−0.03	1.75	2.27
	WLS	0.04	0.04	1.14	1.14
	DR	0.04	0.04	1.14	1.14

Weighting Estimators are Sensitive to Misspecification

Sample size	Estimator	Bias GLM	Bias True	RMSE GLM	RMSE True
(3) Outcome model correct					
$n = 200$	HT	24.25	−0.18	194.58	23.24
	IPW	1.70	−0.26	9.75	4.93
	WLS	−2.29	0.41	4.03	3.31
	DR	−0.08	−0.10	2.67	2.58
$n = 1000$	HT	41.14	−0.23	238.14	10.42
	IPW	4.93	−0.02	11.44	2.21
	WLS	−2.94	0.20	3.29	1.47
	DR	0.02	0.01	1.89	1.13
(4) Both models incorrect					
$n = 200$	HT	30.32	−0.38	266.30	23.86
	IPW	1.93	−0.09	10.50	5.08
	WLS	−2.13	0.55	3.87	3.29
	DR	−7.46	0.37	50.30	3.74
$n = 1000$	HT	101.47	0.01	2371.18	10.53
	IPW	5.16	0.02	12.71	2.25
	WLS	−2.95	0.37	3.30	1.47
	DR	−48.66	0.08	1370.91	1.81

Covariate Balancing Propensity Score (CBPS)

- Idea: Estimate propensity score such that covariates are balanced
- Goal: Robust estimation of parametric propensity score model

- Covariate balancing conditions:

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

- Over-identification via score conditions:

$$\mathbb{E}\left\{\frac{T_i \pi'_\beta(X_i)}{\pi_\beta(X_i)} - \frac{(1-T_i)\pi'_\beta(X_i)}{1-\pi_\beta(X_i)}\right\} = 0$$

- Can be interpreted as another covariate balancing condition

- Combine them with the Generalized Method of Moments or Empirical Likelihood

CBPS Makes Weighting Methods Work Better

	Estimator	Bias GLM	Bias CBPS1	Bias CBPS2	Bias True	RMSE GLM	RMSE CBPS1	RMSE CBPS2	RMSE True
(3) Outcome model correct									
$n = 200$	HT	24.25	1.09	−5.42	−0.18	194.58	5.04	10.71	23.24
	IPW	1.70	−1.37	−2.84	−0.26	9.75	3.42	4.74	4.93
	WLS	−2.29	−2.37	−2.19	0.41	4.03	4.06	3.96	3.31
	DR	−0.08	−0.10	−0.10	−0.10	2.67	2.58	2.58	2.58
$n = 1000$	HT	41.14	−2.02	2.08	−0.23	238.14	2.97	6.65	10.42
	IPW	4.93	−1.39	−0.82	−0.02	11.44	2.01	2.26	2.21
	WLS	−2.94	−2.99	−2.95	0.20	3.29	3.37	3.33	1.47
	DR	0.02	0.01	0.01	0.01	1.89	1.13	1.13	1.13
(4) Both models incorrect									
$n = 200$	HT	30.32	1.27	−5.31	−0.38	266.30	5.20	10.62	23.86
	IPW	1.93	−1.26	−2.77	−0.09	10.50	3.37	4.67	5.08
	WLS	−2.13	−2.20	−2.04	0.55	3.87	3.91	3.81	3.29
	DR	−7.46	−2.59	−2.13	0.37	50.30	4.27	3.99	3.74
$n = 1000$	HT	101.47	−2.05	1.90	0.01	2371.18	3.02	6.75	10.53
	IPW	5.16	−1.44	−0.92	0.02	12.71	2.06	2.39	2.25
	WLS	−2.95	−3.01	−2.98	0.19	3.30	3.40	3.36	1.47
	DR	−48.66	−3.59	−3.79	0.08	1370.91	4.02	4.25	1.81

The Setup for a General Treatment Regime

- $T_i \in \mathcal{T}$: 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

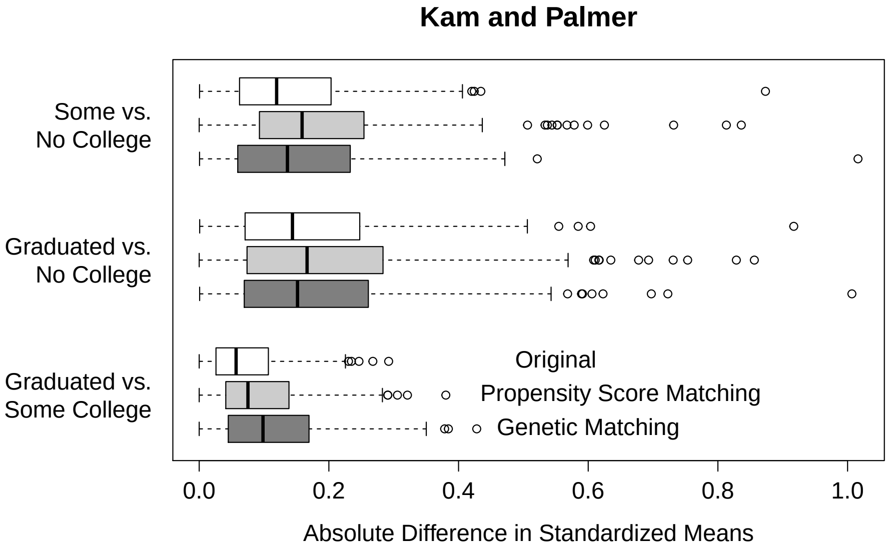
- Checking covariate balance: $\widetilde{T}_i \perp\!\!\!\perp X_i$ does not imply $T_i \perp\!\!\!\perp X_i$

Two Motivating Examples

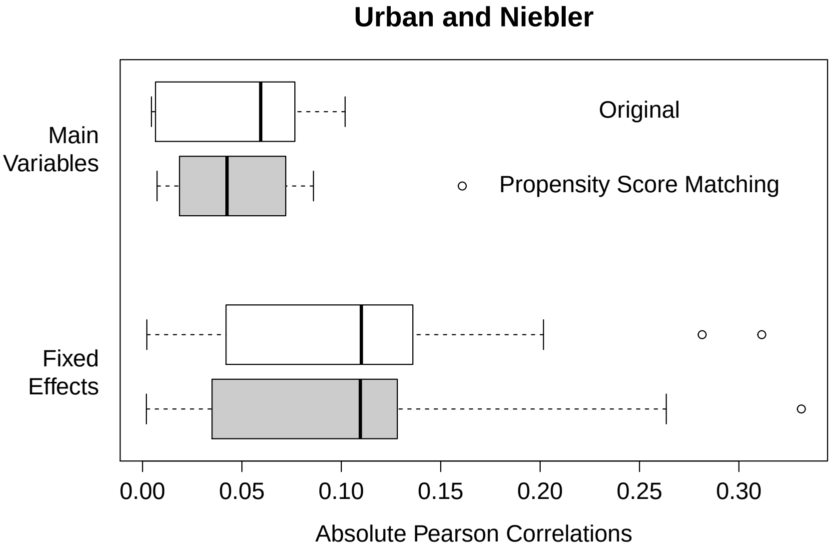
1. Effect of education on political participation
   - Education is assumed to play a key role in political participation
   - T_i: 3 education levels (graduated from college, attended college but not graduated, no college)
   - Original analysis $\rightsquigarrow$ dichotomization (some college vs. no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

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

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 the longitudinal setting with marginal structural models (Imai & Ratkovic, *JASA*, in-press)

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):

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

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

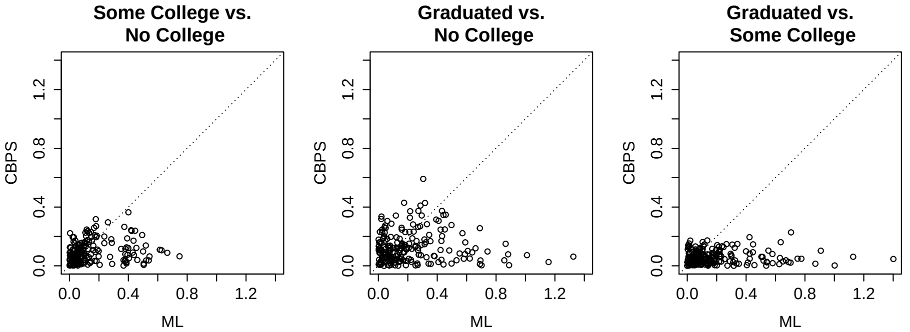
- Sample covariate balancing conditions:

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

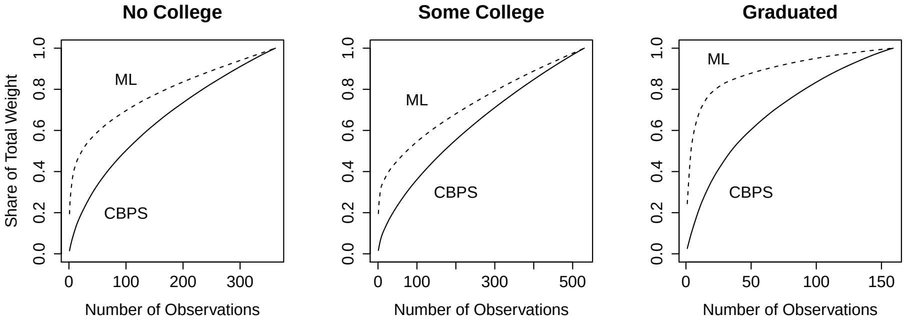
- GMM estimation: covariance matrix can be analytically calculated

Back to the Education Example: CBPS vs. ML

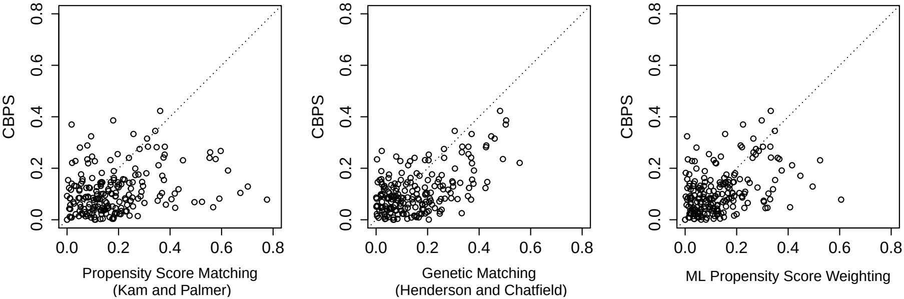
- CBPS achieves better covariate balance



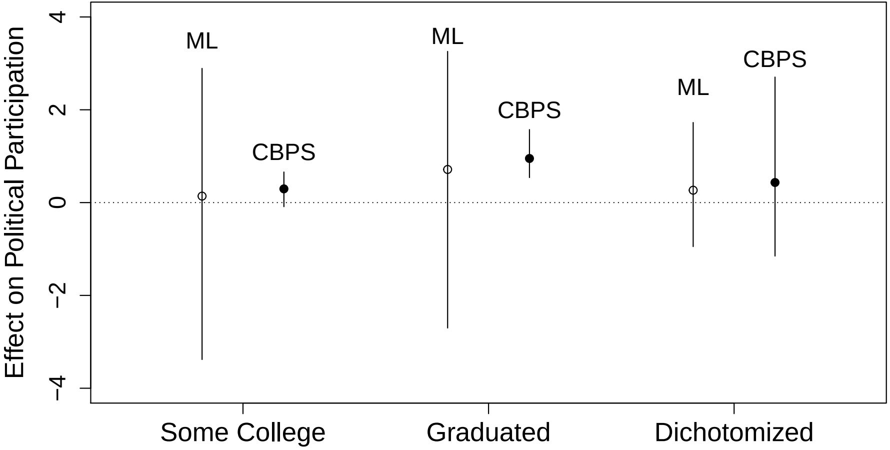
CBPS Avoids Extremely Large Weights



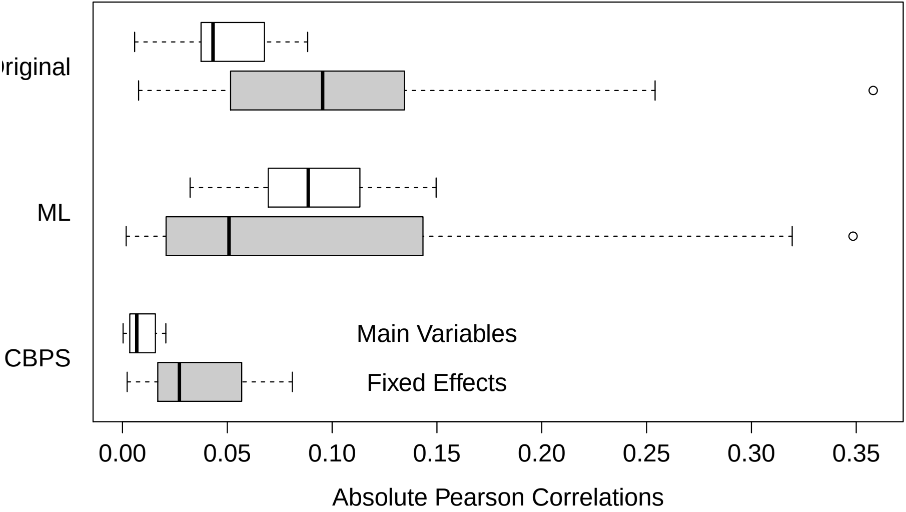
CBPS Balances Well for a Dichotomized Treatment



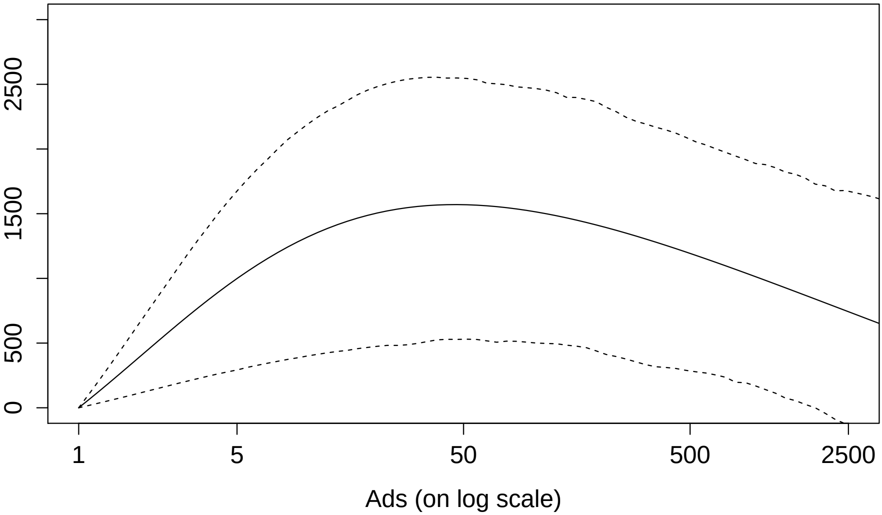
Empirical Results: Graduation Matters, Efficiency Gain



Onto the Advertisement Example



Empirical Finding: Some 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

- Ongoing extensions:
  1. nonparametric estimation via empirical likelihood
  2. generalizing instrumental variables estimates
  3. spatial treatments

References

- “Covariate Balancing Propensity Score.” *Journal of the Royal Statistical Society, Series B*, Vol. 76, pp. 243–263.
- “Robust Estimation of Inverse Probability Weights for Marginal Structural Models.” *Journal of the American Statistical Association*, Forthcoming.
- “Covariate Balancing Propensity Score for General Treatment Regimes.” *Paper* available at **http://imai.princeton.edu**
- Send comments and questions to kimai@princeton.edu