Covariate Balancing Propensity Score

Kosuke Imai

Department of Politics
Princeton University

October 29, 2013

Joint work with Marc Ratkovic

References

This talk is based on the following two papers:

1. "Covariate Balancing Propensity Score" *J. of the Royal Statistical Society, Series B (Methodological)*, in-press
2. "Robust Estimation of Inverse Probability Weights for Marginal Structural Models" working paper

Both papers available at **http://imai.princeton.edu**

Motivation

- Central role of propensity score in causal inference
  - Adjusting for observed confounding in observational studies
  - Generalizing experimental and instrumental variables estimates
- Propensity score tautology
  - sensitivity to model misspecification
  - adhoc specification searches
- Covariate Balancing Propensity Score (CBPS)
  - Estimate the propensity score such that covariates are balanced
  - Inverse probability weights for marginal structural models

Propensity Score

- 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
- Tautology: propensity score methods only work when they work

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

Kang and Schafer (2007, *Statistical Science*)

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

- 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

Sample size	Estimator	Bias GLM	Bias True	RMSE GLM	RMSE 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

		Bias		RMSE	
Sample size	Estimator	GLM	True	GLM	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

Revisiting Kang and Schafer (2007)

		Bias				RMSE			
	Estimator	GLM	CBPS1	CBPS2	True	GLM	CBPS1	CBPS2	True
(1) Both models correct									
$n = 200$	HT	0.33	2.06	−4.74	1.19	12.61	4.68	9.33	23.93
	IPW	−0.13	0.05	−1.12	−0.13	3.98	3.22	3.50	5.03
	WLS	−0.04	−0.04	−0.04	−0.04	2.58	2.58	2.58	2.58
	DR	−0.04	−0.04	−0.04	−0.04	2.58	2.58	2.58	2.58
$n = 1000$	HT	0.01	0.44	−1.59	−0.18	4.92	1.76	4.18	10.47
	IPW	0.01	0.03	−0.32	−0.05	1.75	1.44	1.60	2.22
	WLS	0.01	0.01	0.01	0.01	1.14	1.14	1.14	1.14
	DR	0.01	0.01	0.01	0.01	1.14	1.14	1.14	1.14
(2) Propensity score model correct									
$n = 200$	HT	−0.05	1.99	−4.94	−0.14	14.39	4.57	9.39	24.28
	IPW	−0.13	0.02	−1.13	−0.18	4.08	3.22	3.55	4.97
	WLS	0.04	0.04	0.04	0.04	2.51	2.51	2.51	2.51
	DR	0.04	0.04	0.04	0.04	2.51	2.51	2.52	2.51
$n = 1000$	HT	−0.02	0.44	−1.67	0.29	4.85	1.77	4.22	10.62
	IPW	0.02	0.05	−0.31	−0.03	1.75	1.45	1.61	2.27
	WLS	0.04	0.04	0.04	0.04	1.14	1.14	1.14	1.14
	DR	0.04	0.04	0.04	0.04	1.14	1.14	1.14	1.14

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

Causal Inference with Longitudinal Data

- Setup:
  - units: $i = 1, 2, \ldots, n$
  - time periods: $j = 1, 2, \ldots, J$
  - fixed J with $n \longrightarrow \infty$
  - time-varying binary treatments: $T_{ij} \in \{0, 1\}$
  - treatment history up to time j: $\overline{T}_{ij} = \{T_{i1}, T_{i2}, \ldots, T_{ij}\}$
  - time-varying confounders: X_{ij}
  - confounder history up to time j: $\overline{X}_{ij} = \{X_{i1}, X_{i2}, \ldots, X_{ij}\}$
  - outcome measured at time J: Y_i
  - potential outcomes: $Y_i(\bar{t}_J)$
- Assumptions:
  1. Sequential ignorability

$$Y_i(\bar{t}_J) \perp\!\!\!\perp T_{ij} \mid \overline{T}_{i,j-1} = \bar{t}_{j-1}, \overline{X}_{ij} = \bar{x}_j$$

     where $\bar{t}_J = (\bar{t}_{j-1}, t_j, \ldots, t_J)$
  2. Common support

$$0 < \Pr(T_{ij} = 1 \mid \overline{T}_{i,j-1}, \overline{X}_{ij}) < 1$$

Inverse-Probability-of-Treatment Weighting

- Weighting each observation via the inverse probability of its observed treatment sequence (Robins 1999)
- Potential weights:

$$w_i(\bar{t}_J, \overline{X}_{iJ}(\bar{t}_{J-1})) = \frac{1}{P(\overline{T}_{iJ} = \bar{t}_J \mid \overline{X}_{iJ}(\bar{t}_{J-1}))} = \prod_{j=1}^{J} \frac{1}{P(T_{ij} = t_{ij} \mid \overline{T}_{i,j-1} = \bar{t}_{j-1}, \overline{X}_{ij}(\bar{t}_{j-1}))}$$

- Stabilized potential weights:

$$w_i^*(\bar{t}_J, \overline{X}_{iJ}(\bar{t}_{J-1})) = \frac{P(\overline{T}_{iJ} = \bar{t}_J)}{P(\overline{T}_{iJ} = \bar{t}_J \mid \overline{X}_{iJ}(\bar{t}_{J-1}))}$$

- Observed weights: $w_i = w_i(\overline{T}_{iJ}, \overline{X}_{iJ})$ and $w_i^* = w_i^*(\overline{T}_{iJ}, \overline{X}_{iJ})$

Marginal Structural Models (MSMs)

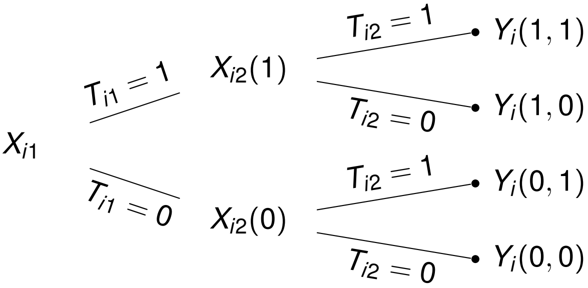
- Consistent estimation of the marginal mean of potential outcome:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\overline{T}_{iJ} = \overline{t}_J\} w_i Y_i \stackrel{p}{\longrightarrow} \mathbb{E}(Y_i(\overline{t}_J))$$

- In practice, researchers fit a weighted regression of Y_i on a function of $\overline{T}_{iJ}$ with regression weight w_i
- Adjusting for $\overline{X}_{iJ}$ leads to post-treatment bias
- MSMs estimate the average effect of any treatment sequence

- **Problem:** MSMs are sensitive to the misspecification of treatment assignment model (typically a series of logistic regressions)
- The effect of misspecification can propagate across time periods
- **Solution:** estimate MSM weights so that covariates are balanced

Two Time Period Case



- time 1 covariates X_{i1}: 3 equality constraints

$$\mathbb{E}(X_{i1}) = \mathbb{E}[\mathbf{1}\{T_{i1} = t_1, T_{i2} = t_2\} w_i X_{i1}]$$

- time 2 covariates X_{i2}: 2 equality constraints

$$\mathbb{E}(X_{i2}(t_1)) = \mathbb{E}[\mathbf{1}\{T_{i1} = t_1, T_{i2} = t_2\} w_i X_{i2}(t_1)]$$

for $t_2 = 0, 1$

Orthogonalization of Covariate Balancing Conditions

	Treatment history: (t_1, t_2)				
Time period	(0,0)	(0,1)	(1,0)	(1,1)	Moment condition
time 1	$+$	$+$	$-$	$-$	$\mathbb{E}\left\{(-1)^{T_{i1}} w_i X_{i1}\right\} = 0$
	$+$	$-$	$+$	$-$	$\mathbb{E}\left\{(-1)^{T_{i2}} w_i X_{i1}\right\} = 0$
	$+$	$-$	$-$	$+$	$\mathbb{E}\left\{(-1)^{T_{i1}+T_{i2}} w_i X_{i1}\right\} = 0$
time 2	$+$	$-$	$+$	$-$	$\mathbb{E}\left\{(-1)^{T_{i2}} w_i X_{i2}\right\} = 0$
	$+$	$-$	$-$	$+$	$\mathbb{E}\left\{(-1)^{T_{i1}+T_{i2}} w_i X_{i2}\right\} = 0$

GMM Estimator (Two Period Case)

- Independence across balancing conditions:

$$\begin{aligned}\hat{\beta} &= \underset{\beta \in \Theta}{\text{argmin}} \ \text{vec}(\mathbf{G})^\top \{\mathbf{I}_3 \otimes \mathbf{W}\}^{-1} \text{vec}(\mathbf{G}) \\ &= \underset{\beta \in \Theta}{\text{argmin}} \ \text{trace}(\mathbf{G}^\top \mathbf{W}^{-1} \mathbf{G})\end{aligned}$$

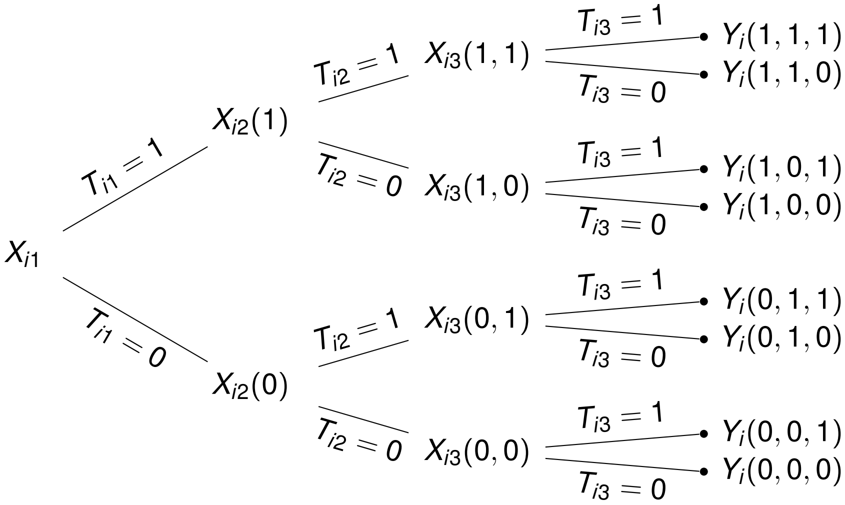
- Sample moment conditions:

$$\mathbf{G} = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} (-1)^{T_{i1}} w_i X_{i1} & (-1)^{T_{i2}} w_i X_{i1} & (-1)^{T_{i1}+T_{i2}} w_i X_{i1} \\ 0 & (-1)^{T_{i2}} w_i X_{i2} & (-1)^{T_{i1}+T_{i2}} w_i X_{i2} \end{bmatrix}$$

- Covariance matrix:

$$\mathbf{W} = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} \mathbb{E}(w_i^2 X_{i1} X_{i1}^\top \mid X_{i1}, X_{i2}) & \mathbb{E}(w_i^2 X_{i1} X_{i2}^\top \mid X_{i1}, X_{i2}) \\ \mathbb{E}(w_i^2 X_{i2} X_{i1}^\top \mid X_{i1}, X_{i2}) & \mathbb{E}(w_i^2 X_{i2} X_{i2}^\top \mid X_{i1}, X_{i2}) \end{bmatrix}$$

Extending Beyond Two Period Case



Generalization of the proposed method to J periods is in the paper

Orthogonalized Covariate Balancing Conditions

Design matrix			Treatment History Hadamard Matrix: (t_1, t_2, t_3)								Time		
			(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)			
T_{i1}	T_{i2}	T_{i3}	h_0	h_1	h_2	h_{12}	h_{13}	h_3	h_{23}	h_{123}	1	2	3
−	−	−	+	+	+	+	+	+	+	+	✗	✗	✗
+	−	−	+	−	+	−	+	−	+	−	✓	✗	✗
−	+	−	+	+	−	−	+	+	−	−	✓	✓	✗
+	+	−	+	−	−	+	+	−	−	+	✓	✓	✗
−	−	+	+	+	+	+	−	−	−	−	✓	✓	✓
+	−	+	+	−	+	−	−	+	−	+	✓	✓	✓
−	+	+	+	+	−	−	−	−	+	+	✓	✓	✓
+	+	+	+	−	−	+	−	+	+	−	✓	✓	✓

- The mod 2 discrete Fourier transform:

$$\mathbb{E}\{(-1)^{T_{i1}+T_{i3}} w_i X_{ij}\} = 0 \quad (\text{6th row})$$

- Connection to the fractional factorial design
  - "Fractional" = past treatment history
  - "Factorial" = future potential treatments

GMM in the General Case

- The same setup as before:

$$\hat{\beta} = \underset{\beta \in \Theta}{\operatorname{argmin}} \ \text{trace}(\mathbf{G}^\top \mathbf{W}^{-1} \mathbf{G})$$

where

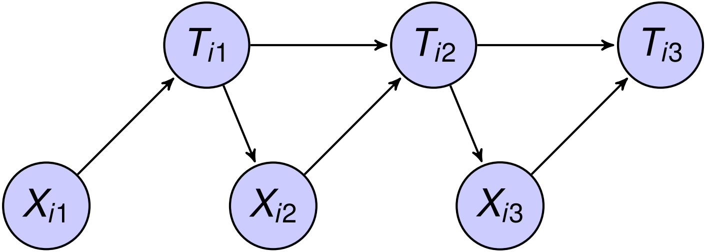
$$\mathbf{G} = \begin{bmatrix} \widetilde{\mathbf{X}}_1^\top \mathbf{M} \mathbf{R}_1 \\ \vdots \\ \widetilde{\mathbf{X}}_J^\top \mathbf{M} \mathbf{R}_J \end{bmatrix} \text{ and } \mathbf{W} = \begin{bmatrix} \mathbb{E}(\widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_1^\top \mid \mathbf{X}) & \cdots & \mathbb{E}(\widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_J^\top \mid \mathbf{X}) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(\widetilde{\mathbf{X}}_J \widetilde{\mathbf{X}}_1^\top \mid \mathbf{X}) & \cdots & \mathbb{E}(\widetilde{\mathbf{X}}_J \widetilde{\mathbf{X}}_J^\top \mid \mathbf{X}) \end{bmatrix}$$

- $\mathbf{M}$ is an $n \times (2^J - 1)$ "model matrix" based on the design matrix
- For each time period j, define $\widetilde{\mathbf{X}}_j$ and "selection matrix" $\mathbf{R}_j$

$$\widetilde{\mathbf{X}}_j = \begin{bmatrix} w_1 X_{1j}^\top \\ w_2 X_{2j}^\top \\ \vdots \\ w_n X_{nj}^\top \end{bmatrix} \quad \text{and} \quad \mathbf{R}_j = \begin{bmatrix} \mathbf{0}_{2^{j-1} \times 2^{j-1}} & \mathbf{0}_{2^{j-1} \times (2^J - 2^{j-1})} \\ \mathbf{0}_{(2^J - 2^{j-1}) \times 2^{j-1}} & \mathbf{I}_{2^J - 2^{j-1}} \end{bmatrix}$$

A Simulation Study with Correct Lag Structure

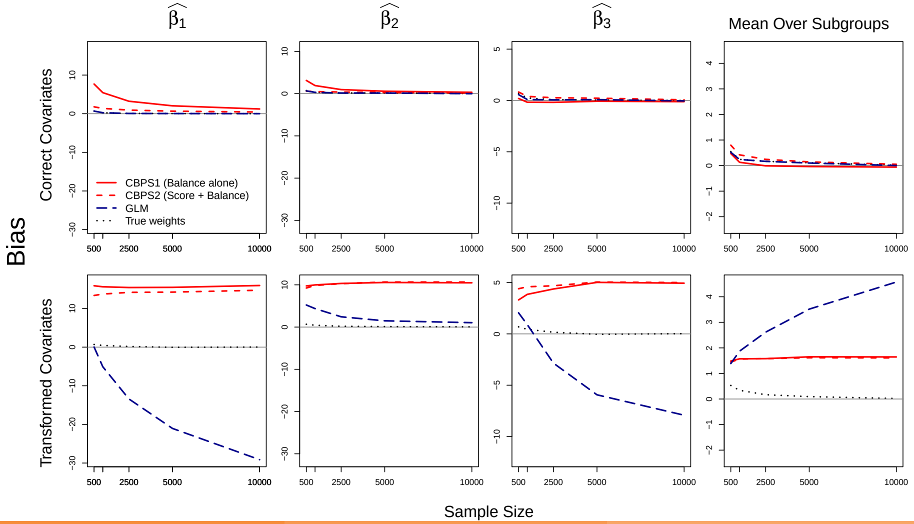
- 3 time periods
- Treatment assignment process:



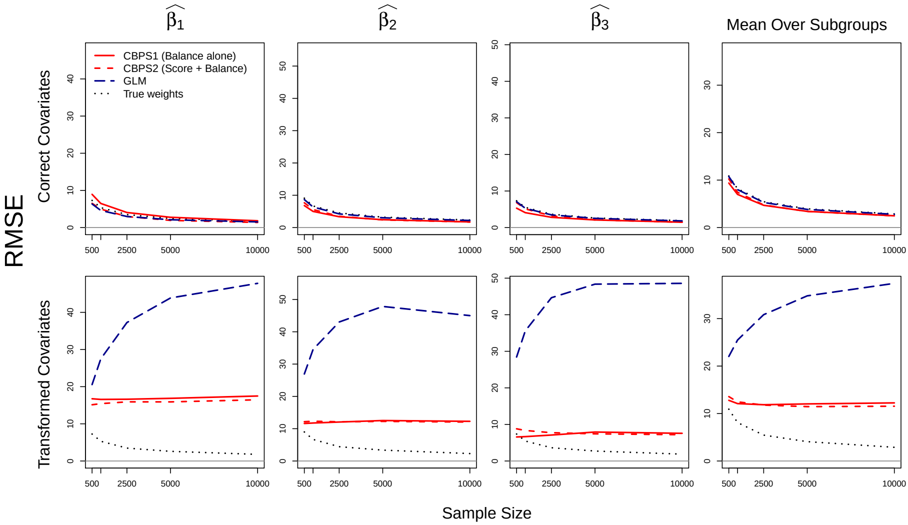
- Outcome: $Y_i = 250 - 10 \cdot \sum_{j=1}^{3} T_{ij} + \sum_{j=1}^{3} \delta^\top X_{ij} + \epsilon_i$
- Functional form misspecification by nonlinear transformation of X_{ij}

Bias

- β_j: regression coefficient for T_{ij} from marginal structural model
- Last column: mean bias for $\mathbb{E}\{Y_i(t_1, t_2, t_3)\}$

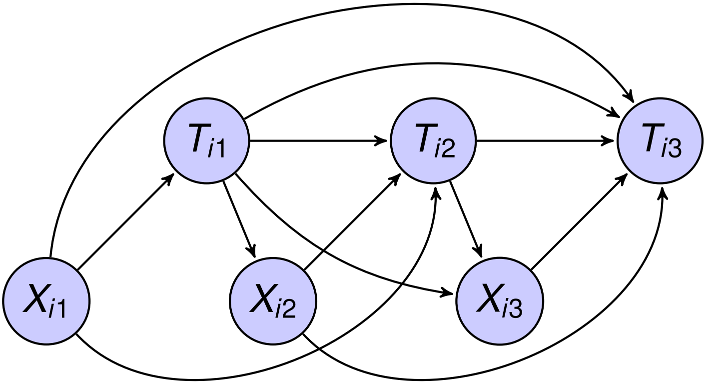


Root Mean Square Error



A Simulation Study with Incorrect Lag Structure

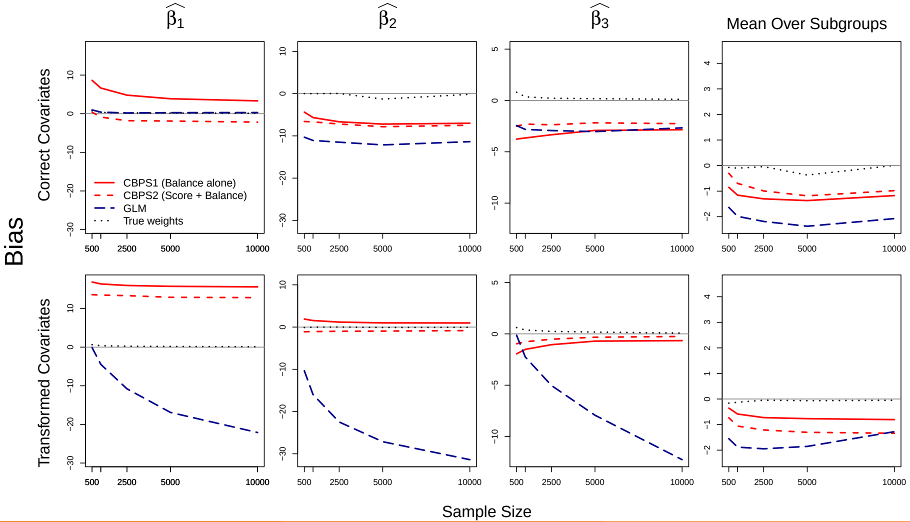
- 3 time periods
- Treatment assignment process:



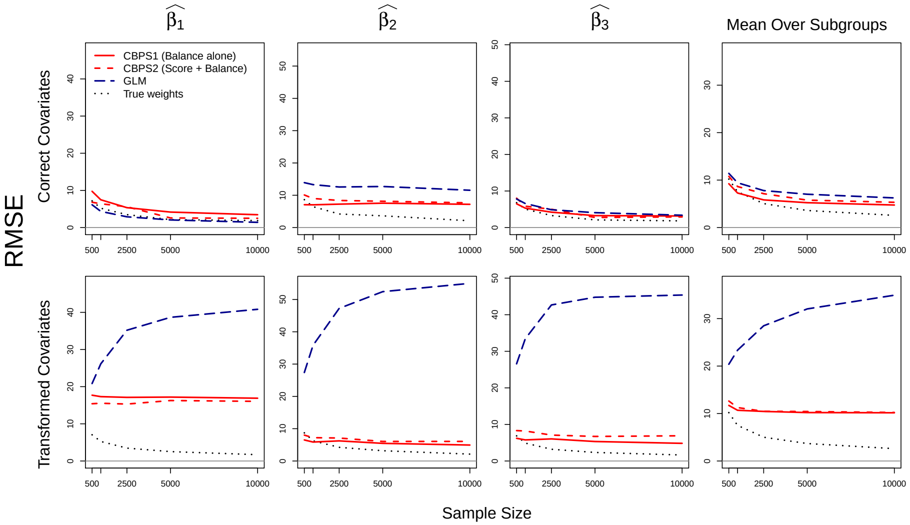
- The same outcome model
- Incorrect lag: only adjusts for previous lag but not all lags
- In addition, the same functional form misspecification of X_{ij}

Bias

- β_j: regression coefficient for T_{ij} from marginal structural model
- Last column: mean bias for $\mathbb{E}\{Y_i(t_1, t_2, t_3)\}$



Root Mean Square Error



Concluding Remarks

- Covariate balancing propensity score:
  1. optimizes covariate balance under the GMM framework
  2. is robust to model misspecification
  3. improves inverse probability weighting methods

- Ongoing work:
  1. Generalized propensity score estimation
  2. Generalizing experimental and instrumental variable estimates
  3. Confounder selection, moment selection

- Open-source software, `CBPS: R Package for Covariate Balancing Propensity Score`, is available at CRAN