Robust Estimation of Inverse Probability Weights for Marginal Structural Models

Kosuke Imai

Princeton University

2013 Joint Statistical Meetings, Montreal

August 5, 2013

Joint work with Marc Ratkovic

Motivation

- Central role of propensity score in causal inference
  - Adjusting for observed confounding in observational studies
  - Generalizing experimental and instrumental variables estimates
- Causal inference in longitudinal data
  - Marginal Structural Models (MSMs)
  - Inverse probability weighting for dynamic treatment regimes
  - Sensitivity to propensity score model misspecification
  - Difficulty of balance checking
- Covariate Balancing Propensity Score (CBPS)
  - Estimate propensity score by optimizing covariate balance
  - Generalize the cross-section case to the longitudinal setting
  - Balances within each *observed* treatment sequence
  - Balances across all *potential* future treatment sequences

CBPS in the Cross-Section Case (JRSSB, in-press)

- Setup:
  - $T_i \in \{0, 1\}$: binary treatment
  - X_i: observed pre-treatment covariates
  - $\Pr(T_i = 1 \mid X_i)$: Propensity score
- Covariate balancing conditions:

$$\mathbb{E}\left\{\frac{\mathbb{1}\{T_i = 0\}X_i}{\Pr(T_i = 0 \mid X_i)}\right\} = \mathbb{E}\left\{\frac{\mathbb{1}\{T_i = 1\}X_i}{\Pr(T_i = 1 \mid X_i)}\right\}$$

- This can be rewritten as:

$$\mathbb{E}\left\{(-1)^{T_i} w_i X_i\right\} = 0 \quad \text{where} \quad w_i = \frac{1}{P(T_i \mid X_i)}$$

- Generalized method of moments (GMM) or empirical likelihood

Marginal Structural Models (MSMs): A Review

- Setup:
  - $T_{i1}, T_{i2} \in \{0, 1\}$: Time one and two binary treatment
  - X_{i1}, X_{i2}: covariates with X_{i2} affected by T_{i1} but not T_{i2}
  - Y_i: Outcome, observed after time two
- The framework and notation generalize to J time periods:
  - Treatment history: $\overline{T}_{ij} = \{T_{i1}, \ldots, T_{ij}\}$
  - Covariate history: $\overline{X}_{ij} = \{X_{i1}, \ldots, X_{ij}\}$
- Assumptions:
  1. Sequential ignorability:

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

     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 Weights for MSMs

- Inverse probability weights:

$$w_i = \frac{1}{P(T_{i1}, T_{i2} \mid X_{i1}, X_{i2})} \quad \text{and} \quad sw_i = \frac{P(T_{i1}, T_{i2})}{P(T_{i1}, T_{i2} \mid X_{i1}, X_{i2})}$$

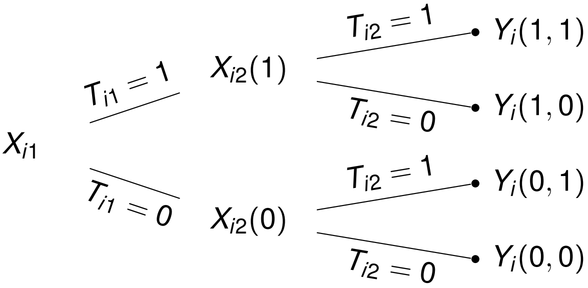
- In the general J period case:

$$w_i = \frac{1}{\prod_{j=1}^{J} P(T_{ij} \mid T_{i,j-1}, \overline{X}_{ij})} \quad \text{and} \quad sw_i = \frac{\prod_{j=1}^{J} P(T_{ij} \mid \overline{T}_{i,j-1})}{\prod_{j=1}^{J} P(T_{ij} \mid T_{i,j-1}, \overline{X}_{ij})}$$

- Typically, propensity scores are estimated by a parametric model
- MSM weights $=$ product of estimated propensity scores

  $\Longrightarrow$ sensitivity to model misspecification
- CBPS: estimate MSM weights so that covariate balance is optimized

Balancing Conditions in the Two Period Case



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

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

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

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

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:

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

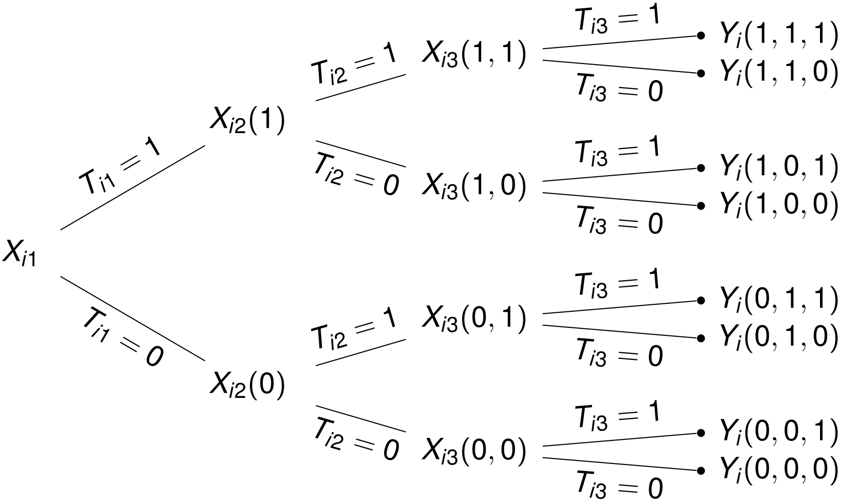
- Sample moment conditions:

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

- 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



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
$-$	$-$	$-$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	✗	✗	✗
$+$	$-$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$	✓	✗	✗
$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$	$+$	$-$	$-$	✓	✓	✗
$+$	$+$	$-$	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	✓	✓	✗
$-$	$-$	$+$	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$-$	✓	✓	✓
$+$	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$	✓	✓	✓
$-$	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$-$	$+$	$+$	✓	✓	✓
$+$	$+$	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	✓	✓	✓

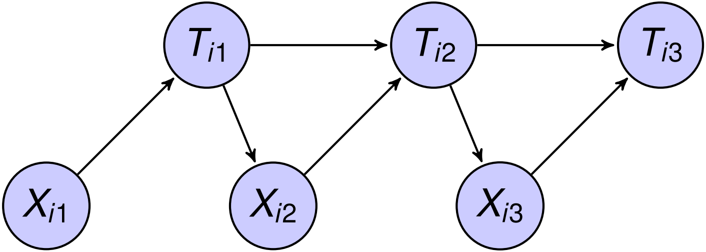
- The mod 2 discrete Fourier transform:

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

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

A Simulation Study with Correct Lag Structure

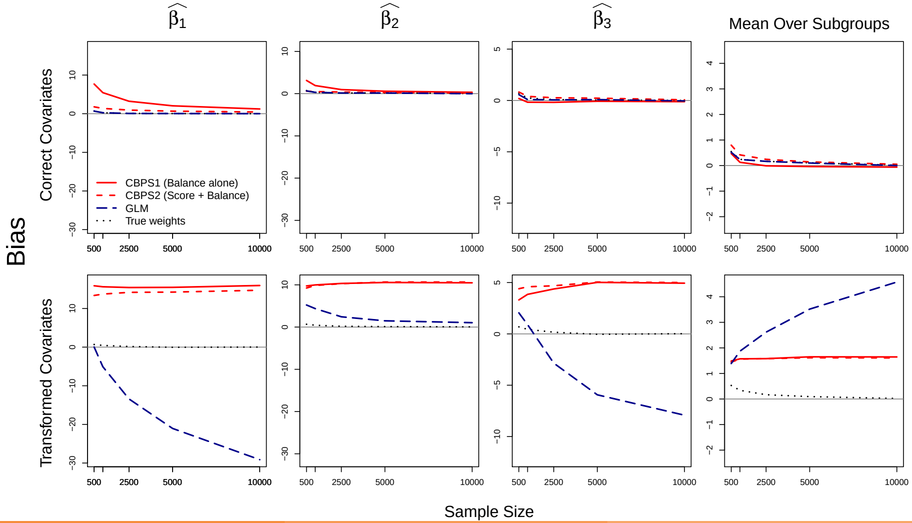
- 3 time periods
- Treatment assignment process (logistic model):



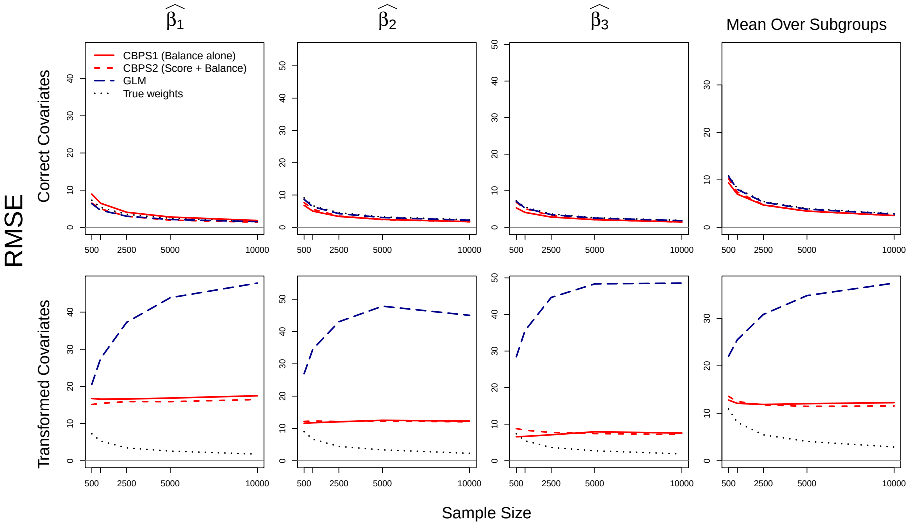
- Outcome: $Y_i = 250 - 10 \cdot \sum_{j=1}^{3} T_{ij} + \sum_{j=1}^{3} \delta^\top X_{ij} + \epsilon_i$
- Functional form misspecification by nonlinear transformation of X_{ij}

Bias

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

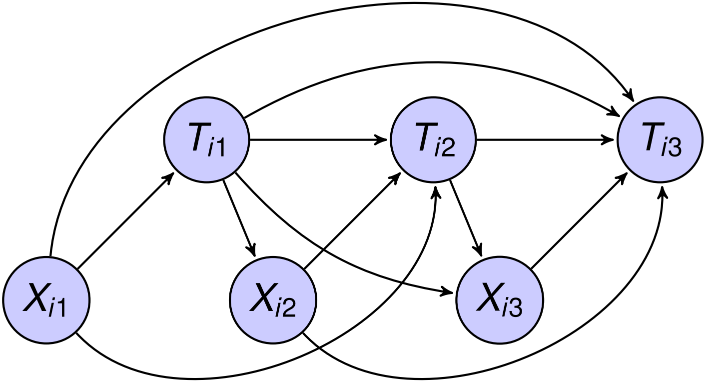


Root Mean Square Error



A Simulation Study with Incorrect Lag Structure

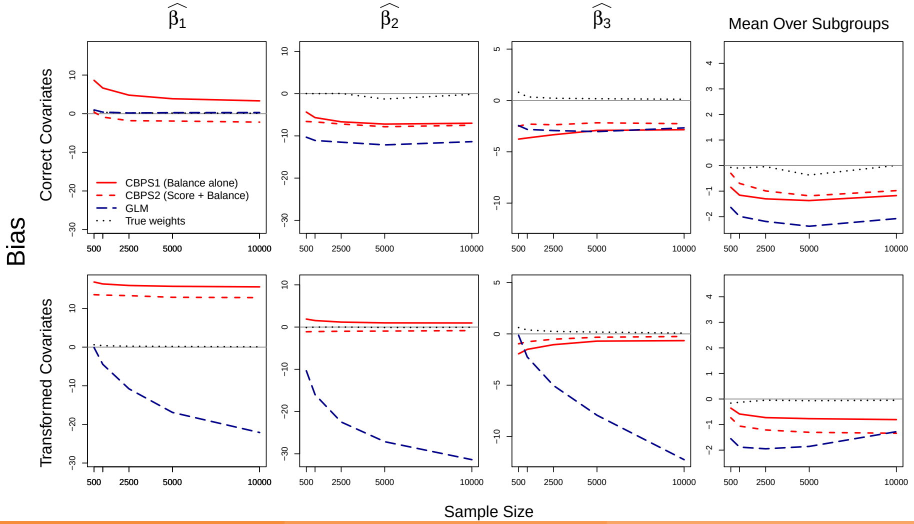
- 3 time periods
- Treatment assignment process (logistic model):



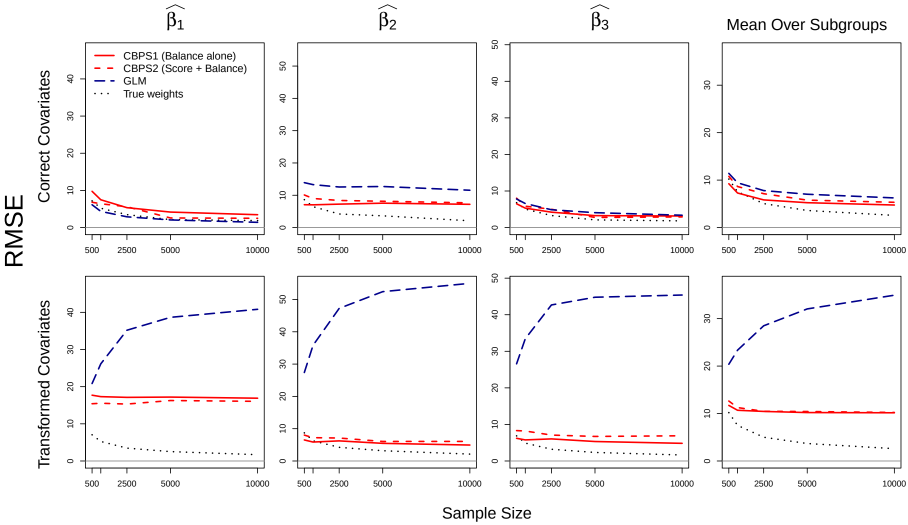
- The same outcome model
- Incorrect lag: only adjusts for previous lag but not all lags
- In addition, the same functional form misspecification of X_{ij}

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 (CBPS):
  1. optimizes covariate balance when estimating propensity score
  2. is more robust to model misspecification than a standard method
  3. improves inverse probability weighting methods including MSMs

- 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

GMM in the General Case

- The same setup as before:

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

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" $\mathbf{R}_j$

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