Covariate Balancing Propensity Score

Kosuke Imai

Princeton University

March 7, 2013
SREE Spring 2013 Conference, Washington D.C.

Joint work with Marc Ratkovic

References

This talk is based on the following two papers:

1. "Covariate Balancing Propensity Score"
2. "Robust Estimation of Inverse Probability Weights for Marginal Structural Models"

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

Motivation

- Causal inference is a central goal of scientific research
- Randomized experiments are not always possible
  $\Longrightarrow$ Causal inference in observational studies
- Experiments often lack external validity
  $\Longrightarrow$ Need to generalize experimental results
- Importance of statistical methods to adjust for confounding factors

Overview of the Workshop

1. **Review:** Propensity score
    - propensity score is a covariate balancing score
    - matching and weighting methods
2. **Problem:** Propensity score tautology
    - sensitivity to model misspecification
    - adhoc specification searches
3. **Solution:** **Covariate balancing propensity score (CBPS)**
    - Estimate propensity score so that covariate balance is optimized
4. **Evidence:** Reanalysis of two prominent critiques
    - Improved performance of propensity score weighting and matching
5. **Extension:** Marginal structural models for longitudinal data
    - CBPS for time-varying treatments and confounders
    - Simulation evidence
6. **Software:** R package `CBPS`

Propensity Score

- Setup:
  - $T_i \in \{0, 1\}$: binary treatment
  - X_i: pre-treatment covariates
  - $(Y_i(1), Y_i(0))$: potential outcomes
  - $Y_i = Y_i(T_i)$: observed outcomes
- Definition: conditional probability of treatment assignment

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

- Balancing property (without assumption):

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

Rosenbaum and Rubin (1983)

- Assumptions:
  1. Overlap:
  $$0 < \pi(X_i) < 1$$
  2. Unconfoundedness:
  $$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid X_i$$

- Propensity score as a dimension reduction tool:
$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid \pi(X_i)$$

Matching and Weighting via Propensity Score

- Propensity score reduces the dimension of covariates
- But, propensity score must be estimated (more on this later)
- Once estimated, simple nonparametric adjustments are possible

- Matching
- Subclassification
- Weighting (Horvitz-Thompson estimator):

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

  often, weights are normalized
- Doubly-robust estimators (Robins *et al.*):

$$\frac{1}{n}\sum_{i=1}^{n}\left[\left\{\hat{\mu}(1,X_i) + \frac{T_i(Y_i - \hat{\mu}(1,X_i))}{\hat{\pi}(X_i)}\right\} - \left\{\hat{\mu}(0,X_i) + \frac{(1-T_i)(Y_i - \hat{\mu}(0,X_i))}{1-\hat{\pi}(X_i)}\right\}\right]$$

- They have become standard tools for applied researchers

Propensity Score Tautology

- Propensity score is unknown
- Dimension reduction is purely theoretical: must model T_i given X_i
- Diagnostics: covariate balance checking
- In practice, adhoc specification searches are conducted
- Model misspecification is always possible

- Theory (Rubin *et al.*): ellipsoidal covariate distributions
  $\Longrightarrow$ equal percent bias reduction
- Skewed covariates are common in applied settings

- Propensity score methods can be sensitive to misspecification

Kang and Schafer (2007, *Statistical Science*)

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

- Setup:
  - 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 to be evaluated:
  1. Horvitz-Thompson
  2. Inverse-probability weighting with normalized weights
  3. Weighted least squares regression
  4. Doubly-robust least squares regression

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

		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

Smith and Todd (2005, *J. of Econometrics*)

- LaLonde (1986; *Amer. Econ. Rev.*):
  - Randomized evaluation of a job training program
  - Replace experimental control group with another non-treated group
  - Current Population Survey and Panel Study for Income Dynamics
  - Many evaluation estimators didn't recover experimental benchmark
- Dehejia and Wahba (1999; *J. of Amer. Stat. Assoc.*):
  - Apply propensity score matching
  - Estimates are close to the experimental benchmark
- Smith and Todd (2005):
  - Dehejia & Wahba (DW)'s results are sensitive to model specification
  - They are also sensitive to the selection of comparison sample

Propensity Score Matching Fails Miserably

- One of the most difficult scenarios identified by Smith and Todd:
  - LaLonde experimental sample rather than DW sample
  - Experimental estimate: \$886 (s.e. $=$ 488)
  - PSID sample rather than CPS sample
- Evaluation bias:
  - Conditional probability of being in the experimental sample
  - Comparison between experimental control group and PSID sample
  - "True" estimate $=0$
  - Logistic regression for propensity score
  - One-to-one nearest neighbor matching with replacement

Propensity score model	Estimates
Linear	-835
	(886)
Quadratic	-1620
	(1003)
Smith and Todd (2005)	-1910
	(1004)

Covariate Balancing Propensity Score

- Idea: Estimate the propensity score such that covariate balance is optimized
- Covariate balancing condition:
  - For the Average Treatment Effect (ATE)

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

  - For the Average Treatment Effect for the Treated (ATT)

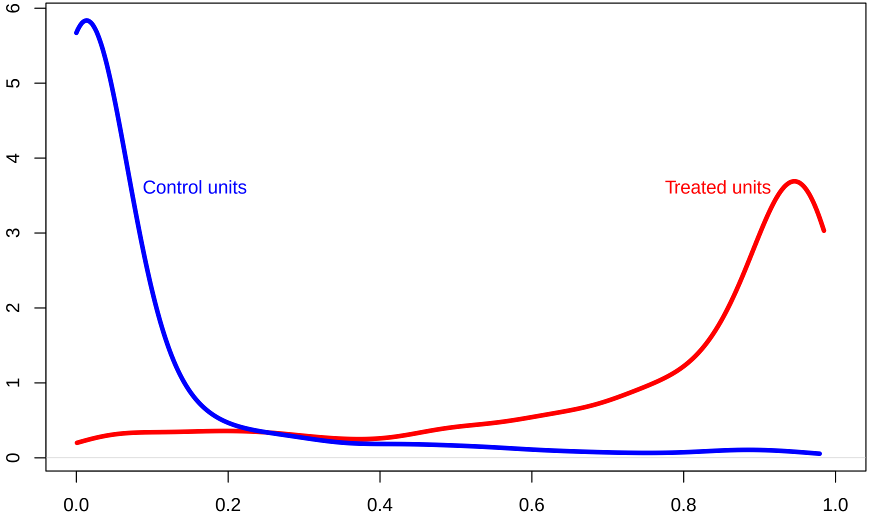
$$\mathbb{E}\left\{T_i\widetilde{X}_i - \frac{\pi_\beta(X_i)(1-T_i)\widetilde{X}_i}{1-\pi_\beta(X_i)}\right\} = 0$$

  where $\widetilde{X}_i = f(X_i)$ is any vector-valued function
- Score condition from maximum likelihood:

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

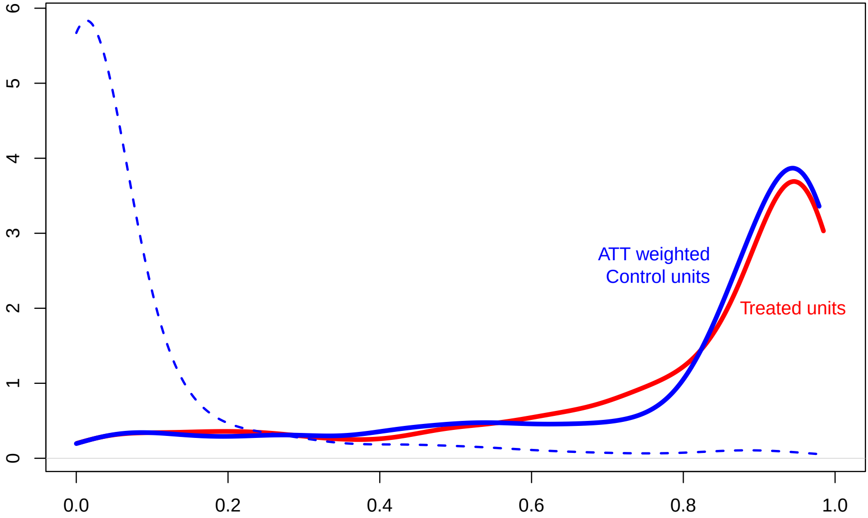
Weighting Control Group to Balance Covariates

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



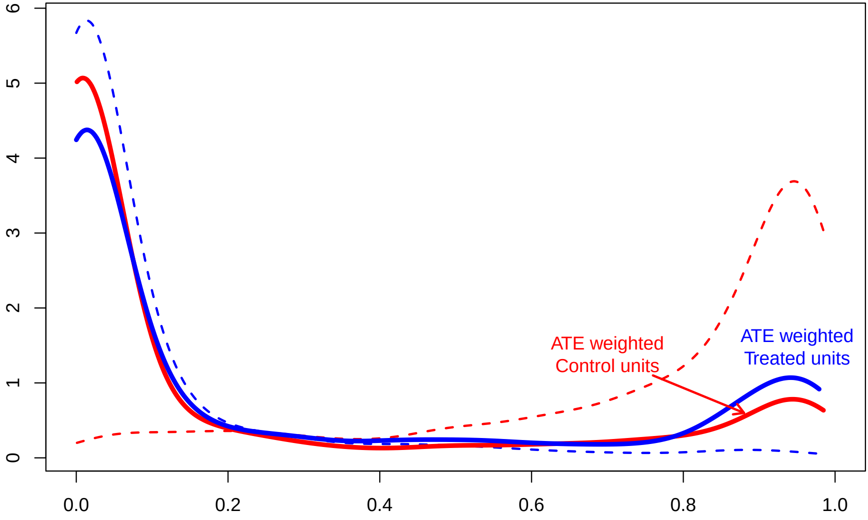
Weighting Control Group to Balance Covariates

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



Weighting Both Groups to Balance Covariates

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



Generalized Method of Moments (GMM) Framework

- Just-identified CBPS: covariate balancing conditions alone
- Over-identified CBPS: combine them with score conditions
- GMM (Hansen 1982):

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

where

$$\bar{g}_\beta(T, X) = \frac{1}{N} \sum_{i=1}^{N} \underbrace{\begin{pmatrix} \text{score condition} \\ \text{balancing condition} \end{pmatrix}}_{g_\beta(T_i, X_i)}$$

- "Continuous updating" GMM estimator with the following Σ:

$$\Sigma_\beta(T, X) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(g_\beta(T_i, X_i) g_\beta(T_i, X_i)^\top \mid X_i)$$

Specification Test and Optimal Matching

- CBPS is overidentified
- Specification test based on Hansen's J-statistic:

$$J \;=\; n\bar{g}_\beta(T,X)^\top \Sigma_\beta(T,X)^{-1}\bar{g}_\beta(T,X) \;\sim\; \chi^2_k$$

  where k is the number of moment conditions

- Can also be used to select matching estimators
- Example: Optimal 1-to-N matching
  - Assume N control units matched with each treated unit
  - Calculate J statistic by downweighting matched control units with weight $1/N$
  - Choose N such that J statistic is minimized

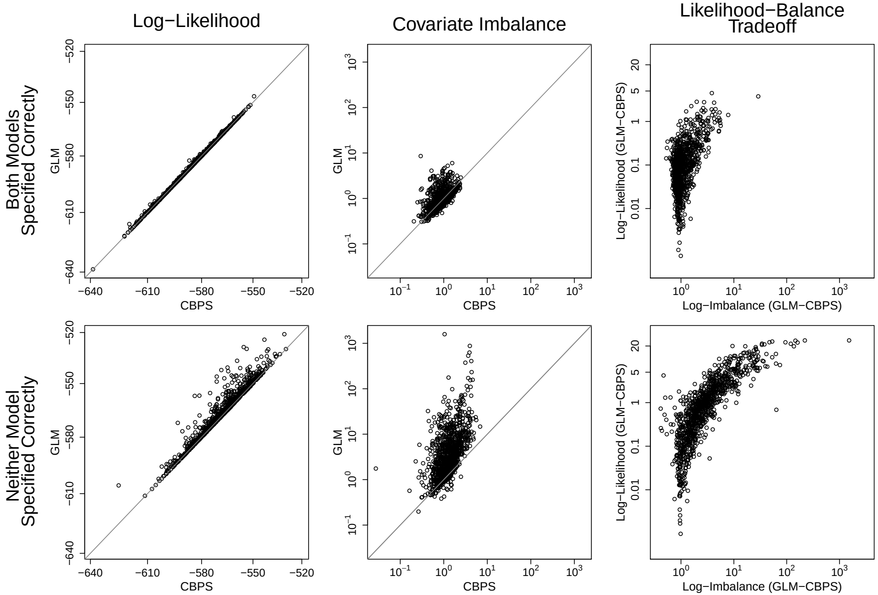
Revisiting Kang and Schafer (2007)

		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

CBPS Sacrifices Likelihood for Better Balance



Revisiting Smith and Todd (2005)

- Evaluation bias: "true" bias $= 0$
- CBPS improves propensity score matching across specifications and matching methods
- However, specification test rejects the null

	1-to-1 Nearest Neighbor			Optimal 1-to-N Nearest Neighbor		
Specification	GLM	Balance	CBPS	GLM	Balance	CBPS
Linear	-835	-559	-302	-885	-257	-38
	(886)	(898)	(873)	(435)	(492)	(488)
Quadratic	-1620	-967	-1040	-1270	-306	-140
	(1003)	(882)	(831)	(406)	(407)	(392)
Smith & Todd	-1910	-1040	-1313	-1029	-672	-32
	(1004)	(860)	(800)	(413)	(387)	(397)

Standardized Covariate Imbalance

- Covariate imbalance in the (Optimal 1–to–N) matched sample
- Standardized difference-in-means

	Linear			Quadratic			Smith & Todd		
	GLM	Balance	CBPS	GLM	Balance	CBPS	GLM	Balance	CBPS
Age	−0.060	−0.035	−0.063	−0.060	−0.035	−0.063	−0.031	0.035	−0.013
Education	−0.208	−0.142	−0.126	−0.208	−0.142	−0.126	−0.262	−0.168	−0.108
Black	−0.087	0.005	−0.022	−0.087	0.005	−0.022	−0.082	−0.032	−0.093
Married	0.145	0.028	0.037	0.145	0.028	0.037	0.171	0.031	0.029
High school	0.133	0.089	0.174	0.133	0.089	0.174	0.189	0.095	0.160
74 earnings	−0.090	0.025	0.039	−0.090	0.025	0.039	−0.079	0.011	0.019
75 earnings	−0.118	0.014	0.043	−0.118	0.014	0.043	−0.120	−0.010	0.041
Hispanic	0.104	−0.013	0.000	0.104	−0.013	0.000	0.061	0.034	0.102
74 employed	0.083	0.051	−0.017	0.083	0.051	−0.017	0.059	0.068	0.022
75 employed	0.073	−0.023	−0.036	0.073	−0.023	−0.036	0.099	−0.027	−0.098
Log-likelihood	−326	−342	−345	−293	−307	−297	−295	−231	−296
Imbalance	0.507	0.264	0.312	0.544	0.304	0.300	0.515	0.359	0.383

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

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

  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:

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

- 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:

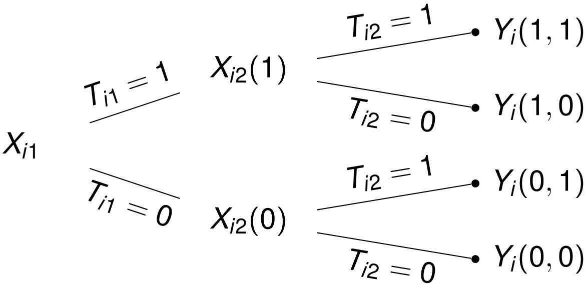
$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\overline{T}_{iJ} = \bar{t}_J\} w_i Y_i \xrightarrow{p} \mathbb{E}(Y_i(\bar{t}_J))$$

- 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

Practical Challenges of Marginal Structural Models

- 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
- Checking covariate balance is difficult
- Balancing covariates at each time period is not sufficient
- E.g., baseline covariates should be balanced across all 2^J groups
- **Solution:** estimate MSM weights so that all covariate balancing conditions are satisfied as much as possible

Two Time Period Case



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

$$\mathbb{E}(X_{i1}) \ = \ \mathbb{E}[\mathbf{1}\{T_{i1} = t_1, T_{i2} = t_2\} w_i(\bar{t}_2, \overline{X}_{i2}(t_1)) \, X_{i1}]$$

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

$$\mathbb{E}(X_{i2}(t_1)) \ = \ \mathbb{E}[\mathbf{1}\{T_{i1} = t_1, T_{i2} = t_2\} w_i(\bar{t}_2, \overline{X}_{i2}(t_1)) \, X_{i2}(t_1)]$$

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 covariate balancing conditions:

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

- Sample moment conditions:

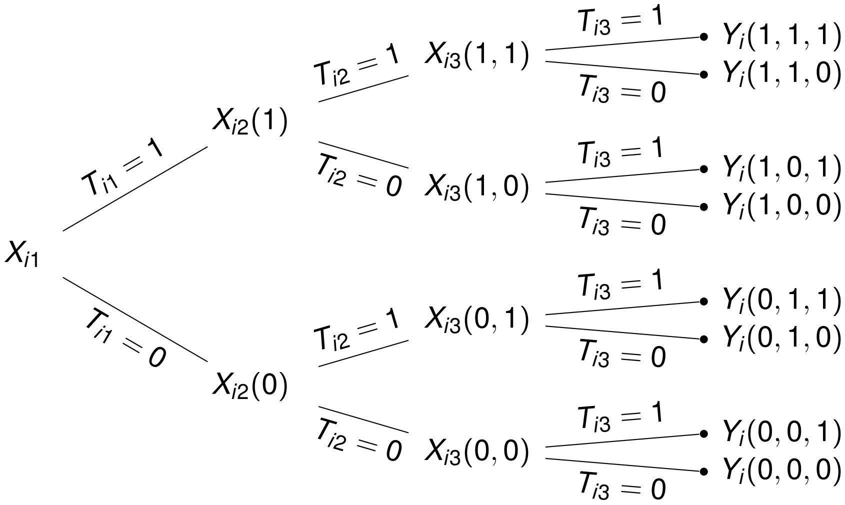
$$\mathbf{G} = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} (-1)^{T_{i1}} w_i X_{i1} & (-1)^{T_{i2}} w_i X_{i1} & (-1)^{T_{i1}+T_{i2}} w_i X_{i1} \\ 0 & (-1)^{T_{i2}} w_i X_{i2} & (-1)^{T_{i1}+T_{i2}} w_i X_{i2} \end{bmatrix}$$

- Covariance matrix (dependence across time periods):

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

- Possible to combine them with score conditions

Extending Beyond Two Period Case



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

Orthogonalized Covariate Balancing Conditions

			Treatment History Hadamard Matrix: (t_1, t_2, t_3)										
Design matrix			(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)	Time		
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
−	−	−	+	+	+	+	+	+	+	+	✗	✗	✗
+	−	−	+	−	+	−	+	−	+	−	✓	✗	✗
−	+	−	+	+	−	−	+	+	−	−	✓	✓	✗
+	+	−	+	−	−	+	+	−	−	+	✓	✓	✗
−	−	+	+	+	+	+	−	−	−	−	✓	✓	✓
+	−	+	+	−	+	−	−	+	−	+	✓	✓	✓
−	+	+	+	+	−	−	−	−	+	+	✓	✓	✓
+	+	+	+	−	−	+	−	+	+	−	✓	✓	✓

- Covariate balancing conditions:

$$\mathbb{E}\{X_{ij}(\bar{t}_{j-1})\} = \mathbb{E}[\mathbf{1}\{\overline{T}_{j-1} = \bar{t}_{j-1}, \underline{T}_{ij} = \underline{t}_j\} w_i(\bar{t}_J, \overline{X}_{iJ}(\bar{t}_{J-1})) X_{ij}(\bar{t}_{j-1})]$$

- The mod 2 discrete Fourier transform:

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

GMM in the General Case

- The same setup as before:

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

where

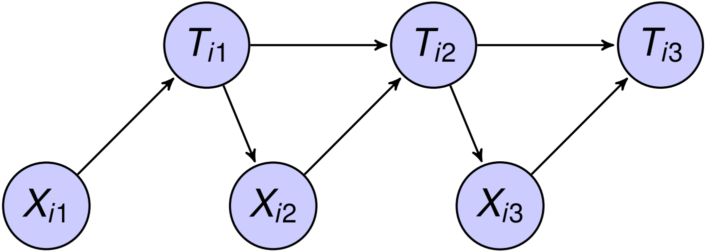
$$\mathbf{G} = \begin{bmatrix} \widetilde{\mathbf{X}}_1^\top \mathbf{M} \mathbf{R}_1 \\ \vdots \\ \widetilde{\mathbf{X}}_J^\top \mathbf{M} \mathbf{R}_J \end{bmatrix} \text{ and } \mathbf{W} = \begin{bmatrix} \mathbb{E}(\widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_1^\top \mid \mathbf{X}) & \cdots &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

- $\mathbf{M}$ 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" $\mathbf{R}_j$

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

A Simulation Study with Correct Lag Structure

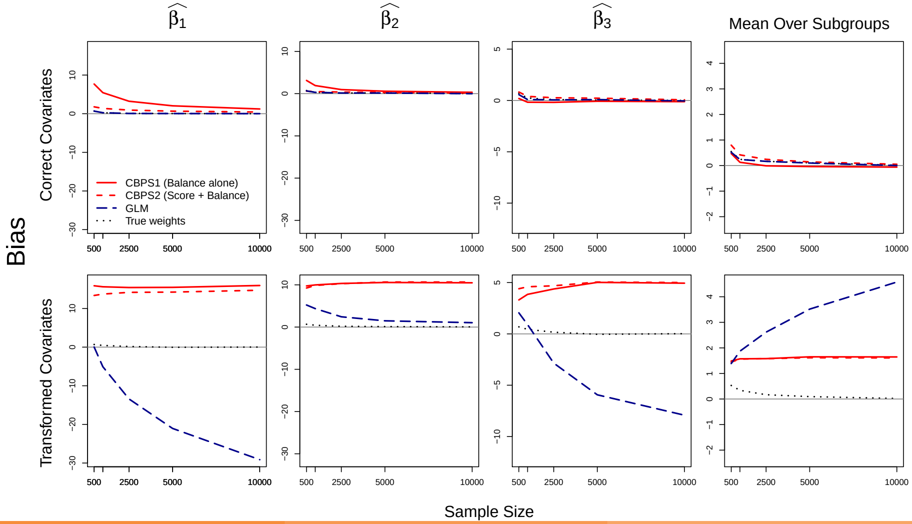
- 3 time periods
- Treatment assignment process:



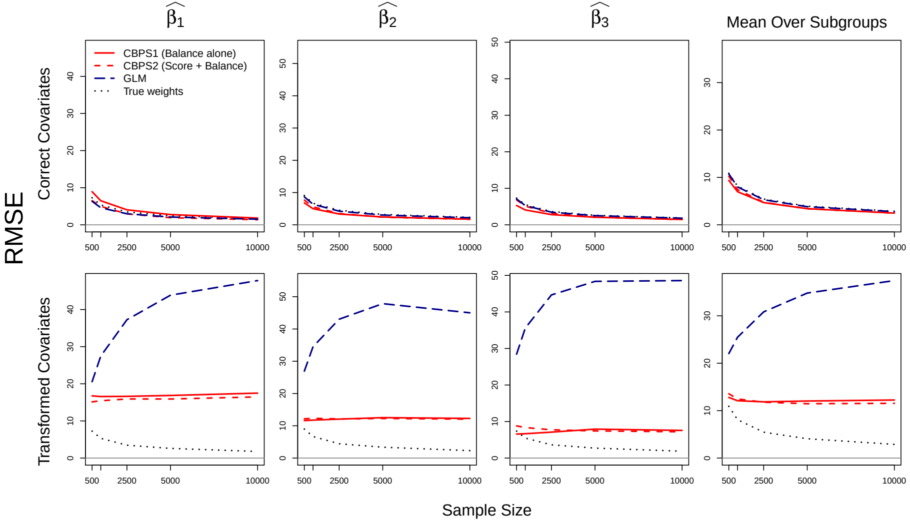
- Outcome: $Y_i = 250 - 10 \cdot \sum_{j=1}^{3} T_{ij} + \sum_{j=1}^{3} \delta^\top X_{ij} + \epsilon_i$
- Functional form misspecification by nonlinear transformation of X_{ij}

Bias

- β_j: the average marginal effect of T_{ij}
- Last column: mean bias for $\mathbb{E}\{Y_i(t_1, t_2, t_3)\}$

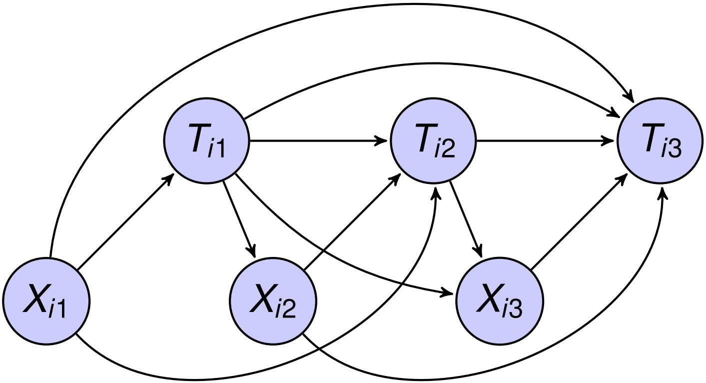


Root Mean Square Error



A Simulation Study with Incorrect Lag Structure

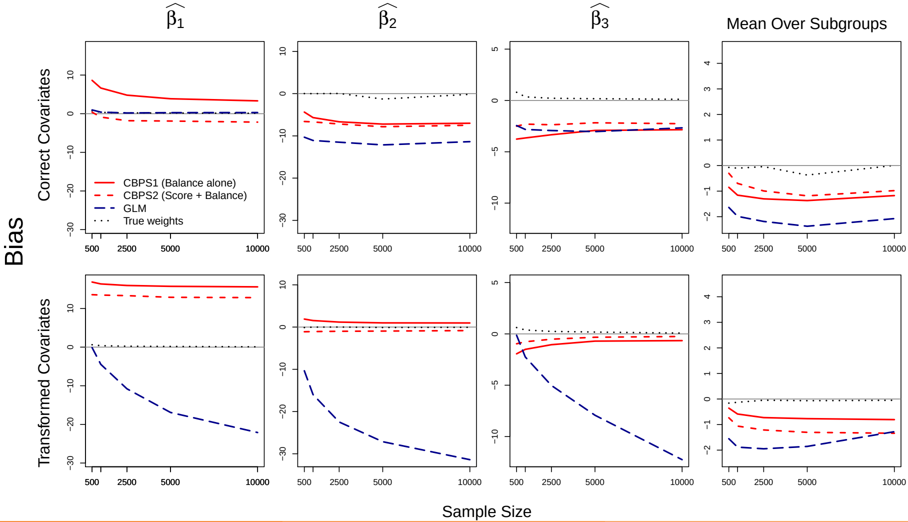
- 3 time periods
- Treatment assignment process:



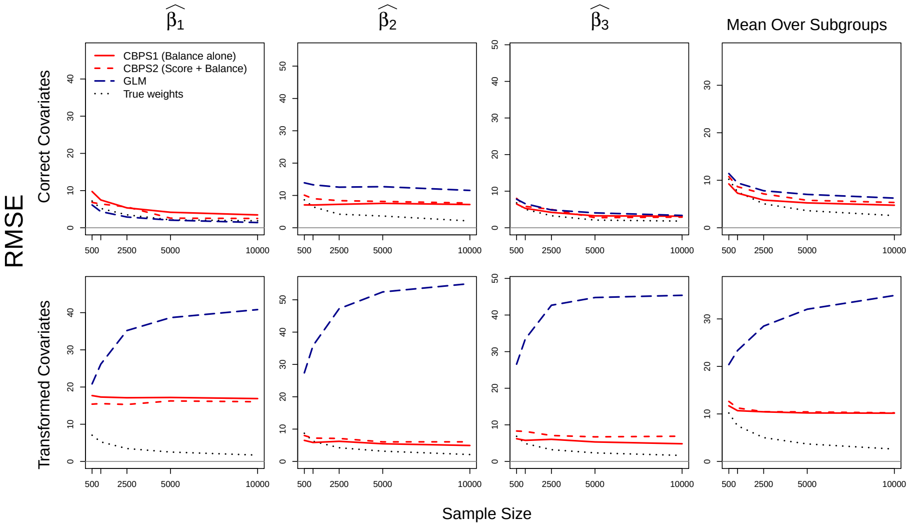
- The same outcome model
- Incorrect lag: only adjusts for previous lag but not all lags
- In addition, the same functional form misspecification of X_{ij}

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



Software: R Package CBPS

```
## upload the package
library("CBPS")
## load the LaLonde data
data(LaLonde)
## Estimate ATT weights via CBPS
fit <- CBPS(treat ~ age + educ + re75 + re74 +
                    I(re75==0) + I(re74==0),
            data = LaLonde, ATT = TRUE)
summary(fit)
## matching via MatchIt
library(MatchIt)
## one to one nearest neighbor with replacement
m.out <- matchit(treat ~ 1, distance = fitted(fit),
                 method = "nearest", data = LaLonde,
                 replace = TRUE)
summary(m.out)
```

Extensions to Other Causal Inference Settings

- Propensity score methods are widely applicable
- This means that CBPS is also widely applicable
- Extensions in progress:
  1. Non-binary treatment regimes
  2. Generalizing experimental estimates
  3. Generalizing instrumental variable estimates
- All of these are situations where balance checking is difficult

Non-binary Treatment Regimes

1. Multi-valued treatment:
   - Propensity score for each value: $\pi_\beta(t, X_i) = \Pr(T_i = t \mid X_i)$
   - Commonly used models: multinomial logit, ordinal logit
   - Inverse probability weighting: weight $= 1/\pi_\beta(T_i, X_i)$
   - Balance covariates across all groups
   - Essentially the same as MSM case: much simpler
2. Continuous and other treatments:
   - Generalized propensity score: $\pi_\beta(t, X_i) = p(T_i = t \mid X_i)$
   - Propensity function: $\psi_\beta(X_i)$ where $p_\psi(T_i = t \mid X_i)$
   - Commonly used models: linear regression, GLMs
   - Outcome analysis:
     - subclassification (Imai and van Dyk)
     - polynomial regression (Hirano and Imbens)
   - Sensitivity to model misspecification, lack of diagnostics
   - Use the same model but balance covariates across binned categories

Generalizing Experimental Estimates

- Lack of external validity for experimental estimates
- Target population $\mathcal{P}$
- Experimental sample: $S_i = 1$ with $i = 1, 2, \ldots, N_e$
- Non-experimental sample: $S_i = 0$ with $i = N_e + 1, \ldots, N$
- Sampling on observables:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp S_i \mid X_i$$

- Propensity score: $\pi_\beta(X_i) = \Pr(S_i = 1 \mid X_i)$
- Outcome analysis: weighted regression for the experimental sample

- Balancing between experimental and non-experimental sample
- You may also balance weighted treatment and control groups within the experimental sample

Review of Instrumental Variables

- Encouragement design (Angrist et al. *JASA*)
- Randomized encouragement: $Z_i \in \{0, 1\}$
- Potential treatment variables: $T_i(z)$ for $z = 0, 1$
- Four principal strata (latent types):
  - compliers $(T_i(1), T_i(0)) = (1, 0)$,
  - non-compliers $\left\{ \begin{array}{cc} always - takers & (T_i(1), T_i(0)) = (1, 1), \\ never - takers & (T_i(1), T_i(0)) = (0, 0), \\ defiers & (T_i(1), T_i(0)) = (0, 1) \end{array} \right.$
- Observed and principal strata:

	$Z_i = 1$	$Z_i = 0$
$T_i = 1$	Complier/Always-taker	Defier/Always-taker
$T_i = 0$	Defier/Never-taker	Complier/Never-taker

- Randomized encouragement as an instrument for the treatment
- Two additional assumptions
  1. Monotonicity: No defiers
  $$T_i(1) \geq T_i(0) \quad \text{for all } i.$$
  2. Exclusion restriction: Instrument (encouragement) affects outcome only through treatment
  $$Y_i(1, t) \ = \ Y_i(0, t) \quad \text{for } t = 0, 1$$
  Zero ITT effect for always-takers and never-takers
- ITT effect decomposition:

$$\begin{aligned} \text{ITT} &= \text{ITT}_c \times \Pr(\text{compliers}) + \text{ITT}_a \times \Pr(\text{always} - \text{takers}) \\ &\quad + \text{ITT}_n \times \Pr(\text{never} - \text{takers}) \\ &= \text{ITT}_c \Pr(\text{compliers}) \end{aligned}$$

- Complier average treatment effect or (LATE): $\text{ITT}_c = \text{ITT} / \Pr(\text{compliers})$

Generalizing Instrumental Variables Estimates

- Compliers may not be of interest
  1. They are a latent type
  2. They depend on the encouragement
- Generalize LATE to ATE
- No unmeasured confounding: ATE $=$ LATE given X_i
- Propensity score: $\pi_\beta(X_i) = \Pr(C_i = \text{complier} \mid X_i)$
- Weighted two-stage least squares with the weight $= 1/\pi_\beta(X_i)$
- Commonly used model: the multinomial mixture (Imbens & Rubin)
- Balance covariates across four observed cells defined by (Z_i, T_i)
  - Weights are based on the probability of different types
  - For example, for the cell with $(Z_i, T_i) = (1, 1)$, use the inverse of $\Pr(C_i = \text{complier} \mid X_i) + \Pr(C_i = \text{always} - \text{taker} \mid X_i)$ as weight

Concluding Remarks

- Covariate balancing propensity score:
  1. simultaneously optimizes prediction of treatment assignment and covariate balance under the GMM framework
  2. is robust to model misspecification
  3. improves propensity score weighting and matching methods
  4. can be extended to various situations
- Open questions:
  1. How to select confounders
  2. How to specify a treatment assignment model
  3. How to choose covariate balancing conditions