Identification and Estimation of Marginal Structural Mean Models with Instrumental Variables

Haben Michael, Yifan Cui, Eric Tchetgen Tchetgen

February 2019

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

Outline

Introduction **Potential outcome framework**
Relaxing SRA Randomization
Simulation No unmeasured confounders (Rubin et al. 80s)
Closing remarks Sequential Randomization Assumption (Robins '80s/'90s)

A motivating example

Sports argument: Bill Belichick is a great
coach: When average players come to play for
him, their performance jumps up; when they
leave, they become average again.

    Causal framework: Players have a potential outcome, their
performance under Belichick and not under Belichick. This
pre/post comparison approximates the difference between the
potential outcomes.

    Sports counter-argument: It's not Belichick. He's just been
fortunate . . .

    Causal framework: The association between Belichick
(treatment) and player performance (outcome) is spurious; it exists
due to a confounder. Is there some variable associated with
treatment and outcome?

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

Sports counter-counter-argument: But there was one year when the quarterback was injured, and new players that year also performed better than usual.

  Causal framework: There are no unmeasured confounders; we can obtain the causal effect by controlling for the known confounder.

Introduction **Potential outcome framework**
Relaxing SRA Randomization
Simulation No unmeasured confounders (Rubin et al. 80s)
Closing remarks Sequential Randomization Assumption (Robins '80s/'90s)

We postulate "potential outcomes," random variables indexed by
treatment level

$$Y_a(\omega), a \in \mathcal{A}$$

interpreted as the response of a unit ω if, possibly contrary to fact,
treatment level a were applied to ω and related to the observed
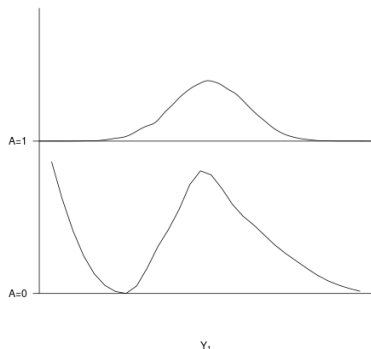data by the consistency axiom

$$Y = Y_A = Y_a\big|_{a=A}$$

Introduction   **Potential outcome framework**
Relaxing SRA   Randomization
Simulation   No unmeasured confounders (Rubin et al. 80s)
Closing remarks   Sequential Randomization Assumption (Robins '80s/'90s)

A "causal effect" can then be stated/defined in terms of the potential outcomes, e.g.,

$$\mathbb{E}(Y_1 - Y_0) \qquad \text{"Average Treatment Effect"}$$

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

Can't say much about Y_1 only knowing $Y_1 \mathbb{1}\{A = 1\}$.

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

We need to be able to say
something about $Y_1 \mathbb{1}\{A = 0\}$
based off of $Y_1 \mathbb{1}\{A = 0\}$
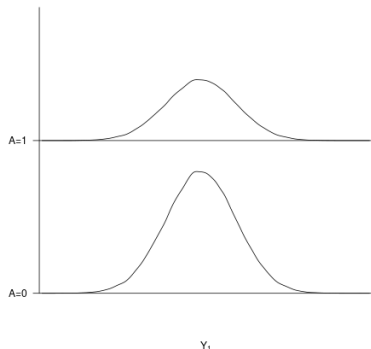
randomization with respect to treatment, $Y_a \perp\!\!\!\perp A$
recover Y_1 by dividing $Y \mathbb{1}\{A = 1\}$ by $P(A = 1)$
in general, $\mathbb{E}(g(Y_a)) = \mathbb{E}\left(\frac{g(Y \mathbb{1}\{A = a\})}{f(A)}\right)$

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

"No Unmeasured Confounders":
randomization holds conditional
on some covariate L. Compute

$Y_1 \mid L = l$, the potential outcome
at each level l, now dividing by
the conditional treatment
probability ("propensity").
Then integrate over L.

$$\mathbb{E}(g(Y_a)) = \mathbb{E}\left(\frac{\mathbb{1}\{A = a\}g(Y)}{f(A \mid L)}\right)$$

Y_1

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)
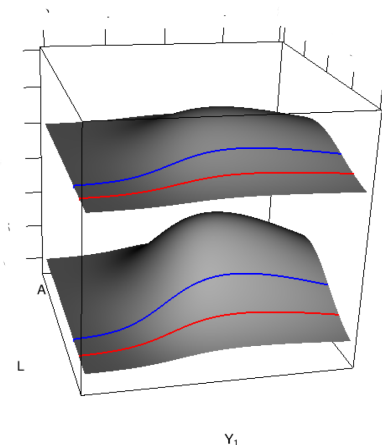
▶ (clones/copies interpretation) At any covariate level $L = l$, if say, $P(A = 1 \mid L = l) = 1/4$, then there are 3 unobserved units for every observed unit and (no unmeasured confounders) these are homogenous as to Y_1

▶ (likelihood perspective)

$$\mathbb{P}(Y = y, A = a, L = l)$$
$$= \mathbb{P}(Y = y \mid A = a, L = l)\mathbb{P}(A = a \mid L = l)\mathbb{P}(L = l)$$
$$= \mathbb{P}(Y_a = y \mid A = a, L = l)\mathbb{P}(A = a \mid L = l)\mathbb{P}(L = l) \text{ (consistency)}$$
$$= \mathbb{P}(Y_a = y \mid L = l)\mathbb{P}(A = a \mid L = l)\mathbb{P}(L = l) \text{ (NUC)}$$

$$\mathbb{P}(Y_a = y, A = a, L = l) / \mathbb{P}(A = a \mid L = l) \text{ (positivity)}$$
$$= \mathbb{P}(Y_a = y, L = l)$$

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

Longitudinal setting: estimating the effect of a treatment regime in the presence of time-varying confounding.

Central example: Estimating time to progression to AIDS under HAART: Physician bases treatment on the patient's CD4 count, that treatment affects CD4 count, which in turn informs a subsequent treatment decision. And CD4 count is prognostic of the outcome.

Introduction	Potential outcome framework
Relaxing SRA	Randomization
Simulation	No unmeasured confounders (Rubin et al. 80s)
Closing remarks	**Sequential Randomization Assumption (Robins '80s/'90s)**
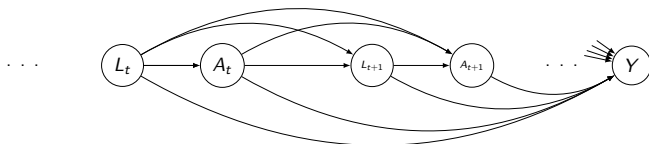
- ▶ T time points $t = 1, \ldots, T$
- ▶ Treatment regime $\overline{A} = (A_1, \ldots, A_T) \in \mathcal{A}^T$ discrete-valued
- ▶ Potential outcomes $\{Y_{\overline{a}}\}$ indexed by fixed treatment regimes $\overline{a} \in \mathcal{A}^T$
- ▶ Observed outcome $Y = Y_{\overline{A}} = \sum_{\overline{a}} \mathbb{1}\{\overline{A} = \overline{a}\} Y_{\overline{a}}$ (consistency)
- ▶ Covariates $\overline{L} = (L_1, \ldots, L_T)$

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

similar targets such as

$$\beta = \mathbb{E}(Y_{\overline{a}}) - \mathbb{E}(Y_{\overline{0}})$$

or

$$\beta_1 : \mathbb{E}(Y_{\overline{a}}) = \beta_0 + \beta_1 \sum_t a_t$$

("marginal structural mean models")

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

Consider using a regression model to estimate β, e.g.:

$$\mathbb{E}(Y \mid \overline{A} = \overline{a}) = \mathbb{E}(Y_{\overline{a}} \mid \overline{A} = \overline{a}) = b_0 + b_1 \sum_t a_t$$

L_{T-1} is a confounder of the subsequent treatment and the outcome, and so should be accounted for, e.g.,

$$\mathbb{E}(Y \mid \overline{A}, L_{T-1}) = b_0 + \gamma_1 L_{T-1} + b_1 \sum_t A_t$$

on the other hand, controlling for L can block the effect of earlier treatment.

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)
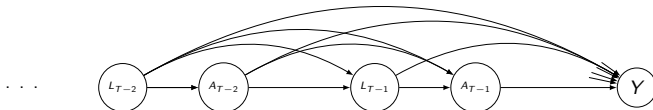
The longitudinal generalization of "no unmeasured confounders" is

▶ "Sequential randomization assumption": for all $\overline{a}$ and t,

$$Y_{\overline{a}} \perp\!\!\!\perp A_t \mid A_1, \ldots, A_{t-1}, L_1, \ldots, L_t$$

Propensity score weights generalize to

$$W_{SRA} = \Pi_{t=1}^{T} f_{A_t \mid \overline{A}_{t-1}, \overline{L}_t}(A_t \mid \overline{A}_{t-1}, \overline{L}_t)$$

SRA will hold when all factors prognostic of Y used by the physicians to determine whether treatment A is given at t are recorded in $\overline{A}_{t-1}, \overline{L}_t$

Introduction
Relaxing SRA
Simulation
Closing remarks

Potential outcome framework
Randomization
No unmeasured confounders (Rubin et al. 80s)
Sequential Randomization Assumption (Robins '80s/'90s)

A marginal structural mean model ("MSMM") is a model on the marginal mean of the potential outcomes,

$$\mathbb{E}(Y_{\overline{a}}) = \mu_\beta(\overline{a})$$

Besides SRA, also assume $0 < \mathbb{P}(A_t = a \mid \overline{L}_{t-1}, \overline{A}_{t-1}) < 1$ when the conditioning event has positive probability

Then (Robins '98)

$$\mathbb{E}((Y - \mu_\beta(\overline{A}))/W_{SRA}) = 0.$$

► $\hat{\beta}$ asymptotically normal (usual regularity conditions)
► Standard software routines can be used, as long as they allow observations to be weighted
► similar change of measure interpretation as NUC theory

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Outline

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Suppose there is some unobserved confounder U, which we would need to have observed in order for "SRA" to hold:

$$Y_{\overline{a}} \perp\!\!\!\perp A_t \mid \overline{A}_{t-1}, \overline{L}_{t-1}, \overline{U}_{t-1} \quad \text{for all } t, \overline{a}$$
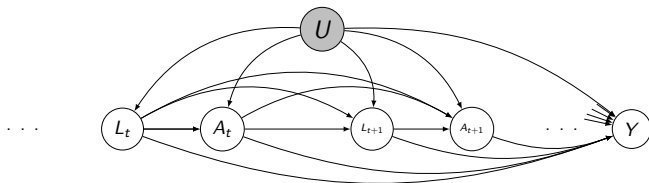
Can we still identify/estimate the causal parameter?

Introduction **Instrumental variables**
Relaxing SRA Main result
Simulation Estimation in practice
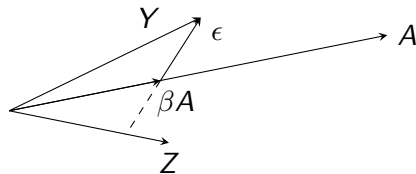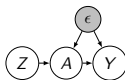Closing remarks Remarks on assumptions

Informally, an IV is a random variable associated with covariates, but orthogonal to the unobserved confounder.

A typical application is OLS with "endogenous error"

$$Y = \beta A + \epsilon$$

Consistency of OLS generally requires ϵ be uncorrelated with A
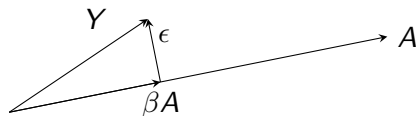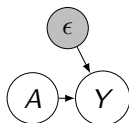
Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Informally, an IV is a random variable associated with covariates, but orthogonal to the unobserved confounder.

A typical application is OLS with "endogenous error"

$$Y = \beta A + \epsilon$$

If in fact the error is correlated with the covariates, OLS is biased
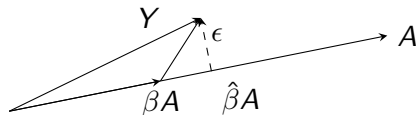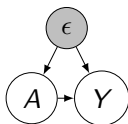
Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Informally, an IV is a random variable associated with covariates, but orthogonal to the unobserved confounder.

A typical application is OLS with "endogenous error"

$$Y = \beta A + \epsilon$$

Suppose we have a random variable Z orthogonal to ϵ, but not to A.

Introduction
Relaxing SRA
Simulation
Closing remarks
Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Examples of instrumental variables:

- assignment to treatment
- physician preference
- draft status
- distance to school/hospital
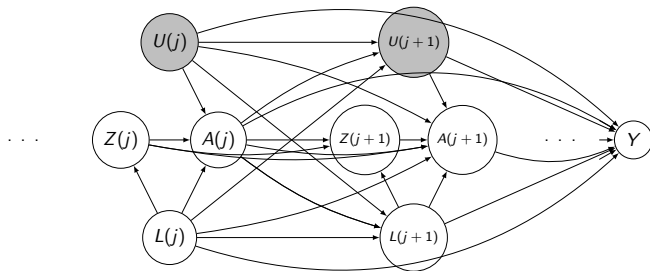
Assumptions

1. $Y_{\overline{a}} \perp\!\!\!\perp A_t \mid \overline{A}_{t-1}, \overline{L}_{t-1}, \overline{U}_{t-1}$ for all $t, \overline{a}$

2. IV assumptions

   2.1 $Y_{\overline{az}} = Y_{\overline{a}}$ a.s. "exclusion restriction"
   2.2 $Z_t \perp\!\!\!\perp \overline{U} \mid \overline{A}_{t-1}, \overline{Z}_{t-1}, \overline{L}_t$ "IV independence A"
   2.3 $\overline{Z} \perp\!\!\!\perp Y_{\overline{a}} \mid \overline{A}, \overline{L}$ "IV independence B"

3. and finally an assumption specific to our problem

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

3. An assumption specific to our problem, either of:

   3.1 Independent Compliance Type:

   $$\mathbb{E}\left[A_t | \overline{U}_t, \overline{L}_t, \overline{A}_{t-1}, \overline{Z}_{t-1}, Z_t = 1\right] - \mathbb{E}\left[A_t | \overline{U}_t, \overline{L}_t, \overline{A}_{t-1}, \overline{Z}_{t-1}, Z_t = 0\right]$$
   $$= \Delta_t\left(\overline{L}_t, \overline{A}_{t-1}, \overline{Z}_{t-1}\right)$$

   or

   3.2 Independent Causal Effect (binary treatment only):

   $$Y_{(a_t=1, a_{t+1}, \ldots, a_T)} - Y_{(a_t=0, a_{t+1}, \ldots, a_T)} \perp\!\!\!\perp \overline{U}_t \mid \overline{L}_t, \overline{A}_{t-1}, \overline{Z}_{t-1}$$

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Weighted Estimating Equation
Define weights by

$$\overline{W} = \prod_{t=1}^{T} (-1)^{1-Z_t} \Delta_t \left(\overline{L}_t, \overline{A}_{t-1}, \overline{Z}_{t-1} \right) f_{Z_t}(Z_t \mid \overline{A}_{t-1}, \overline{Z}_{t-1}\overline{L}_t).$$

Let h denote a vector-valued function of $\overline{A}$ of the same dimension
as β. Under the above assumptions,

$$\mathbb{E}\left(h(\overline{A})(Y - \mu_\beta(\overline{A}))/\overline{W}\right) = \sum_{\overline{a}} h(\overline{a}) \left(\mathbb{E}(Y_{\overline{a}}) - \mu_\beta(\overline{a})\right) (-1)^{T-\sum_j a_j} = 0$$

where the summation is taken over all tuples $\overline{a} \in \{0,1\}^{T-1}$

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

$f_{Z_t | \overline{A}_{t-1}, \overline{Z}_{t-1}, \overline{L}_t}$ and $\Delta_t(\overline{A}_{t-1}, \overline{Z}_{t-1}, \overline{L}_t)$ require modeling/estimation

bootstrap or sandwich variance for inference

weight stabilization analogous to SRA theory

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

3.1 "independent compliance type"

Interpretation: Z_j assignment to treatment or control A_j the treatment actually received assumption is that the difference in proportions of compliance is accounted for by the observed data .

		$A_{Z=0}$	
		0	1
$A_{Z=1}$	0	never-taker	defier
	1	complier	always-taker

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

3.2 "independent causal effect"

assume A is binary. The assumption implies you could obtain the ATE at a time point, i.e., β_1 in the model $E(Y_{(\bar{a}_{j-1}, a_j)}) = \beta_0 + \beta_1 a_j$, using non-IV methods. The theorem allows you to go from here to estimating the causal parameter in an arbitrary mean model.

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Special case

- $T = 1$ time point
- A, Z binary
- $\Delta_j \left(\overline{L}_j, \overline{A}_{j-1}, \overline{Z}_{j-1} \right) = \Delta_j(\overline{A}_{j-1})$
- $f_{Z_j | \overline{A}_{j-1}, \overline{Z}_{j-1}, \overline{L}_j}$ constant

Consider the saturated model

$$\mathbb{E}(Y_a) = \beta_0 + \beta_1 a = \mathbb{E}(Y_0) + (\mathbb{E}(Y_1) - \mathbb{E}(Y_0))a$$

The weights are now $(-1)^{1-Z}$, and

$$\hat{\beta}_1 = \frac{\sum y_t \mathbb{1}\{z_t = 1\} - \sum y_t \mathbb{1}\{z_t = 0\}}{\sum a_t \mathbb{1}\{z_t = 1\} - \sum a_t \mathbb{1}\{z_t = 0\}}$$

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions
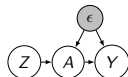
A perspective on the proposed assumptions:

We have an analogue of Robins's g-formula:

$$\mu_\beta(\overline{a}) = \mathbb{E}(Y_{\overline{a}}) =$$
$$\int \mathbb{E}(Y \mid \overline{A} = \overline{a}, \overline{L} = \overline{l}, \overline{U} = \overline{u}) \times$$
$$\prod_j f_{L_j, U_j \mid \overline{A}_{j-1}, \overline{L}_{j-1}, \overline{U}_{j-1}}(l_j, u_j \mid \overline{a}_{j-1}, \overline{l}_{j-1}, \overline{u}_{j-1}) d\nu(l_j, u_j)$$

$$0 = \int \left(\mathbb{E}(Y \mid \overline{A}, \overline{L}, \overline{U}) - \mu_\beta(\overline{A}) \right) \prod_j f_{L_j, U_j \mid \overline{A}_{j-1}, \overline{L}_{j-1}, \overline{U}_{j-1}}(l_j, u_j \mid \overline{a}_{j-1}, \overline{l}_{j-1}, \overline{u}_{j-1}) d\nu(l_j, u_j)$$

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

Decompose the error as

$$Y - \mu_\beta(\overline{A}) = \underbrace{Y - E(Y \mid \overline{A}, \overline{Z}, \overline{L}, \overline{U})}_{\text{exogenous}} + \underbrace{E(Y \mid \overline{A}, \overline{Z}, \overline{L}, \overline{U}) - \mu_\beta(\overline{A})}_{\text{"}\eta\text{", endogenous}}$$

Using the "g-formula" analogue,

$$\eta = \sum_{\overline{a}} \mathbb{1}\{\overline{A} = \overline{a}\} \sum_{t=1}^{T} \phi_t^{(\overline{a})}(\overline{A}_{t-1}, \overline{L}_t) - \mathbb{E}\left(\phi_t^{(\overline{a})}(\overline{A}_{t-1}, \overline{L}_t) \mid \overline{A}_{t-1}, \overline{L}_{t-1}\right)$$

$$= \sum_{\overline{a}} \mathbb{1}\{\overline{A} = \overline{a}\} M^{(\overline{a})}$$

for some $\phi_t^{(\overline{a})}$, a sum of martingales at each level of A

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

We seek functions of the observed data $w(\overline{A}, \overline{L}, \overline{Z})$ to form estimating equations

$$\mathbb{E}(w(\overline{A}, \overline{L}, \overline{Z}) \times (Y - \mu_\beta(\overline{A}))) = \mathbb{E}(w(\overline{A}, \overline{L}, \overline{Z}) \times \eta) = 0$$

We could treat this as an IV problem (η and A are dependent), simply requiring a random variable Z orthogonal to η

Introduction
Relaxing SRA
Simulation
Closing remarks

Instrumental variables
Main result
Estimation in practice
Remarks on assumptions

But we know more about the structure of η.

Under the proposed conditions the quantities $1/\overline{W}$ are orthogonal to η

Outline

"Independent compliance type" assumption holds, "independent causal effect" does not hold
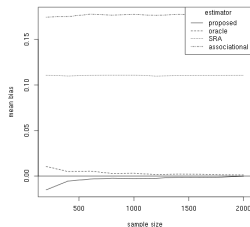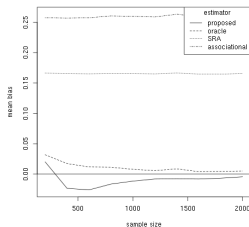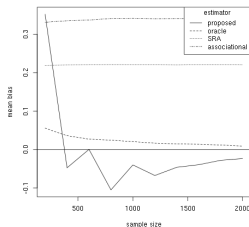
J=2   J=3   J=4

Mean bias versus sample size of the weighted estimator, for J=2, 3, and 4, time points, compared with oracle (weights including observed and unobserved confounders), SRA (weights including observed confounders), and associational (no weighting) estimators.
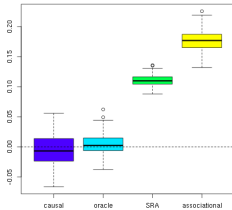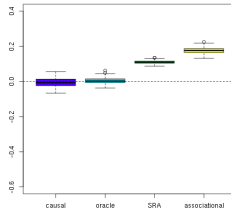
J=2   J=2 (axis rescaled)   J=3
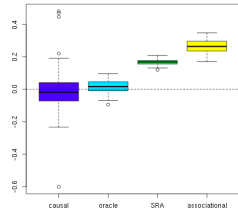
N=500

Outline

target application: SMART trials

see Wharton tech report (Tchetgen Tchetgen, Michael, Cui '18)
for

- identification of the parameters of any marginal structural models, e.g., failure time model or quantile model
- semiparametric efficient, multiply robust estimator partially protects against model misspecification in that the estimator is consistent whenever any one of three sets of nuisance parameters are consistently estimated

References

Haben Michael, Yifan Cui, and Eric J. Tchetgen Tchetgen.

A simple weighted approach for instrumental variable estimation of marginal strucural mean models.

In progress, 2019.

E. J Tchetgen Tchetgen, H. Michael, and Y. Cui.

Marginal Structural Models for Time-varying Endogenous Treatments: A Time-Varying Instrumental Variable Approach.

ArXiv e-prints, September 2018.

James Robins.

Marginal Structural Models.

In 1997 Proceedings of the American Statistical Association, pages 1–10 of 1998 Section on Bayesian Statistical Science, 1997.

Linbo Wang and Eric Tchetgen Tchetgen.

Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80(3):531–550, 2018.