Towards Assessing Time-Varying Causal Effect Moderation in Experimental and Observational Studies

Daniel Almirall, Oral Defense Examination

University of Michigan, Department of Statistics

PhD Dissertation Committee:

Susan Murphy (Chair, UMich),

Stephen Raudenbush (UChicago),

Thomas Ten Have (UPenn),

& Yu Xie (UMich)

April 18, 2007

Contents

1 Effect Moderation in One Time Point

Part I of the dissertation focuses on effect moderation in one-time point as a warm-up, and to get ideas straight in the simplest setting.

The data structure in one time point is (S, A, Y).

Examples: $A =$ treatment, $Y =$ outcome, $S =$ pre-A covariate

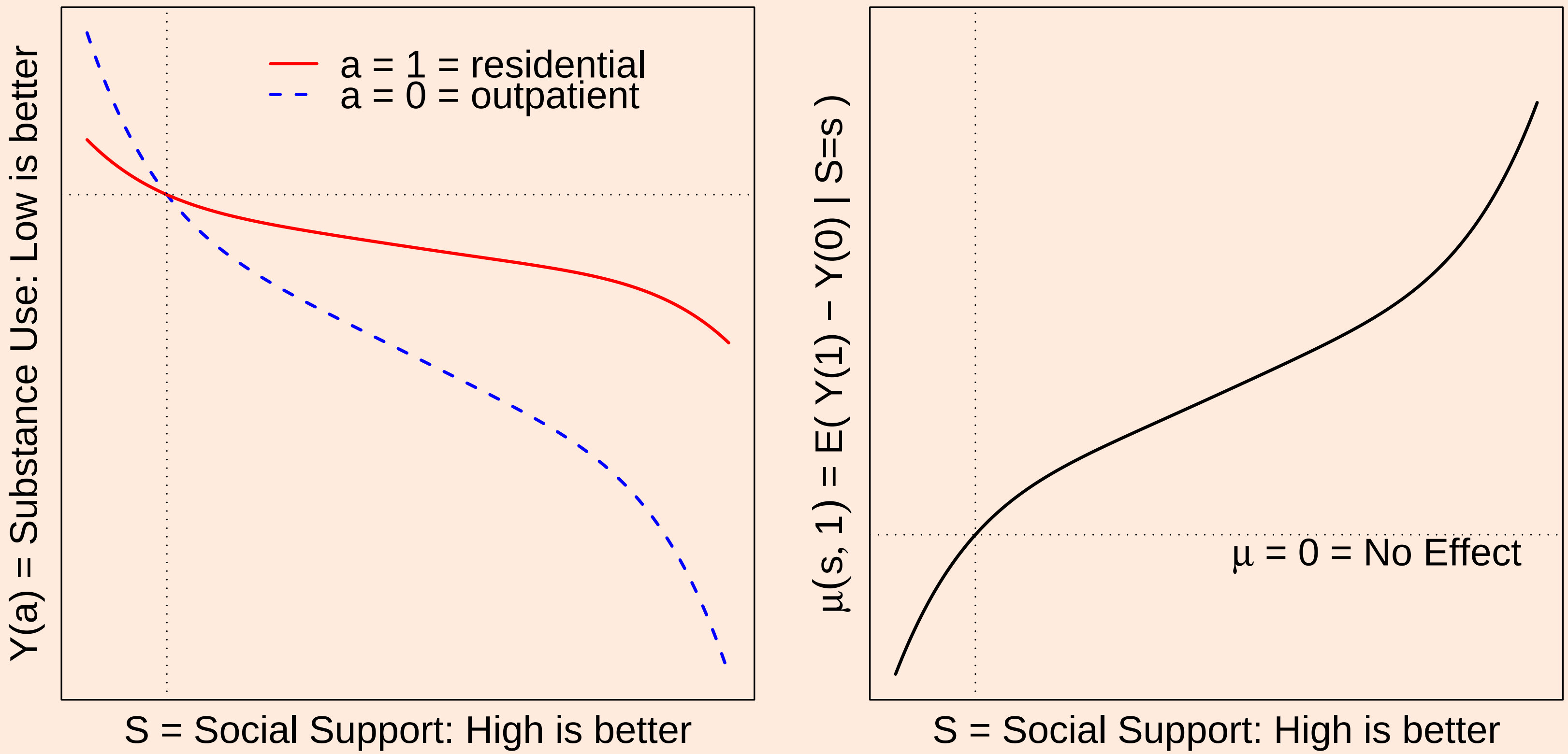
A	Y	S
Medication vs. Placebo	Depression	Suicidal?
SAT Coaching?	SAT Math Score	Gender
Inpatient vs. Outpatient	Substance Abuse	Social Support

Warm-up: Suppose we want the effect of A on Y. Why condition on (adjust for) pre-treatment variables S?

1. **Confounding**: S is correlated with both A and Y. In this case, S is known as a "confounder" of the effect of A on Y.
2. **Precision**: S may be a pre-treatment measure of Y, or any other variable highly correlated with Y.
3. **Missing Data**: The outcome Y is missing for some units, S and A predict missingness, and S is associated with Y.
4. **Effect Heterogeneity**: S may moderate, temper, or specify the effect of A on Y. In this case, S is known as a "moderator" of the effect of A on Y. Formalized in next slide.

Definition & Illustration of Effect Moderation

$$\mu(s,a) \equiv E(Y(a) - Y(0) \mid S = s)$$



Causal Effect Moderation in Context: Relevance?

Theoretical Implication: Understanding the heterogeneity of the effects of causes enhances our understanding of various (or competing) scientific theories; and it may suggest new scientific (etiologic) hypotheses to be tested.

Practical Implication: Identifying types, or subgroups, of individuals for which treatment is not effective may suggest altering the treatment to suit the needs of that particular type of individual.

An Elaboration of Yu Xie's Variability Principle: We really want $Y_i(a) - Y_i(0) \; \forall \; i$. We settle for "groupings" of effects (here, groupings by S); $\mu(s,a)$ "comes closer" than $E(Y(a) - Y(0))$.

2 Mean Model in One Time Point

Decompose the conditional mean $E(Y(a) \mid S)$ as follows:

$$\begin{aligned} E(Y(a) \mid S = s) = {} & E(Y(0) \mid S = 0) \\ & + \Big(E(Y(0) \mid S = s) - E(Y(0) \mid S = 0)\Big) \\ & + E(Y(a) - Y(0) \mid S = s) \\ = {} & \eta_0 + \phi(s) + \mu(s, a). \end{aligned}$$

The intercept η_0 and the function $\phi(s)$ are non-causal. They are known as nuisance functions. $\phi(s)$ is the "associational main effect" of S on $Y(0)$.

Prototypical Linear Parametric Model

We use β for our causal parameters of interest:

$$\begin{aligned} E(Y(a) \mid S) &= \eta_0 + \phi(S) + \mu(S, a; \beta) \\ &= \eta_0 + \phi(S) + aH\beta \end{aligned}$$

where H is a function of S.

Sometimes we parameterize $\phi(S)$ using $\phi(S; \eta_{-0}) = G\eta_{-0}$, where G is a function of S.

Example: Let $G = (S)$ and $H = (1, S)$:

$$E(Y(a) \mid S = s) = \eta_0 + \eta_1 s + a \times (\beta_1 + \beta_2 s).$$

If a and S are binary, then this is the fully saturated model.

3 Estimation in One Time Point

The dissertation discusses three estimators for β in $\mu(S, a; \beta)$:

1. Traditional Regression
2. Semi-parametric Estimation Method: Robins' E-Estimator
3. Inverse Probability of Treatment Weighted (IPTW) Regression

We discuss these (and more) in turn, supposing that

1. a is binary (0,1), and
2. True model for $\mu(s, a)$ is $\mu(S, a; \beta) = aH\beta$ for some H.
   Example: $H = (1, S) \Rightarrow aH\beta = a(\beta_1 + \beta_2 s)$.

An important consideration in estimation is how A comes about.

Traditional Ordinary Least Squares Regression

Recall true model: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$.

Useful when S is sole confounder, and have good model for $\phi(s)$.

Requires model for nuisance function: $\phi(S; \eta_{-0}) = G\eta_{-0}$.

Regress $Y \sim [1, G, A \times H]$ to get $(\widehat{\eta}, \widehat{\beta})$. The $\widehat{\beta}$ estimates solve

$$0 = \mathbb{P}_n\left(\left(Y - \eta_0 - G\eta_{-0} - AH\beta\right)AH^T\right).$$

$\widehat{\beta}$ unbiased for β if $\phi(S; \eta_{-0}) = G\eta_{-0}$ is true model for $\phi(s)$ and $A \perp \{Y(0), Y(1)\}$ given S.

Semi-parametric E-Estimator

Recall true model: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$.

Useful when S is sole confounder, but we have no model for $\phi(s)$.

Does NOT require model for nuisance function $\phi(s)$.

Get $\widehat{\beta}$ by solving the following estimating equations

$$0 = \mathbb{P}_n\Bigg(\Big(Y - \widehat{b}(S;\xi) - AH\beta\Big)\Big(A - \widehat{p}(S;\alpha)\Big)H^T\Bigg),$$

where $\widehat{b}(S;\xi)$ is a guess for $E(Y - AH\beta \mid S) = \eta_0 + \phi(S)$.

$\widehat{\beta}$ unbiased for β if $p(S;\alpha)$ is true model for $Pr(A = 1 \mid S)$, and $A \perp \{Y(0), Y(1)\}$ given S. (Discuss double-robustness.)

IPT Weighted Regression (WLS)

Recall true model: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$.

Useful when we have measured confounders V ($\supset S$).

Requires model for nuisance function: $\phi(S; \eta_{-0}) = G\eta_{-0}$.

Regress $Y \overset{\widehat{w}}{\sim} [1, G, A \times H]$ to get $(\widehat{\eta}, \widehat{\beta})$, where weights are

$$w(V, A) = A \times \frac{Pr(A = 1 \mid S)}{Pr(A = 1 \mid V)} + (1 - A) \times \frac{Pr(A = 0 \mid S)}{Pr(A = 0 \mid V)}.$$

$\widehat{\beta}$ unbiased for β if $\phi(S; \eta_{-0}) = G\eta_{-0}$ is true model for $\phi(s)$, and $A \perp \{Y(0), Y(1)\}$ given V.

An Overview of Estimation Strategies

Model A: $E(Y(a) \mid S) = \eta_0 + \phi(S) + aH\beta$

Model B: $E(Y(a) \mid V) = \phi_0 + \phi(V) + aH\beta$

H is always a function of S Ex: $H = (1, S)$

	Model A			**Model B**
	No Confnders	S is Sole Confnder	Confnders $V^\sharp$	Modrtrs S, Confndrs V
ϕ Is Known	OLS*	OLS*	IPTW Regression*	OLS
ϕ Is Not Known	OLS (if $S \perp A$)	E-estimtr*†	IPTW E-estimtr*†	E-estimtr$^\sharp$

*just discussed †need $Pr(A = 1 \mid S)$ $^\sharp$need $Pr(A = 1 \mid V)$

4 Application in One Time Point

$n = 1984$ adolescents that are substance abusers.

Motivation: American Society of Addiction Medicine (ASAM) Patient Placement Criteria (PPC)

Two levels of care (LOC):

$A = 0$ outpatient,

$A = 1$ residential

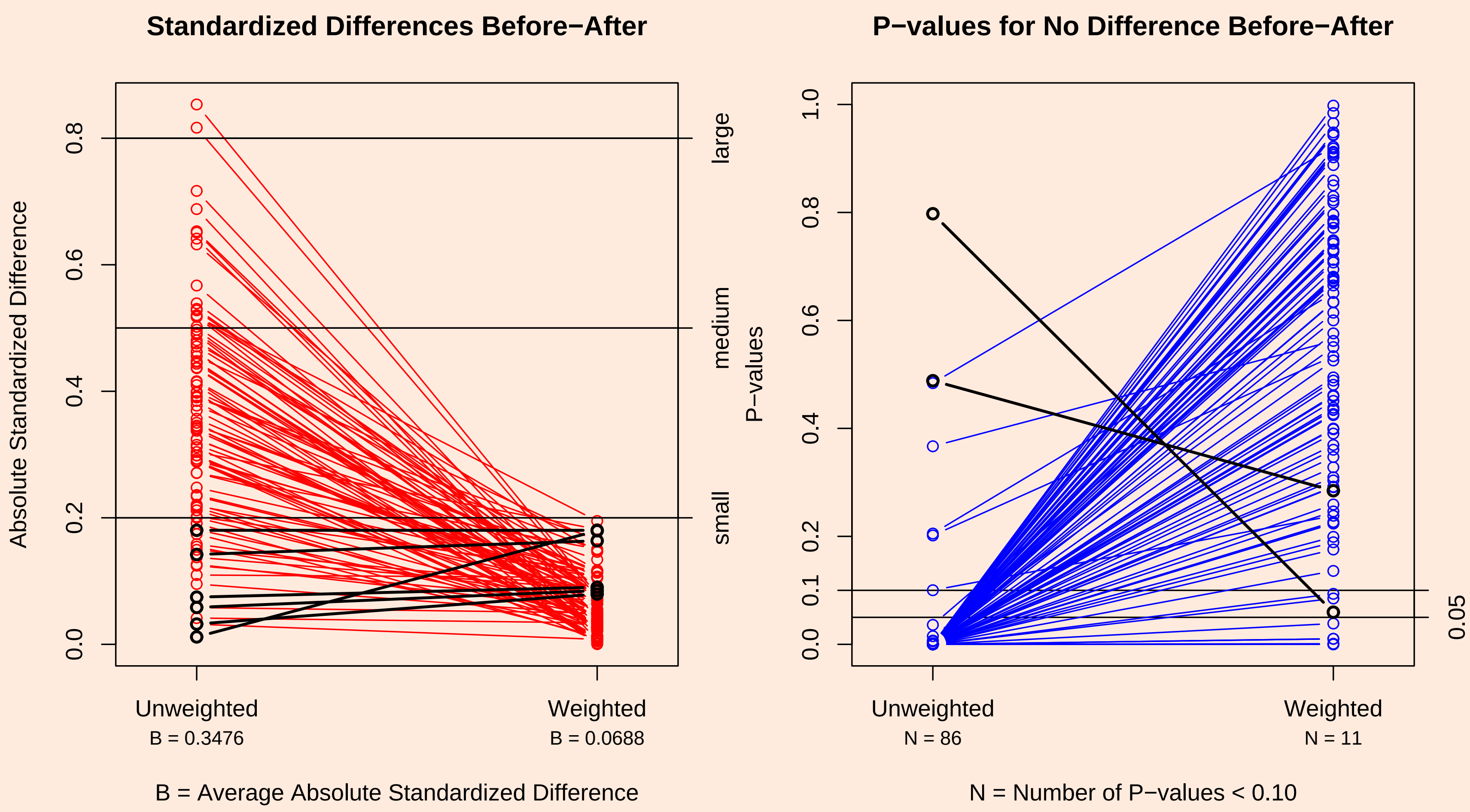
Illustrate methodology using:

$S =$ Needle Frequency Index (hi is bad)

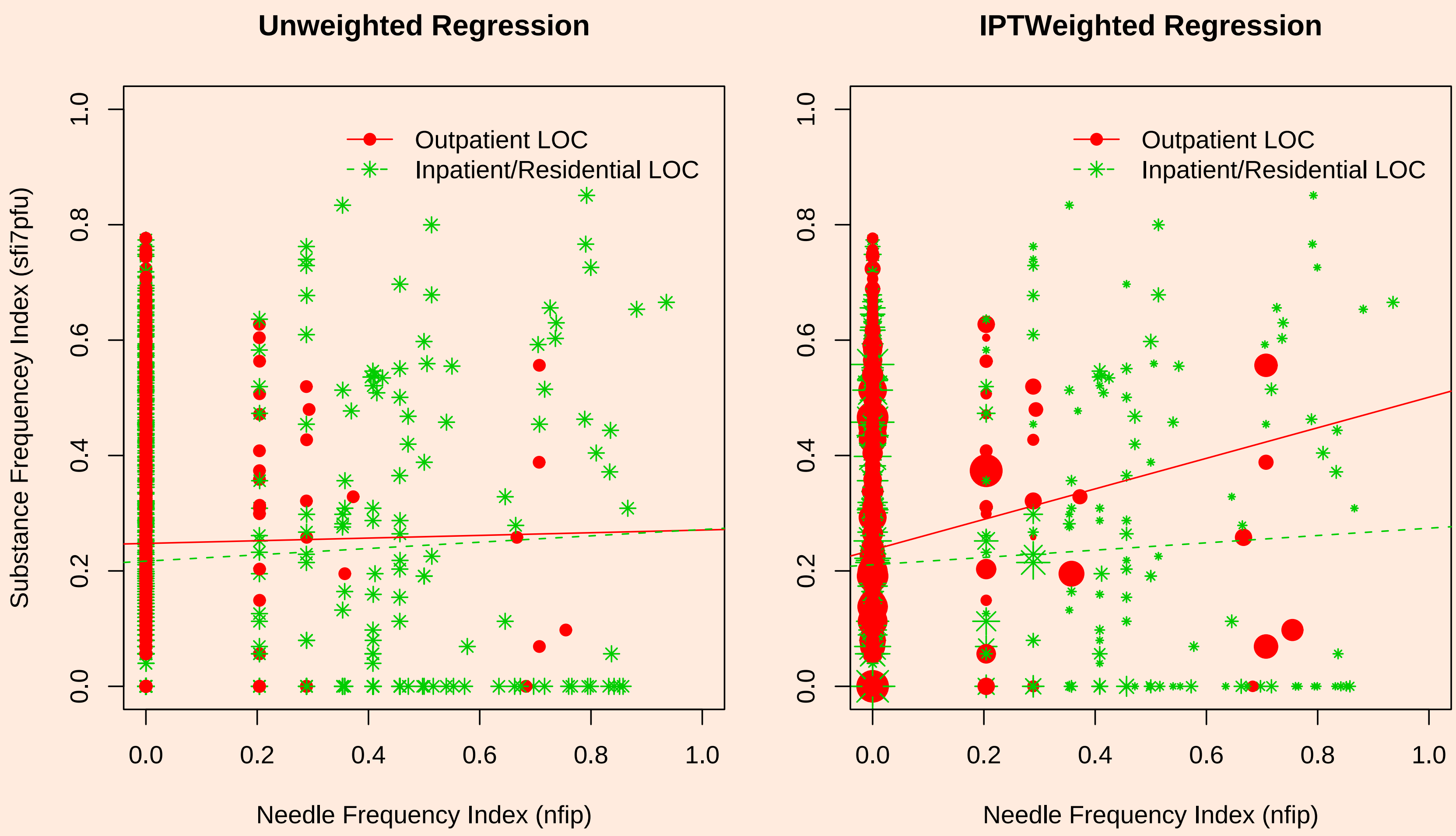
$Y =$ Substance Frequency Index (hi is bad)

$V =$ 86 covariates to adjust for (possible confounders)

Covariate Balance Before-After Weighting



Effect Moderation by S = Needle Frequency Index



5 Time-Varying Effect Moderation

Part II of the dissertation focuses on time-varying effect moderation.

The data structure in the time-varying setting is $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

Running Example, from the PROSPECT Study:

(a_1, a_2)	Time-varying treatment pattern; a_t is binary (0,1)
$Y(a_1, a_2)$	Depression at the end of the study; continuous
S_1	Suicidal Ideation at baseline visit; continuous
$S_2(a_1)$	Suicidal Ideation at second visit; continuous

As a Decomposition of the Marginal Causal Effect

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

Consider the following arithmetic decomposition of the causal effect of (a_1, a_2) on Y, using the covariates $\bar{S}_2(a_1)$:

$$E\big[Y(a_1,a_2) - Y(0,0)\big] = E\Big[E\big[Y(a_1,a_2) - Y(a_1,0) \mid \bar{S}_2(a_1)\big]\Big] + E\Big[E\big[Y(a_1,0) - Y(0,0) \mid S_1\big]\Big].$$

The inner expectations represent the **conditional intermediate causal effects** μ_1 and μ_2, respectively.

Formal Definition of Time-Varying Causal Effects

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

Conditional Intermediate Causal Effect at $t = 1$:

$$\begin{aligned}\mu_{\mathbf{1}}(\mathbf{s_1}, \mathbf{a_1}) &\equiv E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1] \\ &= a_1 \times H_1\beta_1 \qquad \text{(Linear Parameterization)} \\ &\overset{\text{say}}{=} a_1 \times (\beta_1 + \beta_2 s_1) \qquad \text{(Sample Model)}\end{aligned}$$

Conditional Intermediate Causal Effect at $t = 2$:

$$\begin{aligned}\mu_{\mathbf{2}}(\bar{\mathbf{s}}_{\mathbf{2}}, \bar{\mathbf{a}}_{\mathbf{2}}) &\equiv E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2] \\ &= a_2 \times H_2\beta_2 \qquad \text{(Linear Parameterization)} \\ &\overset{\text{say}}{=} a_2 \times (\beta_3 + \beta_4(s_1 + s_2)/2) \qquad \text{(Sample Model)}\end{aligned}$$

6 Robins' Structural Nested Mean Model

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

The SNMM for the conditional mean of $Y(a_1, a_2)$ given $\bar{S}_2(a_1)$ is:

$$
\begin{aligned}
& E\big[Y(a_1, a_2) \mid \bar{S}_2(a_1)\big] \\
&\qquad = E[Y(0,0)] + \Big\{E[Y(0,0) \mid S_1] - E[Y(0,0)]\Big\} \\
&\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y}(\mathbf{a_1}, \mathbf{0}) - \mathbf{Y}(\mathbf{0}, \mathbf{0}) \mid \mathbf{S_1}\big]\Big\} \\
&\qquad\qquad + \Big\{E[Y(a_1, 0) \mid \bar{S}_2(a_1)] - E[Y(a_1, 0) \mid S_1]\Big\} \\
&\qquad\qquad\qquad + \Big\{\mathbf{E}\big[\mathbf{Y}(\mathbf{a_1}, \mathbf{a_2}) - \mathbf{Y}(\mathbf{a_1}, \mathbf{0}) \mid \bar{\mathbf{S}}_\mathbf{2}(\mathbf{a_1})\big]\Big\}
\end{aligned}
$$

Identifying the Causal and Nuisance Functionals

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

The decomposition can be written more succinctly as:

$$E\big[Y(a_1, a_2) \mid \bar{S}_2(a_1) = \bar{s}_2\big] = \mu_0 + \epsilon_1(s_1) + \boldsymbol{\mu_1(s_1, a_1)}$$
$$+ \epsilon_2(\bar{s}_2, a_1) + \boldsymbol{\mu_2(\bar{s}_2, \bar{a}_2)},$$ where

- $\epsilon_2(\bar{s}_2, a_1) = E[Y(a_1, 0) \mid \bar{S}_2(a_1) = \bar{s}_2] - E[Y(a_1, 0) \mid S_1 = s_1]$,
- $\epsilon_1(s_1) = E[Y(0,0) \mid S_1 = s_1] - E[Y(0,0)]$,
- $\mu_2(\bar{s}_2, a_2, 0) = 0$ and $\mu_1(s_1, 0) = 0$,
- $E_{S_2|S_1}[\epsilon_2(\bar{s}_2, a_1) \mid S_1 = s_1] = 0$, and $E_{S_1}[\epsilon_1(s_1)] = 0$.

7 Estimation in Time-Varying Setting

Recall that parametric models for our causal estimands μ_1 and μ_2 are based on the set of parameters $\beta = (\beta_1', \beta_2')'$.

The dissertation considers two estimators for β:

1. Proposed 2-Stage Regression Estimator
2. Robins' Semi-parametric G-Estimator
   - Relies on 2-Stage Estimator for starting values

Both estimators rely on Robins' **Consistency** and **Sequential Ignorability** assumptions. The 2-Stage Estimator relies more on modelling assumptions. We discuss these in turn, but first ...

What's wrong with the Traditional Estimator?

An **Example** of The **Traditional Estimator**: Apply OLS with

$$E(Y \mid \bar{S}_2 = \bar{s}_2, \bar{A}_2 = \bar{a}_2) = \beta_0^* + \eta_1 s_1 + a_1 \times (\beta_1^* + \beta_2^* s_1) \\ + \eta_2 s_2 + a_2 \times (\beta_3^* + \beta_4^*(s_1 + s_2)/2)$$

- Possibly incorrectly specified nuisance functions.

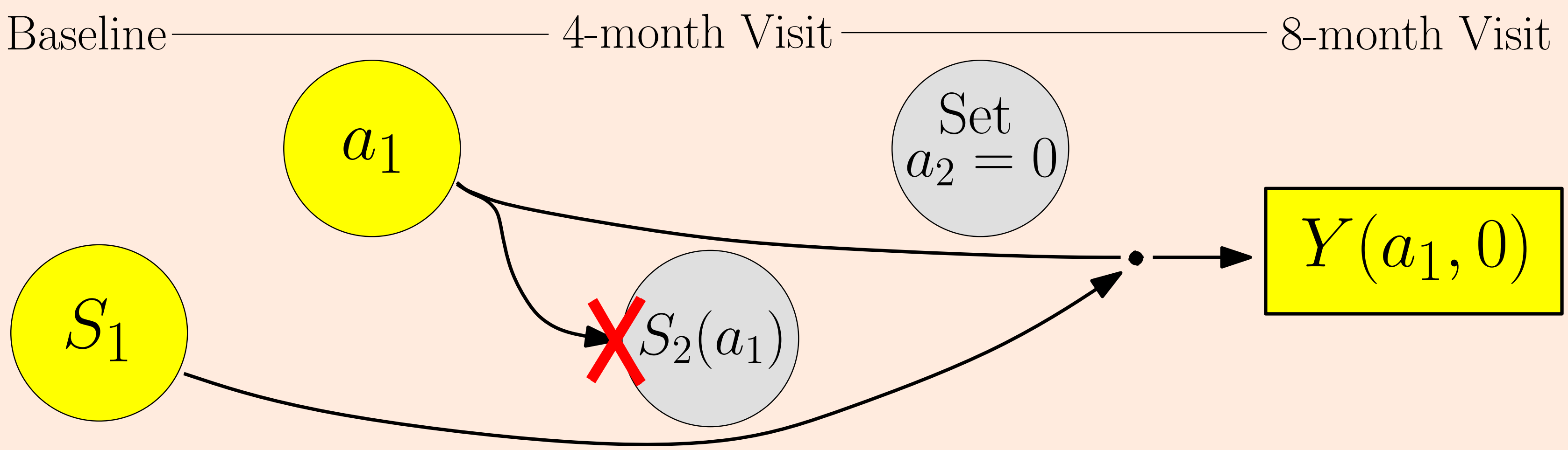
- Two problems arise when using the traditional regression estimator.

- These problems occur even in the absence of time-varying confounders (that is, even under Sequential Ignorability).

First problem with the Traditional Estimator

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

Wrong Effect

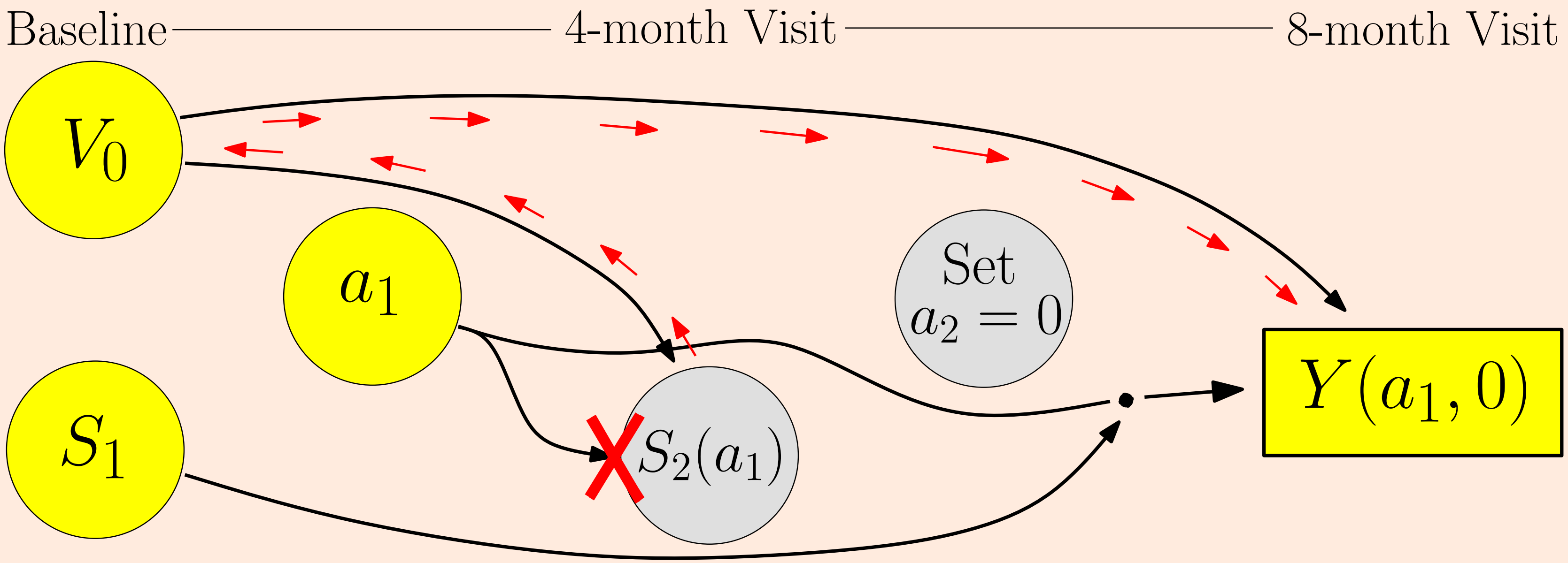


What about the effect transmitted through $S_2(a_1)$?

Second problem with the Traditional Estimator

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

Spurious Effect



Berkson's paradox; Judea Pearl's backdoor criterion

Parameterizing the Nuisance Functions

So we must parameterize the nuisance functions correctly.

Recall the constraints on the nuisance functions:

- $\epsilon_2(\bar{s}_2, a_1) = E[Y(a_1, 0) \mid \bar{S}_2(a_1) = \bar{s}_2] - E[Y(a_1, 0) \mid S_1 = s_1]$,
- $\epsilon_1(s_1) = E[Y(0, 0) \mid S_1 = s_1] - E[Y(0, 0)]$,
- $E_{S_2|S_1}[\epsilon_2(\bar{s}_2, a_1) \mid S_1 = s_1] = 0$, and $E_{S_1}[\epsilon_1(s_1)] = 0$.

Example parameterizations for the nuisance functions:

$$\epsilon_1(s_1) \overset{\text{say}}{=} \eta_{1,1}\big(s_1 - E(S_1)\big)$$
$$\epsilon_2(\bar{s}_2, a_1) \overset{\text{say}}{=} \big(\eta_{2,1} + \eta_{2,2} s_1\big)\big(s_2 - E(S_2(a_1) \mid S_1 = s_1)\big)$$

Proposed 2-Stage Regression Estimator

Recall that $E\big[Y(a_1, a_2) \mid \bar{S}_2(a_1) = \bar{s}_2\big] = \mu_2(\bar{s}_2, \bar{a}_2; \beta_2)$ $+ \epsilon_2(\bar{s}_2, a_1; \eta_2, \gamma_2) + \mu_1(s_1, a_1; \beta_1) + \epsilon_1(s_1; \eta_1, \gamma_1) + \mu_0.$

1. We have models for the μ's: $A_1 H_1 \beta_1$ and $A_2 H_2 \beta_2$; Set aside
2. Model $m_1(\gamma_1) = E(S_1)$, estimate γ_1 with GLM; model $m_2(s_1, a_1; \gamma_2) = E(S_2(a_1) \mid S_1 = s_1)$, estimate γ_2 with GLM
3. Construct residuals $\hat{\delta}_1 = s_1 - \hat{m}_1(\hat{\gamma}_1)$ and $\hat{\delta}_2 = s_2 - \hat{m}_2(s_1, a_1; \hat{\gamma}_2)$
4. Construct models for ϵ's: $G_1 \hat{\delta}_1 \eta_1 = G_1^* \eta_1$ and $G_2 \hat{\delta}_2 \eta_2 = G_2^* \eta_2$
5. Obtain $\hat{\beta}$ and $\hat{\eta}$ using OLS of $Y \sim [1, G_1^*, A_1 H_1, G_2^*, A_2 H_2]$

Robins' Semi-parametric G-Estimator

Robins' G-Estimator is the solution to these estimating equations:

$$0 = \mathbb{P}_n \Bigg\{ \Big(Y - A_2 H_2 \beta_2 - b_2(\bar{S}_2, A_1)\Big) \times \Big(A_2 - p_2(\bar{S}_2, A_1)\Big) \times \begin{pmatrix} 0 \\ H_2' \end{pmatrix}'$$

$$+ \Big(Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 - b_1(S_1)\Big) \times \Big(A_1 - p_1(S_1)\Big) \times \begin{pmatrix} H_1' \\ \Delta'(H_1) \end{pmatrix}' \Bigg\}$$

$$\Delta(H_1) = E\left[H_2 A_2 \middle| S_1, A_1 = 1\right] - E\left[H_2 A_2 \middle| S_1, A_1 = 0\right]$$

$$b_2(\bar{S}_2, A_1) = E\left[Y - A_2 H_2 \beta_2 \mid \bar{S}_2, A_1\right]$$

$$p_2(\bar{S}_2, A_1) = Pr\left[A_2 = 1 \mid \bar{S}_2, A_1\right]$$

$$b_1(S_1) = E\left[Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 \mid S_1\right]$$

$$p_1(S_1) = Pr\left[A_1 = 1 \mid S_2\right]$$

8 Bias-Variance Trade-off

This discussion assumes true models for the causal effects, the μ_ts:

Robins' G-Estimator is unbiased if either p_t or b_t are correctly specified. So-called *double-robustness* property.

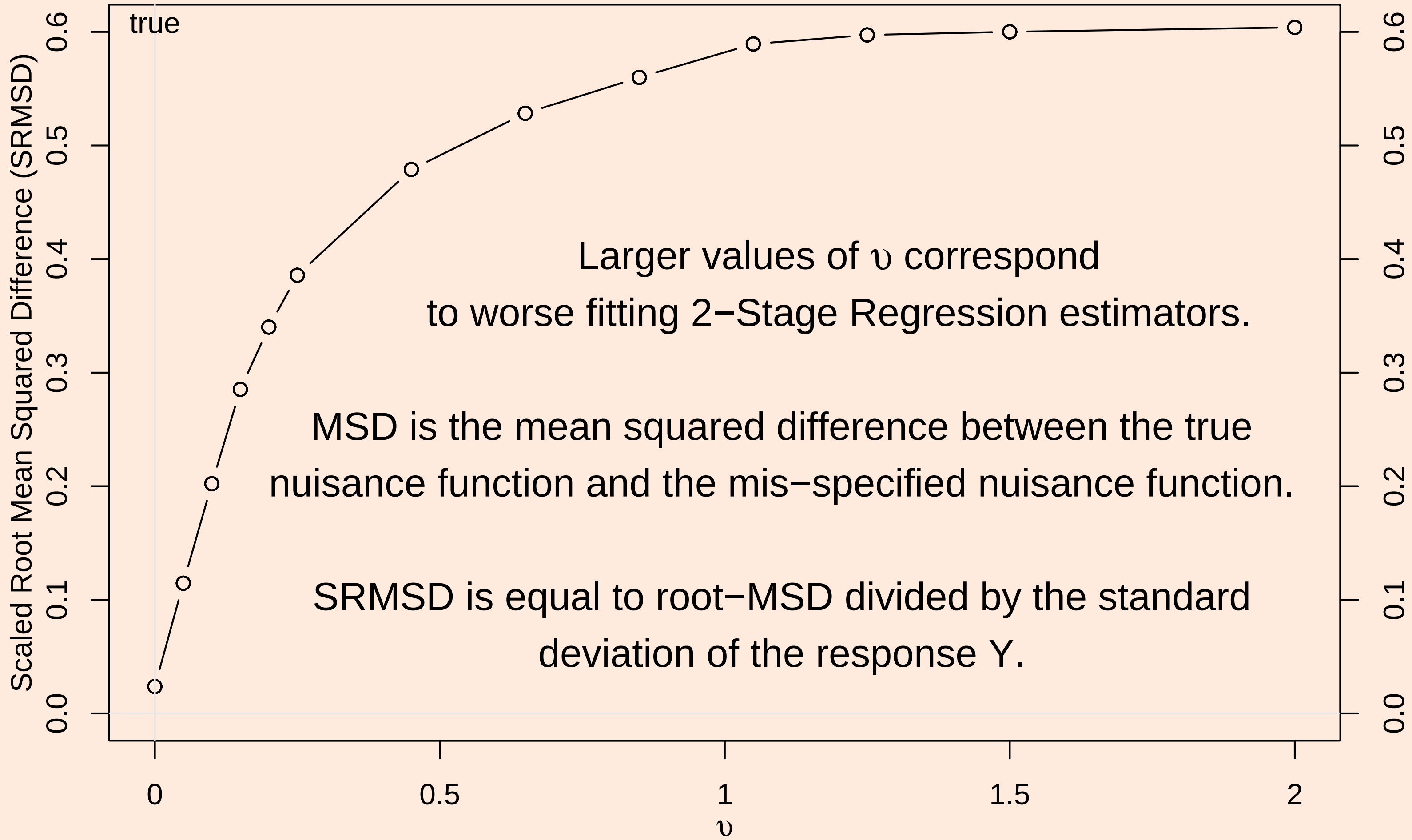
Robins' G-Estimator is semi-parametric efficient if p_t, b_t, and Δ are all correctly specified.

2-Stage Regression Estimator is unbiased only if the nuisance functions are correctly specified.

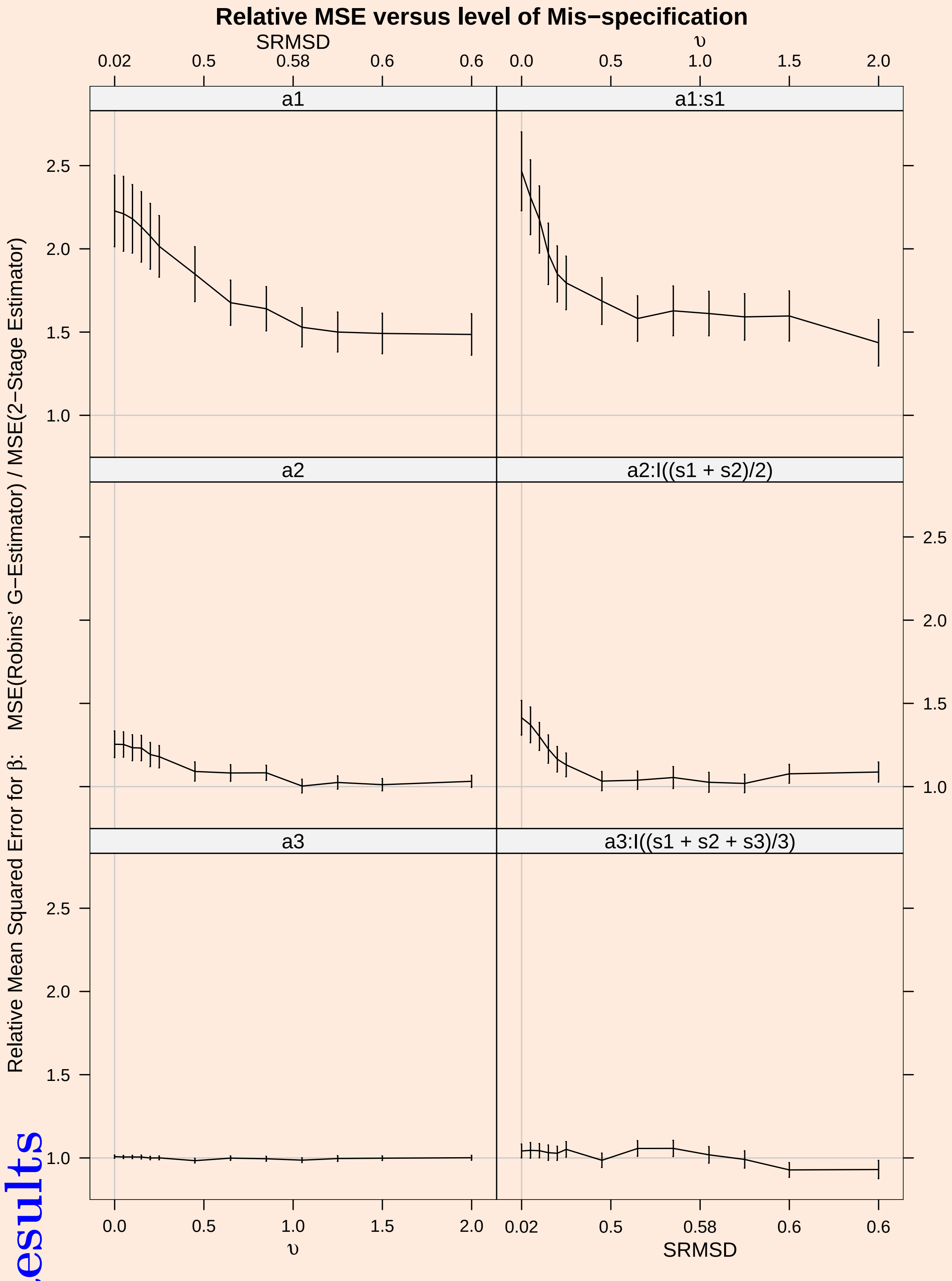
2-Stage Regression Estimator with correctly specified nuisance **is more efficient than** G-Estimator

But what happens as we mis-specify the nuisance functions?

Mis-specifying ϵ_t's using $S^* = S \times N(1, \mathbf{sd} = \nu)$



Results



Relative MSE versus level of Mis−specification
SRMSD
υ
a1
a1:s1
a2
a2:I((s1 + s2)/2)
a3
a3:I((s1 + s2 + s3)/3)
Relative Mean Squared Error for β: MSE(Robins' G−Estimator) / MSE(2−Stage Estimator)
0.02 0.5 0.58 0.6 0.6
0.0 0.5 1.0 1.5 2.0
1.0 1.5 2.0 2.5

9 Illustration in Time-Varying Setting

$n = 543$ geriatric primary care patients from PROSPECT Study

$K = 3$: visits to clinic at baseline, 4-, 8-, and 12-months

Data structure is $\{S_1, A_1, S_2, A_2, S_3, A_3, Y\}$

$S_t =$ suicidal ideation, $A_t =$ adherence, $Y =$ 12-month depression

Adherence is defined as ever meeting with a health specialist

Monotonic adherence pattern: $(0,0,0), (1,0,0), (1,1,0), (1,1,1)$

Assumptions:

- Consistency/SUTVA
- Sequential Ignorability given $\bar{S}_3$ (very likely violated)
- Modeling assumptions

Models Used in the Illustration

Causal effects: (expect $\beta_{t0} < 0$ and $\beta_{t1} > 0$)

1. $\mu_1(S_1, a_1) = a_1\left(\beta_{10} + \beta_{11} S_1\right)$,
2. $\mu_2(\bar{S}_2, \bar{a}_2) = a_2\left(\beta_{20} + \beta_{21}\left(S_1 + S_2\right)/2\right)$, and
3. $\mu_3(\bar{S}_3, \bar{a}_3) = a_3\left(\beta_{30} + \beta_{31}\left(S_1 + S_2 + S_3\right)\right)/3)$.

Nuisance functions:

1. $\epsilon_1(S_1) = \eta_{1,1} \times (S_1 - \gamma_{1,1})$
2. $\epsilon_2(\bar{S}_2, A_1) = (\eta_{2,1} + \eta_{2,2} S_1) \times (S_2 - (\gamma_{2,1} + \gamma_{2,2} S_1 + \gamma_{2,3} A_1 + \gamma_{2,4} S_1 A_1 + \gamma_{2,4} S_1^2))$
3. $\epsilon_3(\bar{S}_3, \bar{A}_2) = (\eta_{3,1} + \eta_{3,2} S_2) \times (S_3 - (\gamma_{3,1} + \gamma_{3,2} S_1 + \gamma_{3,3} A_1 + \gamma_{3,4} S_1 A_1 + \gamma_{3,5} S_2 + \gamma_{3,6} S_1^2 + \gamma_{3,7} S_2^2))$.

Results

Parameters		2-Stage Estimator			Robins' G-Estimator		
		$\widehat{\beta}$	$\widehat{SE}(\widehat{\beta})$	$Pr(>\lvert z\rvert)$	$\widetilde{\beta}$	$\widehat{SE}(\widetilde{\beta})$	$Pr(>\lvert z\rvert)$
Intercept	β_0	3.12	0.07	< 0.01	*	*	*
ϵ_1	$\eta_{1,1}$	0.20	0.07	< 0.01	*	*	*
μ_1	$\beta_{1,0}$	0.04	0.34	0.90	0.17	0.45	0.70
	$\beta_{1,1}$	−0.09	0.22	0.68	0.32	0.36	0.38
ϵ_2	$\eta_{2,1}$	0.13	0.14	0.93	*	*	*
	$\eta_{2,2}$	−0.04	0.07	0.50	*	*	*
μ_2	$\beta_{2,0}$	0.78	0.45	0.08	1.05	0.46	0.20
	$\beta_{2,1}$	−0.76	0.44	0.09	−1.46	0.51	< 0.01
ϵ_3	$\eta_{3,1}$	0.21	0.12	0.08	*	*	*
	$\eta_{3,2}$	0.04	0.10	0.72	*	*	*
μ_3	$\beta_{3,0}$	−0.98	0.32	< 0.01	−1.09	0.33	< 0.01
	$\beta_{3,1}$	1.05	0.40	< 0.01	1.28	0.39	< 0.01

* G-Estimation does not produce estimates for the nuisance functions.

10 Future Work

In point-treatment setting:

1. Dimension reduction/feature construction for eff. moderation
2. Using latent variable methods to help me achieve this
3. I dream of writing for TAS "A Survey of Methods for Causal Effect Moderation in One Time Point"

In the time-varying setting:

1. Top Priority: Clean up bias-variance stuff & get this out
2. JASA case study of MSMs & SNMMs and their estimators (including IPT Weighted versions of the SNMM estimators)

Thank you!
More Questions?

11 Extra Slides

Limitations

- Need more intuition in the simulations: what are the scenarios under which 2-Stage dominates? are we really mis-specifying enough?

- The PROSPECT analysis should start off with the saturated SNMM using a binary suicidal ideation measure (note that the 2-Stage Estimator cannot possibly be mis-specified in this case)

- I never dealt with the randomization variable in PROSPECT.

The Generative Model in Simulations

$nits = 1000$ simulated data sets each of size $n = 500$

1. $\delta_1 \sim \widehat{res}_1$. Then $S_1 \longleftarrow 0.40 + \delta_1$.
2. $Z \longleftarrow$ Bin$(n, p = 0.50)$. Then $A_1 \longleftarrow 0$ if $Z = 0$; otherwise
$$A_1 \longleftarrow \text{Bin}(n, p_1 = \Lambda(1.0 - 0.24 s_1))$$
3. $\delta_2 \sim \text{N}_n(0, \text{sd} = 0.75)$. Generate S_2 by setting
$$S_2 \longleftarrow 0.27 + 0.41 s_1 + 0.01 a_1 - 0.01 s_1^2 - 0.27 s_1 a_1 + \delta_2.$$
4. Set $A_2 \longleftarrow 0$ if $A_1 = 0$; otherwise
$$A_2 \longleftarrow \text{Bin}(n, p_2 = \Lambda(1.0 + 0.40 s_1 - 0.31 s_2)).$$

5. $\delta_3 \sim \mathsf{N}_n(0, \mathsf{sd} = 0.51)$. Generate S_3 by setting

$$S_3 \longleftarrow 0.17 + 0.10 s_1 - 0.25 a_1 + 0.30 s_2 - 0.75 a_2 + 0.05 s_1^2 \\ - 0.04 s_2^2 - 0.1 a_1 s_1 + \delta_3.$$

6. Set $A_3 \longleftarrow 0$ if $A_2 = 0$; otherwise

$$A_3 \longleftarrow \mathsf{Bin}(n, p_3 = \Lambda(1.0 - 0.2 s_1 - 0.3 s_2 + 0.4 s_3)).$$

SNMM Generated as follows:

$$Y \longleftarrow \mathsf{intercept} + \epsilon_1^{\mathsf{TRUE}}(s_1; \eta_1) + a_1(\beta_{1,1} + \beta_{1,2} s_1) \\ + \epsilon_2^{\mathsf{TRUE}}(\bar{s}_2, a_1; \eta_2) + a_2(\beta_{2,1} + \beta_{2,2}(s_1 + s_2)/2) \\ + \epsilon_3^{\mathsf{TRUE}}(\bar{s}_3, \bar{a}_2; \eta_3) + a_3(\beta_{1,1} + \beta_{3,2}(s_1 + s_2 + s_3)/3) + \delta_y,$$

where

1. intercept = 3.55
2. $\beta_{1,1} = \beta_{2,1} = \beta_{3,2} = 0.30$,
3. $\beta_{1,2} = \beta_{2,2} = \beta_{3,1} = -0.30$,
4. δ_y is a random sample of size n from $\mathsf{N}(0, \mathsf{sd} = 0.7)$,

and where the true nuisance functions are defined as

1. $\epsilon_1^{\mathsf{TRUE}}(s_1; \eta_1) = 0.45 \times \delta_1$,
2. $\epsilon_2^{\mathsf{TRUE}}(\bar{s}_2, a_1; \eta_2) =$
   $(0.30 + 0.20 s_1 + 0.15 a_1 + 0.15 a_1 s_1 + 1.0 \sin(4.5 s_1)) \times \delta_2$,
3. $\epsilon_3^{\mathsf{TRUE}}(\bar{s}_3, \bar{a}_2; \eta_3) =$
   $(0.40 - 0.30 s_2 + 0.30 a_2 + 0.60 a_2 s_2 + 1.6 \sin(2.5 s_2)) \times \delta_3$.

Scaled Root Mean Squared Difference

This is how we measured **amount of mis-specification**:

$$\mathsf{SRMSD}(\nu) = \sqrt{\frac{E\left(\sum_{t=3}^{K} \epsilon_t^{\mathsf{TRUE}} - \sum_{t=1}^{K} \epsilon_t^{\nu}(\widehat{\eta}, \widehat{\gamma})\right)^2}{Var(Y)}},$$

where ν corresponds to a mis-specified 2-Stage Regression Estimator.

The expectation E and variance Var in SRMSD are over the data $D = (\bar{S}_3, \bar{A}_3, Y)$ for fixed $(\widehat{\eta}, \widehat{\gamma})$.

Calculated via Monte Carlo integration.

I claim SRMSD has an "effect-size-like" interpretation.

Time-varying Confounding in PROSPECT

	Before Weighting			After Weighting
	Variable Name	Absolute Effect Size†	Sign	Effect Size‡
A_1 = HSANY_4				
	HAMDA_0	0.77	+	0.18
	RE_0	0.64	-	-0.05
	RE16N_0	0.66	-	-0.05
	MCS_0	0.53	-	-0.08
	MMSE2_0	0.45	+	0.22
	SSI_0	0.40	-	-0.29
A_2 = HSANY_8				
	CAD_4	0.80	+	0.27
	DYSTH_0	0.69	-	-0.66
	OPS_0	0.49	-	-0.07
	HAMDA_4	0.55	-	0.07
	CAD_0	0.51	+	0.47
	POSAF_0	0.53	+	-0.28
A_3 = HSANY_12				
	CAD_8	0.90	+	0.37
	WHITE1	0.71	+	0.08
	RP16N_0	0.60	+	0.28