Structural Nested Mean Models for Assessing Time-Varying Effect Moderation

Daniel Almirall
Center for Health Services Research, Durham VAMC
& Duke University Medical, Dept. of Biostatistics

Joint work with Thomas Ten Have (UPenn)
& Susan Murphy (UMich)

JSM – Salt Lake City, Utah – July 31, 2007
## Contents

1. Warm-up: Why do we condition on pre-treatment variables \( S \)? ........................................ 3
2. Time-Varying Effect Moderation ................................................................. 6
3. Robins’ Structural Nested Mean Model ..................................................... 8
4. Estimation in Time-Varying Setting .......................................................... 9
5. Bias-Variance Trade-off ........................................................................ 16
6. Conclusions ......................................................................................... 23
1 Warm-up: Why do we condition on pre-treatment variables $S$?
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.
Effect Moderation in One Time Point

Definition: Fix \( a \neq 0 \). Let

\[
\mu(s, a) \equiv E[Y(a) - Y(0) \mid S = s].
\]

If \( \mu(s, a) \) is non-constant in \( s \), then \( S \) is said to be a moderator of the effect of \( a \) on \( Y \).

Example: If the structural model for the conditional mean of \( Y(a) \) given \( S \) is

\[
E[Y(a) \mid S = s] = \beta_0 + \gamma_1 s + \beta_1 a + \beta_2 sa,
\]

then \( \mu(s, a) = \beta_1 a + \beta_2 sa \). In this example, \( S \) is a moderator of the effect of \( a \) on \( Y \) if \( \beta_2 \neq 0 \).
The focus of this paper is 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
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 \( t = 1 \):

\[
\mu_1(s_1, a_1) \equiv E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1] = a_1 \times H_1\beta_1 \quad \text{(Linear Parameterization)}
\]

say

\[
= a_1 \times (\beta_{11} + \beta_{12}s_1) \quad \text{(Sample Model)}
\]

Conditional Intermediate Causal Effect at \( t = 2 \):

\[
\mu_2(\bar{s}_2, \bar{a}_2) \equiv E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2] = a_2 \times H_2\beta_2 \quad \text{(Linear Parameterization)}
\]

say

\[
= a_2 \times (\beta_{21} + \beta_{22}(s_1 + s_2)/2) \quad \text{(Sample Model)}
\]
3 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:

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

where

\[
\cdot \epsilon_2(\bar{s}_2, a_1) = E[Y(a_1, 0) \mid \bar{S}_2(a_1) = \bar{s}_2] - E[Y(a_1, 0) \mid S_1 = s_1],
\]

\[
\cdot \epsilon_1(s_1) = E[Y(0, 0) \mid S_1 = s_1] - E[Y(0, 0)],
\]

\[
\cdot \mu_2(\bar{s}_2, a_2, 0) = 0 \text{ and } \mu_1(s_1, 0) = 0,
\]

\[
\cdot E_{S_2 \mid S_1} [\epsilon_2(\bar{s}_2, a_1) \mid S_1 = s_1] = 0, \text{ and } E_{S_1} [\epsilon_1(s_1)] = 0.
\]
Recall that parametric models for our causal estimands $\mu_1$ and $\mu_2$ are based on the set of parameters $\beta = (\beta'_1, \beta'_2)'$.

The dissertation considers two estimators for $\beta$:

1. Proposed 2-Stage Regression Estimator
2. Robins’ Semi-parametric G-Estimator

*** We can use 2-Stage Estimator for starting guesses ***

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

Baseline \(\rightarrow\) 4-month Visit \(\rightarrow\) 8-month Visit

\[ V_0 \]

\[ S_1 \]

\[ a_1 \]

\[ S_2(a_1) \]

Set \( a_2 = 0 \)

\[ Y(a_1, 0) \]

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:

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

Example parameterizations for the nuisance functions:

\[
\epsilon_1(s_1) \overset{\text{say}}{=} \eta_{1,1}(s_1 - E(S_1))
\]

\[
\epsilon_2(\bar{s}_2, a_1) \overset{\text{say}}{=} (\eta_{2,1} + \eta_{2,2}s_1)(s_2 - E(S_2(a_1) \mid S_1 = s_1))
\]
Proposed 2-Stage Regression Estimator

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

1. We have models for the \( \mu \)'s: \( A_1H_1\beta_1 \) and \( A_2H_2\beta_2 \); Set aside

2. Model \( m_1(\gamma_1) = E(S_1) \), estimate \( \gamma_1 \) with GLM; model \( m_2(s_1, a_1; \gamma_2) = E(S_2(a_1) \mid S_1 = s_1) \), estimate \( \gamma_2 \) 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 \( \epsilon \)'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 \( Y \sim [1, G_1^*, A_1H_1, G_2^*, A_2H_2] \)
Robins’ Semi-parametric G-Estimator

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

\[
0 = \mathbb{P}_n \left\{ \left( Y - A_2 H_2 \beta_2 - b_2(\bar{S}_2, A_1) \right) \times \left( A_2 - p_2(\bar{S}_2, A_1) \right) \times \begin{pmatrix} 0 \\ H_2' \end{pmatrix}' \right. \\
+ \left( Y - A_2 H_2 \beta_2 - A_1 H_1 \beta_1 - b_1(S_1) \right) \times \left( A_1 - p_1(S_1) \right) \times \begin{pmatrix} H_1' \\ \Delta'(H_1) \end{pmatrix}' \} \\
\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 \middle| \bar{S}_2, A_1 \right] \\
p_2(\bar{S}_2, A_1) = Pr \left[ A_2 = 1 \middle| \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 \middle| S_1 \right] \\
p_1(S_1) = Pr \left[ A_1 = 1 \middle| S_2 \right]
5 Bias-Variance Trade-off

This discussion assumes true models for the causal effects, the \( \mu_t \)s:

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 \( \Delta \) 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?
A Simulation Study of the Bias-Variance Trade-off

1. Generate \( Y \mid (\bar{S}_3, \bar{A}_3) \) SNMM; use \( n = 600 \)

2. Move away from true model fit in a smooth and principled way

3. Fit 2-Stage Estimator and Robins’ G-Estimator

4. Repeat \( N = 1000 \) times

5. Compare bias and variance of the 2-Stage Estimator versus the G-Estimator as you move away from true fit
**Mis-specifying $\epsilon_t$’s using $S^* = S \times N(1, \text{sd} = \nu)$**

**Exploration by Multiplicative Error Method**
Results

Note two different fits for Robins’ G-Estimator:

1. One using “bad” starting guesses (i.e., setting $b_t = 0$)
2. The other using the results of the 2-Stage Estimator as starting guesses
5 Bias-Variance Trade-off
5 Bias-Variance Trade-off

![Graph showing the trade-off between bias and variance with standard deviation.

- **G-Estimator:**
  - 2-Stage Guess
  - G-Estimator: b_t = 0

- **RMSD:**
  - 0.02
  - 0.23
  - 0.33
  - 0.38
  - 0.4
  - 0.41

- **Standard Deviation:**
  - 0.1
  - 0.2
  - 0.3
  - 0.4

- **Legend:**
  - Solid line: G-Estimator: 2-Stage Guess
  - Dashed line: G-Estimator: b_t = 0
  - Triangle: 2-Stage
  - Circle: G-Estimator: b_t = 0

- **Axes:**
  - X-axis: Standard Deviation
  - Y-axis: RMSD
6 Conclusions

1. The bias-variance trade-off is real and important.
2. The G-Estimator relies on good starting guesses.
3. With bad $b_t$ guesses, the G-Estimator never dominates in terms of MSE.
4. With $b_t$ guesses from 2-Stage Estimator, the G-Estimator starts to dominate when amount of mis-specification is between small and medium.
5. Standard errors are difficult for both estimators.
6. The 2-Stage Estimator is easy to carry out, intuitive, and can stand alone as its own estimator.
Thank you!
More Questions?