Structural Nested Mean Models for Assessing Time-Varying Effect Moderation

Daniel Almirall

University of Michigan, Department of Statistics

& Institute for Social Research

Joint work with Thomas Ten Have (UPenn)

& Susan Murphy (UMich)

March 27, 2006

Contents

1 Introduction & Motivation

In any study of the effects of a treatment A on a response Y, there are three reasons to adjust for concomitant 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. **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.

Effect Moderation in $K = 1$ 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 $\beta_2 \neq 0$.

Focus of this paper is Time-Varying Effect Moderation

This paper offers **a method for studying effect moderation** in the longitudinal setting **when treatments, as well as the covariates that modify the effect of treatment, vary over time**.

2 The PROSPECT Study

Consider $K = 2$, with patient visits at 4 and 8 months and the following data structure: $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$.

(a_1, a_2)	Adherence to assigned treatment pattern
$Y(a_1, a_2)$	Depression at the end of the study
S_1	Baseline covariates: both person-specific characteristics and baseline measures of intermediate outcomes, e.g., $S_1 =$ (Race, Gender, Suicidal ideation at $t = 1$, Cognitive status at $t = 1$)
$S_2(a_1)$	Time $t = 2$ measures of the intermediate outcomes, e.g., $S_2 =$ (Suicidal ideation at $t = 2$, Cognitive status at $t = 2$)

3 Time-Varying Effect Moderation

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

Intermediate Effect at $t = 1$: Study how a_1 affects Y on average, as a function of S_1:

$$\mu_1(s_1, a_1) \equiv E[Y(a_1, 0) - Y(0, 0) \mid S_1 = s_1] \overset{\text{say}}{=} \beta_1 a_1 + \beta_2 s_1 a_1$$

Intermediate Effect at $t = 2$: Study how a_2 affects Y on average, as a function of S_1, a_1, and S_2:

$$\mu_2(s_1, a_1, s_2, a_2) \equiv E[Y(a_1, a_2) - Y(a_1, 0) \mid S_1 = s_1, S_2(a_1) = s_2]$$
$$\overset{\text{say}}{=} \beta_3 a_2 + \beta_4 s_1 a_2 + \beta_5 s_2 a_2 + \beta_6 a_1 a_2$$

4 Robins' Structural Nested Mean Model

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

We can express the SNMM for the conditional mean of $Y(a_1, a_2)$ given $\bar{S}_2(a_1)$ as:

$$
\begin{aligned}
E\left[Y(a_1, a_2) \mid \bar{S}_2(a_1)\right] \\
&= \left\{E\left[Y(a_1, a_2) - Y(a_1, 0) \mid \bar{S}_2(a_1)\right]\right\} \\
&+ \left\{E[Y(a_1, 0) \mid \bar{S}_2(a_1)] - E[Y(a_1, 0) \mid S_1]\right\} \\
&+ \left\{E\left[Y(a_1, 0) - Y(0, 0) \mid S_1\right]\right\} \\
&+ \left\{E[Y(0, 0) \mid S_1] - E[Y(0, 0)]\right\} + E[Y(0, 0)]
\end{aligned}
$$

Identifying the Causal and Nuisance Functionals

Recall the data structure $\{S_1, a_1, S_2(a_1), a_2, Y(a_1, a_2)\}$. This decomposition can be written more succinctly as:

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

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 $E_{S_1}[\epsilon_1(s_1)] = 0$.

5 Estimation

The article discusses two estimators for β:

1. 2-Stage Regression Estimator
2. Robins' Semi-parametric G-Estimator

Both estimators rely on Robins' **Consistency** and **Sequential Ignorability** assumptions for causal inference in the time-varying setting. The 2-Stage Estimator makes stronger modelling assumptions.

2-Stage Estimator

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 $E_{S_1}[\epsilon_1(s_1)] = 0$.

A parametrization of the nuisance functions may take the form:

$$\epsilon_1(s_1, a_1) \overset{\text{say}}{=} \eta_{1,1}\big(s_1 - E(S_1)\big)$$

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

2-Stage Estimator - Algorithm

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

Steps (given models for the μ's):

1. Model $m_1(\gamma_1) = E(S_1)$ and estimate γ_1 using OLS; model $m_2(s_1, a_1; \gamma_2) = E(S_2(a_1) \mid S_1 = s_1)$ and estimate γ_2 using OLS.
2. Construct residuals $\hat{s}_1^* = s_1 - \hat{m}_1(\hat{\gamma}_1)$ and $\hat{s}_2^* = s_2 - \hat{m}_2(s_1, a_1; \hat{\gamma}_2)$.
3. Construct models for the ϵ's based on the residuals.
4. Obtain $\hat{\beta}$ and $\hat{\eta} = (\hat{\eta}_1^T, \hat{\eta}_2^T)^T$ using OLS of $Y \sim \mu$'s and ϵ's.

Robins' Semi-parametric Efficient Estimator

Robins' estimator is the solution to the following set of estimating equations:

$$0 = P_n\Bigg[\Big(Y - g_2(\bar{S}_2, A_1) - (H_2\beta_2)\big(A_2 - p_2(\bar{S}_2, A_1)\big)\Big) \times [H_2^T, 0]^T \Big(A_2 - p_2(\bar{S}_2, A_1)\Big) + \Big(Y - H_2 A_2 \beta_2 - g_1(S_1) - (H_1\beta_1)\big(A_1 - p_1(S_1)\big)\Big) \times \left[\Delta(H_1)^T, H_1^T\right]^T \Big(A_1 - p_1(S_1)\Big)\Bigg],$$

where...

Robins' Estimator (continued)

...where

- H_t summarizes the history $(\bar{S}_t, \bar{A}_{t-1})$ (e.g., as in a design matrix); and make up the models for the intermediate effect functions, as in $\mu_1 = H_1 A_1 \beta_2$ and $\mu_2 = H_2 A_2 \beta_2$,
- $\Delta(H_1) = E\left[H_2 A_2 \middle| S_1, A_1 = 1\right] - E\left[H_2 A_2 \middle| S_1, A_1 = 0\right]$
- $g_2(\bar{S}_2, A_1) = E\left[Y \mid \bar{S}_2, A_1\right]$ & $g_1(S_1) = E\left[Y - H_2 A_2 \mid S_1\right]$
- $p_2(\bar{S}_2, A_1) = E\left[A_2 \mid \bar{S}_2, A_1\right]$ & $p_1(S_1) = E\left[A_1 \mid S_2\right]$

6 Coming Soon, Future Work, Etc.

1. Forthcoming work describes bias-variance trade-off
2. Many possible avenues for future work include
   (a) Dealing with time-varying confounders
   (b) Extensions to longitudinal outcomes
   (c) Other Estimators: MLE, Orthogonal 2-Stage Estimator
   (d) Other decompositions of the total causal effect
3. Challenging Issues
   (a) Power, multicolinearity, effect size issues
   (b) Aliasing with high-ordered main effects
   (c) Multiple comparison issues