Chapter 14: G-Estimation of Structural Nested Models

Contents

1 14.1 The Structure of Structural Nested Models (pp. 189-192) 2
1.1 Comparison to Previous Approaches . . . 2

2 14.2 Rank Preservation (pp. 192-194) 2
2.1 Implications . . . 3
2.2 Example: Smoking Cessation and Weight . . . 3

3 14.3 The G-Null Hypothesis (pp. 194-196) 3
3.1 Creating the Pseudo-Outcome . . . 3

4 14.4 Estimating the Causal Effect (pp. 196-199) 4
4.1 G-Estimation Algorithm . . . 4
4.2 Example: Simple Model . . . 4

5 14.5 G-Estimation with Model Misspecification (pp. 199-201) 5
5.1 Robustness Properties . . . 5
5.2 Comparison to Other Methods . . . 5

6 14.6 Estimating the Average Causal Effect (pp. 201-202) 5
6.1 From Conditional to Marginal Effects . . . 5
6.2 Effect in Specific Subgroups . . . 6

7 14.7 Structural Nested Models with Two or More Parameters (pp. 202-204) 6
7.1 General SNMM . . . 6
7.2 Choice of Estimating Functions . . . 6

8 14.8 Censoring and Missing Data (pp. 204-206) 7
8.1 Censoring Weights . . . 7
8.2 Joint Treatment and Censoring Weights . . . 7

9 14.9 Marginal vs Conditional Effects (pp. 206) 7
9.1 Three Types of Effects . . . 7
9.2 Methods and Natural Estimands . . . 8

10 Summary 8

This chapter introduces **G-estimation**, a method for estimating the parameters of **structural nested models (SNMs)**. Unlike IP weighting and standardization, G-estimation does not directly model either the treatment mechanism or the outcome mechanism. Instead, it models the **causal effect itself**, making it robust to certain types of model misspecification.

This chapter is based on Hernán and Robins (2020, chap. 14, pp. 189-206).

Key innovation: G-estimation is based on the idea that if we correctly remove the causal effect of treatment, the residuals should be independent of treatment conditional on confounders. This leads to estimating equations that do not require full specification of treatment or outcome models.

1 14.1 The Structure of Structural Nested Models (pp. 189-192)

Structural nested models directly parameterize the causal effect rather than the mean outcome or treatment probability.

Definition 1.1 (Structural Nested Mean Model)**.** A **structural nested mean model (SNMM)** specifies how the mean of Y^a differs from the mean of $Y^{a'}$ as a function of treatment and covariates:

$$\mathrm{E}\left[Y^a - Y^{a'} \mid L\right] = \gamma(a, a'; \psi, L)$$

For dichotomous treatment with $a = 1$ and $a' = 0$:

$$\mathrm{E}[Y^1 - Y^0 \mid L] = \psi_0 + \psi_1^\top L$$

where $\psi = (\psi_0, \psi_1)$ are the parameters of interest.

1.1 Comparison to Previous Approaches

Marginal structural model (IP weighting):

$$\mathrm{E}[Y^a] = \beta_0 + \beta_1 a$$

Models the mean outcome under treatment a.

Outcome regression (standardization):

$$\mathrm{E}[Y \mid A, L] = \beta_0 + \beta_1 A + \beta_2^\top L + \beta_3^\top (A \times L)$$

Models the conditional mean outcome given treatment and confounders.

Structural nested model (G-estimation):

$$\mathrm{E}[Y^1 - Y^0 \mid L] = \psi_0 + \psi_1^\top L$$

Models the conditional causal effect directly.

Interpretation:

- ψ_0: Average causal effect when $L = 0$
- ψ_1: Effect modification - how the causal effect changes with L
- The model is "nested" because it can be embedded within a more general model for Y^a

Advantage: If the effect model is correct, we get consistent estimates even if we don't correctly specify the full outcome distribution.

2 14.2 Rank Preservation (pp. 192-194)

G-estimation relies on the assumption that treatment affects everyone in the same direction (though possibly by different amounts).

Definition 2.1 (Rank Preservation)**. Rank preservation** (also called **monotonicity** or **no qualitative interaction**) assumes:
If $Y_i^1 > Y_j^1$, then $Y_i^0 > Y_j^0$ for all individuals i, j.
Equivalently: Treatment does not reverse the ranking of individuals with respect to the outcome.

2.1 Implications

Allowed under rank preservation: - Individual causal effects $Y_i^1 - Y_i^0$ can differ across individuals - Some individuals can have large effects, others small effects - Effects can vary with covariates L

NOT allowed under rank preservation: - Treatment helps some individuals ($Y_i^1 > Y_i^0$) and harms others ($Y_j^1 < Y_j^0$) - Qualitative interactions where treatment reverses rankings

2.2 Example: Smoking Cessation and Weight

Rank preservation: - Some people gain more weight than others when quitting - But quitting increases weight for everyone (or at least doesn't decrease it for anyone)

Violation: - Some people gain weight when quitting, others lose weight when quitting

Why this matters:

Rank preservation is stronger than the average null hypothesis $\mathrm{E}[Y^1] = \mathrm{E}[Y^0]$. It's closer to the sharp null hypothesis $Y_i^1 = Y_i^0$ for all i.

While strong, rank preservation is often plausible for outcomes where we expect treatment to affect everyone in the same direction (even if by different amounts).

3 14.3 The G-Null Hypothesis (pp. 194-196)

The key idea of G-estimation: under the null hypothesis of no causal effect with specific parameters, we can construct a pseudo-outcome that is independent of treatment.

Definition 3.1 (G-Null Hypothesis)**.** For a given parameter value ψ, define the **G-null hypothesis** $H_0(\psi)$:

$$H_0(\psi) : Y^1 - Y^0 = \psi_0 + \psi_1^\top L \text{ for all individuals}$$

Under rank preservation, this is equivalent to:

$$H_0(\psi) : Y_i^1 - Y_i^0 = \psi_0 + \psi_1^\top L_i \text{ for all } i$$

3.1 Creating the Pseudo-Outcome

Under $H_0(\psi)$, we can construct:

$$H(\psi) = Y - A(\psi_0 + \psi_1^\top L)$$

Key property: If $H_0(\psi)$ is true, then:

$$H(\psi) = Y^0 \text{ for all individuals}$$

Since Y^0 is the potential outcome under no treatment, it should be independent of actual treatment A given confounders L:

$$H(\psi) \perp\!\!\!\perp A \mid L$$

Intuition:

1. $H(\psi)$ "removes" the effect of treatment from Y using parameter ψ
2. If we remove the correct effect, what remains is the untreated potential outcome Y^0
3. Under conditional exchangeability, $Y^0 \perp\!\!\!\perp A \mid L$
4. So the correct ψ makes $H(\psi)$ independent of A conditional on L

This is the **estimating principle**: Find ψ such that $H(\psi) \perp\!\!\!\perp A \mid L$.

4 14.4 Estimating the Causal Effect (pp. 196-199)

G-estimation finds the value of ψ that makes $H(\psi)$ independent of A conditional on L.

4.1 G-Estimation Algorithm

Step 1: Specify a structural nested model

$$\mathrm{E}[Y^1 - Y^0 \mid L] = \psi_0 + \psi_1^\top L$$

Step 2: For a candidate value ψ, compute pseudo-outcome

$$H(\psi) = Y - A(\psi_0 + \psi_1^\top L)$$

Step 3: Test whether $H(\psi) \perp\!\!\!\perp A \mid L$ by fitting

$$\mathrm{E}[H(\psi) \mid A, L] = \alpha_0 + \alpha_1 A + \alpha_2^\top L$$

Step 4: The correct ψ is the one that makes $\alpha_1 = 0$

Step 5: In practice, solve the estimating equation:

$$\sum_{i=1}^{n} A_i[Y_i - A_i(\psi_0 + \psi_1^\top L_i)] = 0$$

or more generally:

$$\sum_{i=1}^{n} U_i(\psi)[Y_i - A_i(\psi_0 + \psi_1^\top L_i)] = 0$$

where $U_i(\psi)$ is an appropriate function (often $U_i = A_i$ or $U_i = A_i(1, L_i)^\top$).

4.2 Example: Simple Model

SNMM: $\mathrm{E}[Y^1 - Y^0] = \psi_0$ (constant effect)

Estimating equation:

$$\sum_{i=1}^{n} A_i(Y_i - A_i\psi_0) = 0$$

Solution:

$$\hat{\psi}_0 = \frac{\sum_i A_i Y_i}{\sum_i A_i^2} = \frac{\sum_i A_i Y_i}{n_1}$$

where $n_1 = \sum_i A_i$ is the number of treated individuals.

This is the mean outcome among the treated when there is no confounding.

With confounders: The estimating equation becomes more complex:

$$\sum_{i=1}^{n} [A_i - \hat{E}[A \mid L_i]][Y_i - A_i(\psi_0 + \psi_1^\top L_i)] = 0$$

This requires: 1. Estimating $\mathrm{E}[A \mid L]$ (e.g., via logistic regression) 2. Solving for ψ using the estimating equation

The estimator is **doubly robust** in some cases: consistent if either the effect model or the treatment model is correct.

5 14.5 G-Estimation with Model Misspecification (pp. 199-201)

G-estimation has robustness properties that differ from IP weighting and standardization.

5.1 Robustness Properties

When SNMM is correctly specified: - G-estimation is consistent even if $\mathrm{E}[A \mid L]$ is misspecified - Need to correctly model the effect $\mathrm{E}[Y^1 - Y^0 \mid L]$, not the full outcome model $\mathrm{E}[Y \mid A, L]$

When treatment model is correctly specified: - G-estimation is consistent even if the effect model is misspecified in certain ways - Specific robustness depends on the choice of estimating function $U(\psi)$

Double robustness: - Some G-estimators are doubly robust: consistent if either the effect model or a working model for $\mathrm{E}[H(\psi) \mid A, L]$ is correct - This is similar to doubly robust IP weighted estimators

5.2 Comparison to Other Methods

Method	Requires Correct	Robust To
IP weighting	$\Pr[A \mid L]$	Outcome model misspec.
Standardization	$\mathrm{E}[Y \mid A, L]$	Treatment model misspec.
G-estimation	$\mathrm{E}[Y^1 - Y^0 \mid L]$	Full outcome model misspec.
Doubly robust	Either model	One model misspecification

Practical implication:

If you have strong subject-matter knowledge about how the effect varies with covariates, but are uncertain about the full outcome model, G-estimation can be more robust than standardization.

However, G-estimation requires rank preservation, which is a strong assumption that IP weighting and standardization do not require.

6 14.6 Estimating the Average Causal Effect (pp. 201-202)

From the SNMM, we can compute the average causal effect.

6.1 From Conditional to Marginal Effects

SNMM: $\mathrm{E}[Y^1 - Y^0 \mid L] = \psi_0 + \psi_1^\top L$

Average causal effect:

$$\mathrm{E}[Y^1 - Y^0] = E_L[\mathrm{E}[Y^1 - Y^0 \mid L]] = E_L[\psi_0 + \psi_1^\top L] = \psi_0 + \psi_1^\top \mathrm{E}[L]$$

Estimator:

$$\widehat{\mathrm{E}[Y^1 - Y^0]} = \hat{\psi}_0 + \hat{\psi}_1^\top \bar{L}$$

where $\bar{L} = n^{-1} \sum_i L_i$ is the sample mean of L.

6.2 Effect in Specific Subgroups

Effect at $L = \ell$:

$$\mathrm{E}[Y^1 - Y^0 \mid L = \ell] = \psi_0 + \psi_1^\top \ell$$

Effect in the treated (ATT):

$$\mathrm{E}[Y^1 - Y^0 \mid A = 1] = \psi_0 + \psi_1^\top \mathrm{E}[L \mid A = 1]$$

Estimator for ATT:

$$\mathrm{E}[Y^1 - \widehat{Y^0} \mid A = 1] = \hat{\psi}_0 + \hat{\psi}_1^\top \bar{L}_{A=1}$$

where $\bar{L}_{A=1}$ is the mean of L among the treated.

Flexibility: The SNMM $\mathrm{E}[Y^1 - Y^0 \mid L]$ describes effect modification. From this, we can compute:

- Overall average effect $\mathrm{E}[Y^1 - Y^0]$
- Effects in subgroups defined by L
- ATT or ATU (average treatment effect in the untreated)

This is similar to the g-formula, but based on modeling the effect rather than the full outcome.

7 14.7 Structural Nested Models with Two or More Parameters (pp. 202-204)

SNMMs can include multiple effect modifiers.

7.1 General SNMM

$$\mathrm{E}[Y^1 - Y^0 \mid L] = \psi_0 + \psi_1 L_1 + \psi_2 L_2 + \psi_3 L_1 L_2 + ...$$

Parameters: $\psi = (\psi_0, \psi_1, \psi_2, \psi_3, ...)$

Estimating equations: Need as many equations as parameters

$$\sum_{i=1}^{n} U_{ij}(\psi)[Y_i - A_i(\psi_0 + \psi_1 L_{i1} + \psi_2 L_{i2} + ...)] = 0$$

for $j = 1, 2, \ldots, p$ where p is the number of parameters.

7.2 Choice of Estimating Functions

Common choices for U_{ij}:

1. **Simple**: $U_i = (A_i, A_i L_{i1}, A_i L_{i2}, ...)^\top$
2. **Optimal**: $U_i = (A_i - \mathrm{E}[A \mid L_i])(1, L_{i1}, L_{i2}, ...)^\top$
3. **Doubly robust**: More complex functions that achieve double robustness

The choice affects: - Efficiency (variance of estimator) - Robustness properties - Computational complexity

Practical consideration:

For simple SNMMs (e.g., constant effect ψ_0 or effect linear in L), the estimating equations can be solved analytically or via simple iterative methods.

For complex SNMMs with many parameters and interactions, solving the estimating equations requires numerical methods (e.g., Newton-Raphson).

8 14.8 Censoring and Missing Data (pp. 204-206)

G-estimation extends to handle censoring and missing outcomes.

8.1 Censoring Weights

Let $C = 1$ if censored, $C = 0$ if observed.

Assumption: $C \perp\!\!\!\perp Y^a \mid A, L$ (censoring independent of potential outcomes given treatment and covariates)

Weighted estimating equation:

$$\sum_{i:C_i=0} \frac{1}{\Pr[C_i = 0 \mid A_i, L_i]} U_i(\psi)[Y_i - A_i\gamma(A_i, 0; \psi, L_i)] = 0$$

This weights each uncensored observation by the inverse probability of being uncensored.

8.2 Joint Treatment and Censoring Weights

When we have both confounding and censoring:

$$\sum_{i:C_i=0} W_i U_i(\psi)[Y_i - A_i\gamma(A_i, 0; \psi, L_i)] = 0$$

where:

$$W_i = \frac{1}{\Pr[A_i \mid L_i] \times \Pr[C_i = 0 \mid A_i, L_i]}$$

Or using stabilized weights for improved stability.

Extensions:

1. **Time-varying censoring**: G-estimation extends to longitudinal settings with time-varying treatment and censoring (Part III)
2. **Informative censoring**: If censoring depends on unobserved factors, more complex methods are needed (e.g., sensitivity analysis)
3. **Multiple imputation**: Can combine G-estimation with multiple imputation for missing data

9 14.9 Marginal vs Conditional Effects (pp. 206)

G-estimation naturally estimates conditional effects $\mathrm{E}[Y^1 - Y^0 \mid L]$. We can average to get marginal effects.

9.1 Three Types of Effects

Marginal effect (population average):

$$\mathrm{E}[Y^1 - Y^0]$$

Conditional effect (within levels of L):

$$\mathrm{E}[Y^1 - Y^0 \mid L]$$

Individual effect:

$$Y_i^1 - Y_i^0$$

9.2 Methods and Natural Estimands

Method	Natural Estimand	To Get Other Estimands
IP weighting	Marginal effect	Model $\mathrm{E}[Y^a \mid V]$ for conditional
Standardization	Conditional effect	Average over L for marginal
G-estimation	Conditional effect	Average over L for marginal

Advantage of SNMMs: By modeling $\mathrm{E}[Y^1 - Y^0 \mid L]$ directly, G-estimation provides natural inference for effect modification while still allowing marginal effect estimation.

Choosing a method:

- If primarily interested in marginal effects: IP weighting may be most natural
- If interested in effect modification and conditional effects: G-estimation or standardization
- If interested in both: G-estimation provides a unified framework
- For robustness: Consider doubly robust methods that combine approaches

10 Summary

Key concepts introduced:

1. **Structural nested models**: Model the causal effect directly rather than the outcome or treatment mechanism
2. **Rank preservation**: Assumption that treatment doesn't reverse individual rankings
3. **G-null hypothesis**: Under the correct parameters, the pseudo-outcome $H(\psi)$ equals Y^0
4. **G-estimation**: Find ψ that makes $H(\psi)$ independent of A given L
5. **Robustness**: G-estimation is robust to outcome model misspecification (requires effect model to be correct)
6. **Effect modification**: SNMMs naturally model how effects vary with covariates
7. **Censoring**: G-estimation extends to handle missing data via inverse probability weighting

Comparison of methods:

Aspect	IP Weighting	Standardization	G-Estimation
Models	Treatment mechanism	Outcome mechanism	Causal effect
Estimand	Marginal effect	Conditional effect	Conditional effect
Assumptions	Conditional exchangeability	Conditional exchangeability	+ Rank preservation
Robustness	Treatment model	Outcome model	Effect model

Advantages of G-estimation:

- Models the quantity of scientific interest (the causal effect) directly
- Robust to certain outcome model misspecifications
- Natural for effect modification
- Can be doubly robust

Limitations:

- Requires rank preservation (stronger than exchangeability alone)
- Can be computationally more intensive
- Less familiar to many practitioners
- Requires solving estimating equations (no closed form in general)

Practical advice:

1. Use G-estimation when you have strong subject-matter knowledge about effect modification
2. Consider it alongside IP weighting and standardization as a sensitivity analysis
3. Be careful about rank preservation assumption - assess plausibility in your context
4. For simple models, G-estimation can be implemented with standard software
5. For complex models, may need specialized software or custom programming

Looking ahead: Chapter 15 discusses outcome regression and propensity scores in more detail, and Chapter 16 introduces instrumental variables, another approach for dealing with unmeasured confounding.

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Chapman & Hall/CRC. https://miguelhernan.org/whatifbook.