Sandwich estimator derivations for natural effect model parameters

Johan Steen

October 13, 2025

Weighting-based approach

Stratum-specific effects

Models

natural effect model $\mu_1(X, x^*, C; \beta)$:

$$\text{e.g., } g\big[\mathrm{E}\{Y(X, M(x^*))|C\}\big] = \beta_0 + \beta_1 X + \beta_2 x^* + \beta_3 C$$

working model $\mu_2(X, C; \theta)$:

$$\text{e.g., } g\big[\mathrm{E}(M|X, C)\big] = \theta_0 + \theta_1 X + \theta_2 C$$

Estimating equations

$$U_1(X_i, x^*, C_i; \beta, \theta) = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x^*_{ij}, C_i; \beta)}{\partial \beta} \Sigma^{-1}_{\mu_1} \frac{\Pr(M_i|x^*_{ij}, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)} \cdot \big[Y_i - \mu_1(X_i, x^*_{ij}, C_i; \beta)\big]$$

$$U_2(X_i, C_i; \theta) = \frac{\partial \mu_2(X_i, C_i; \theta)}{\partial \theta} \Sigma^{-1}_{\mu_2} \big[M_i - \mu_2(X_i, C_i; \theta)\big]$$

with k the number of replications or hypothetical values x^* for each observation unit i and Σ_{μ_i} the residual variance-covariance matrix for model μ_i.

Sandwich estimator

Let $\zeta = (\beta, \theta)$ and $\tilde{U} = (U_1, U_2)$. The sandwich estimator variance-covariance matrix can then be written as

$$n^{-1} \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} \mathrm{Var}(\tilde{U}) \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-T}$$

with n the total sample size of the original dataset,

$$\mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} = \begin{bmatrix} \mathrm{E}\left(-\frac{\partial U_1}{\partial \beta}\right) & \mathrm{E}\left(-\frac{\partial U_1}{\partial \theta}\right) \\ 0 & \mathrm{E}\left(-\frac{\partial U_2}{\partial \theta}\right) \end{bmatrix}^{-1}$$

and

$$\frac{\partial U_1}{\partial \beta} = -k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x^*_{ij}, C_i; \beta)}{\partial \beta} \Sigma^{-1}_{\mu_1} \frac{\Pr(M_i|x^*_{ij}, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)} \frac{\partial \mu_1(X_i, x^*_{ij}, C_i; \beta)}{\partial \beta}$$

$$\frac{\partial U_1}{\partial \theta} = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x^*_{ij}, C_i; \beta)}{\partial \beta} \Sigma^{-1}_{\mu_1} \frac{\partial}{\partial \theta} \left(\frac{\Pr(M_i|x^*_{ij}, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)}\right) \big[Y_i - \mu_1(X_i, x^*_{ij}, C_i; \beta)\big]$$

$$\frac{\partial U_2}{\partial \theta} = -\frac{\partial \mu_2(X_i, C_i; \theta)}{\partial \theta} \Sigma^{-1}_{\mu_2} \frac{\partial \mu_2(X_i, C_i; \theta)}{\partial \theta}$$

Population-average effects

Models

natural effect model $\mu_1(X, x^*; \beta)$:

$$\text{e.g., } g\big[\mathrm{E}\{Y(X, M(x^*))\}\big] = \beta_0 + \beta_1 X + \beta_2 x^*$$

working model $\mu_2(X, C; \theta)$:

$$\text{e.g., } g\big[\mathrm{E}(M|X, C)\big] = \theta_0 + \theta_1 X + \theta_2 C$$

working model $\mu_3(C; \tau)$:

$$\text{e.g., } g\big[\mathrm{E}(X|C)\big] = \tau_0 + \tau_1 C$$

Estimating equations

$$U_1(X_i, x^*; \beta, \theta, \tau) = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x^*_{ij}; \beta)}{\partial \beta} \Sigma^{-1}_{\mu_1} \Pr(X_i|C_i; \tau)^{-1} \cdot \frac{\Pr(M_i|x^*_{ij}, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)} \cdot \big[Y_i - \mu_1(X_i, x^*_{ij}; \beta)\big]$$

$$U_2(X_i, C_i; \theta) = \frac{\partial \mu_2(X_i, C_i; \theta)}{\partial \theta} \Sigma^{-1}_{\mu_2} \big[M_i - \mu_2(X_i, C_i; \theta)\big]$$

$$U_3(C_i; \tau) = \frac{\partial \mu_3(C_i; \tau)}{\partial \tau} \Sigma^{-1}_{\mu_3} \big[X_i - \mu_3(C_i; \tau)\big]$$

Sandwich estimator

Let $\zeta = (\beta, \theta, \tau)$ and $\tilde{U} = (U_1, U_2, U_3)$. The sandwich estimator variance-covariance matrix can then be written as

$$n^{-1} \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} \mathrm{Var}(\tilde{U}) \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-T}$$

with

$$\mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} = \begin{bmatrix} \mathrm{E}\left(-\frac{\partial U_1}{\partial \beta}\right) & \mathrm{E}\left(-\frac{\partial U_1}{\partial \theta}\right) & \mathrm{E}\left(-\frac{\partial U_1}{\partial \tau}\right) \\ 0 & \mathrm{E}\left(-\frac{\partial U_2}{\partial \theta}\right) & 0 \\ 0 & 0 & \mathrm{E}\left(-\frac{\partial U_3}{\partial \tau}\right) \end{bmatrix}^{-1}$$

and

$$\begin{aligned}
\frac{\partial U_1}{\partial \beta} &= -k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x_{ij}^*; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \Pr(X_i|C_i;\tau)^{-1} \cdot \frac{\Pr(M_i|x_{ij}^*, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)} \\
&\qquad \cdot \frac{\partial \mu_1(X_i, x_{ij}^*; \beta)}{\partial \beta} \\
\frac{\partial U_1}{\partial \theta} &= k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x_{ij}^*; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \Pr(X_i|C_i;\tau)^{-1} \cdot \frac{\partial}{\partial \theta} \left(\frac{\Pr(M_i|x_{ij}^*, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)} \right) \\
&\qquad \cdot \left[Y_i - \mu_1(X_i, x_{ij}^*; \beta)\right] \\
\frac{\partial U_1}{\partial \tau} &= k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(X_i, x_{ij}^*; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \frac{\partial \Pr(X_i|C_i;\tau)^{-1}}{\partial \tau} \cdot \frac{\Pr(M_i|x_{ij}^*, C_i; \theta)}{\Pr(M_i|X_i, C_i; \theta)} \\
&\qquad \cdot \left[Y_i - \mu_1(X_i, x_{ij}^*; \beta)\right] \\
\frac{\partial U_2}{\partial \theta} &= -\frac{\partial \mu_2(X_i, C_i; \theta)}{\partial \theta} \Sigma_{\mu_2}^{-1} \frac{\partial \mu_2(X_i, C_i; \theta)}{\partial \theta} \\
\frac{\partial U_3}{\partial \tau} &= -\frac{\partial \mu_3(C_i; \tau)}{\partial \tau} \Sigma_{\mu_3}^{-1} \frac{\partial \mu_3(C_i; \tau)}{\partial \tau}
\end{aligned}$$

Imputation-based approach

Stratum-specific effects

Models

natural effect model $\mu_1(x, X, C; \beta)$:

$$\text{e.g., } g\big[\mathrm{E}\{Y(x, M(X))|C\}\big] = \beta_0 + \beta_1 x + \beta_2 X + \beta_3 C$$

working model $\mu_2(X, M, C; \gamma)$:

$$\text{e.g., } g\big[\mathrm{E}(Y|X, M, C)\big] = \gamma_0 + \gamma_1 X + \gamma_2 M + \gamma_3 C$$

Estimating equations

$$U_1(x, X_i, C_i; \beta, \gamma) = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i, C_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \cdot \big[\mu_2(x_{ij}, M_i, C_i; \gamma) - \mu_1(x_{ij}, X_i, C_i; \beta)\big]$$

$$U_2(X_i, M_i, C_i; \gamma) = \frac{\partial \mu_2(X_i, M_i, C_i; \gamma)}{\partial \gamma} \Sigma_{\mu_2}^{-1} \big[Y_i - \mu_2(X_i, M_i, C_i; \gamma)\big]$$

Sandwich estimator

Let $\zeta = (\beta, \gamma)$ and $\tilde{U} = (U_1, U_2)$. The sandwich estimator variance-covariance matrix can then be written as

$$n^{-1} \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} \mathrm{Var}(\tilde{U}) \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-T}$$

with

$$\mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} = \begin{bmatrix} \mathrm{E}\left(-\frac{\partial U_1}{\partial \beta}\right) & \mathrm{E}\left(-\frac{\partial U_1}{\partial \gamma}\right) \\ 0 & \mathrm{E}\left(-\frac{\partial U_2}{\partial \gamma}\right) \end{bmatrix}^{-1}$$

and

$$\frac{\partial U_1}{\partial \beta} = -k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i, C_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \frac{\partial \mu_1(x_{ij}, X_i, C_i; \beta)}{\partial \beta}$$

$$\frac{\partial U_1}{\partial \gamma} = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i, C_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \frac{\partial \mu_2(x_{ij}, M_i, C_i; \gamma)}{\partial \gamma}$$

$$\frac{\partial U_2}{\partial \gamma} = -\frac{\partial \mu_2(X_i, M_i, C_i; \beta)}{\partial \gamma} \Sigma_{\mu_2}^{-1} \frac{\partial \mu_2(X_i, M_i, C_i; \beta)}{\partial \gamma}$$

Population-average effects

Models

natural effect model $\mu_1(x, X; \beta)$:

$$\text{e.g., } g\big[\mathrm{E}\{Y(x, M(X))\}\big] = \beta_0 + \beta_1 x + \beta_2 X$$

working model $\mu_2(X, M, C; \gamma)$:

$$\text{e.g., } g\big[\mathrm{E}(Y|X, M, C)\big] = \gamma_0 + \gamma_1 X + \gamma_2 M + \gamma_3 C$$

working model $\mu_3(C; \tau)$:

$$\text{e.g., } g\big[\mathrm{E}(X|C)\big] = \tau_0 + \tau_1 C$$

Estimating equations

$$U_1(x, X_i; \beta, \gamma, \tau) = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \Pr(X_i | C_i; \tau)^{-1} \cdot \big[\mu_2(x_{ij}, M_i, C_i; \gamma) - \mu_1(x_{ij}, X_i; \beta)\big]$$

$$U_2(X_i, M_i, C_i; \gamma) = \frac{\partial \mu_2(X_i, M_i, C_i; \gamma)}{\partial \gamma} \Sigma_{\mu_2}^{-1} \big[Y_i - \mu_2(X_i, M_i, C_i; \gamma)\big]$$

$$U_3(C_i; \tau) = \frac{\partial \mu_3(C_i; \tau)}{\partial \tau} \Sigma_{\mu_3}^{-1} \big[X_i - \mu_3(C_i; \tau)\big]$$

Sandwich estimator

Let $\zeta = (\beta, \gamma, \tau)$ and $\tilde{U} = (U_1, U_2, U_3)$. The sandwich estimator variance-covariance matrix can then be written as

$$n^{-1} \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} \mathrm{Var}(\tilde{U}) \cdot \mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-T}$$

with

$$\mathrm{E}\left(-\frac{\partial \tilde{U}}{\partial \zeta}\right)^{-1} = \begin{bmatrix} \mathrm{E}\left(-\frac{\partial U_1}{\partial \beta}\right) & \mathrm{E}\left(-\frac{\partial U_1}{\partial \gamma}\right) & \mathrm{E}\left(-\frac{\partial U_1}{\partial \tau}\right) \\ 0 & \mathrm{E}\left(-\frac{\partial U_2}{\partial \gamma}\right) & 0 \\ 0 & 0 & \mathrm{E}\left(-\frac{\partial U_3}{\partial \tau}\right) \end{bmatrix}^{-1}$$

and

$$\frac{\partial U_1}{\partial \beta} = -k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \Pr(X_i|C_i)^{-1} \frac{\partial \mu_1(x_{ij}, X_i; \beta)}{\partial \beta}$$

$$\frac{\partial U_1}{\partial \gamma} = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \Pr(X_i|C_i)^{-1} \frac{\partial \mu_2(x_{ij}, M_i, C_i; \gamma)}{\partial \gamma}$$

$$\frac{\partial U_1}{\partial \tau} = k^{-1} \sum_{j=1}^{k} \frac{\partial \mu_1(x_{ij}, X_i; \beta)}{\partial \beta} \Sigma_{\mu_1}^{-1} \frac{\partial \Pr(X_i|C_i; \tau)^{-1}}{\partial \tau} \cdot \big[\mu_2(x_{ij}, M_i, C_i; \gamma) - \mu_1(x_{ij}, X_i; \beta)\big]$$

$$\frac{\partial U_2}{\partial \gamma} = -\frac{\partial \mu_2(X_i, M_i, C_i; \beta)}{\partial \gamma} \Sigma_{\mu_2}^{-1} \frac{\partial \mu_2(X_i, M_i, C_i; \beta)}{\partial \gamma}$$

$$\frac{\partial U_3}{\partial \tau} = -\frac{\partial \mu_3(C_i; \tau)}{\partial \tau} \Sigma_{\mu_3}^{-1} \frac{\partial \mu_3(C_i; \tau)}{\partial \tau}$$