Asymptotic variance of the weighted pseudo-likelihood estimator in the presence of nuisance parameters

1 Unstabilized weights

Let

$$q(\theta; x, y, z, \gamma) \equiv \sum_{i=1}^{n} \sum_{a} \mathbf{1}_{\{z_i = z^{(a)}\}} \frac{l(\theta; \mathbf{y}_{ia})}{p(z^{(a)} \mid x_i, \gamma)}$$

be the weighted pseudo-log-likelihood for marginal structural model parameters θ and

$$m(\gamma; x, z) \equiv \sum_{i=1}^{n} \log p(z_i \mid x_i, \gamma)$$

the log-likelihood for treatment assignment model parameters γ. Let the corresponding maximum likelihood estimates be $\hat{\gamma} \equiv \arg\max_{\gamma} m(\gamma; x, z)$ and $\hat{\theta} \equiv \arg\max_{\theta} q(\theta; x, y, z, \hat{\gamma})$. In addition, let

$$\begin{aligned}
U^{\theta}(\theta; \gamma) &\equiv \partial q(\theta; x, y, z, \gamma)/\partial\theta, \\
I^{\theta\theta}(\theta; \gamma) &\equiv \partial^2 q(\theta; x, y, z, \gamma)/\partial\theta^2, \\
I^{\theta\gamma}(\theta; \gamma) &\equiv \partial^2 q(\theta; x, y, z, \gamma)/\partial\theta\partial\gamma, \\
U^{\gamma}(\gamma) &\equiv \partial m(\gamma; x, z)/\partial\gamma, \\
I^{\gamma\gamma}(\gamma) &\equiv \partial^2 m(\gamma; x, z)/\partial\gamma^2.
\end{aligned}$$

We are interested in the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta_0)$. Following Robins et al. (1992, p. 494), a first order Taylor expansion around the true parameter values (θ_0, γ_0) gives

$$\begin{aligned}
0 = \frac{1}{n} U^{\theta}(\hat{\theta}; \hat{\gamma}) &\approx \frac{1}{n} U^{\theta}(\theta_0; \gamma_0) + \frac{1}{n} I^{\theta\theta}(\theta_0; \gamma_0)(\hat{\theta} - \theta_0) + \frac{1}{n} I^{\theta\gamma}(\theta_0; \gamma_0)(\hat{\gamma} - \gamma_0) \\
&\approx \frac{1}{n} U^{\theta}(\theta_0; \gamma_0) + E[I_i^{\theta\theta}(\theta_0; \gamma_0)](\hat{\theta} - \theta_0) + E[I_i^{\theta\gamma}(\theta_0; \gamma_0)](\hat{\gamma} - \gamma_0)
\end{aligned}$$

or

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx E[-I_i^{\theta\theta}(\theta_0; \gamma_0)]^{-1} \left[\frac{\sqrt{n}}{n} U^{\theta}(\theta_0; \gamma_0) + E[I_i^{\theta\gamma}(\theta_0; \gamma_0)] \sqrt{n}(\hat{\gamma} - \gamma_0) \right].$$

Here, by another Taylor expansion around γ_0,

$$\sqrt{n}(\hat{\gamma}-\gamma_0) \approx E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}\frac{\sqrt{n}}{n}U^{\gamma}(\gamma_0),$$

which substituted back to the previous expression gives

$$\sqrt{n}(\hat{\theta}-\theta_0) \approx E[-I_i^{\theta\theta}(\theta_0;\gamma_0)]^{-1}\left[\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left\{U_i^{\theta}(\theta_0;\gamma_0)+E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^{\gamma}(\gamma_0)\right\}\right].$$

By the central limit theorem it then follows that

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{\mathrm{d}} N(0, E[-I_i^{\theta\theta}(\theta_0;\gamma_0)]^{-1}V[B_i(\theta_0,\gamma_0)]E[-I^{\theta\theta}(\theta_0;\gamma_0)']^{-1}),$$

where

$$B_i(\theta_0,\gamma_0) \equiv U_i^{\theta}(\theta_0;\gamma_0)+E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^{\gamma}(\gamma_0).$$

Since $E[B_i(\theta_0,\gamma_0)]=0$, the variance $V[B_i(\theta_0,\gamma_0)]$ can be further written as

$$\begin{aligned}
V[B_i(\theta_0,\gamma_0)] &\equiv V[U_i^{\theta}(\theta_0;\gamma_0)+E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^{\gamma}(\gamma_0)] \\
&= E\big[\{U_i^{\theta}(\theta_0;\gamma_0)+E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^{\gamma}(\gamma_0)\} \\
&\qquad \times \{U_i^{\theta}(\theta_0;\gamma_0)+E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^{\gamma}(\gamma_0)\}'\big] \\
&= E[U_i^{\theta}(\theta_0;\gamma_0)U_i^{\theta}(\theta_0;\gamma_0)'] \\
&\quad + E[U_i^{\theta}(\theta_0;\gamma_0)U_i^{\gamma}(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1\prime}E[I_i^{\theta\gamma}(\theta_0;\gamma_0)'] \\
&\quad + E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[U_i^{\gamma}(\gamma_0)U_i^{\theta}(\theta_0;\gamma_0)'] \\
&\quad + E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[U_i^{\gamma}(\gamma_0)U_i^{\gamma}(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1\prime}E[I_i^{\theta\gamma}(\theta_0;\gamma_0)'],
\end{aligned}$$

where the final term

$$\begin{aligned}
&E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[U_i^{\gamma}(\gamma_0)U_i^{\gamma}(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1\prime}E[I_i^{\theta\gamma}(\theta_0;\gamma_0)'] \\
&= E[I_i^{\theta\gamma}(\theta_0;\gamma_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[I_i^{\theta\gamma}(\theta_0;\gamma_0)'],
\end{aligned}$$

since $m(\gamma;x,z)$ is a log-likelihood and $E[U_i^{\gamma}(\gamma_0)U_i^{\gamma}(\gamma_0)']=E[-I_i^{\gamma\gamma}(\gamma_0)]$. In addition, we have that

$$
\begin{aligned}
&E[U_i^\theta(\theta;\gamma)U_i^\gamma(\gamma)'] \\
&= \int_{\mathbf{y}_i,x_i} \sum_{z_i} \sum_a \mathbf{1}_{\{z_i=z^{(a)}\}} \frac{l'(\theta;\mathbf{y}_{ia})}{p(z^{(a)} \mid x_i,\gamma)} \frac{\partial \log p(z_i \mid x_i,\gamma)}{\partial\gamma} p(z_i \mid x_i) p(\mathrm{d}\mathbf{y}_i, \mathrm{d}x_i) \\
&= \sum_a \int_{\mathbf{y}_{ia},x_i} \frac{l'(\theta;\mathbf{y}_{ia})}{p(z^{(a)} \mid x_i,\gamma)} \frac{\partial \log p(z^{(a)} \mid x_i,\gamma)}{\partial\gamma} p(z^{(a)} \mid x_i) p(\mathrm{d}\mathbf{y}_{ia}, \mathrm{d}x_i) \\
&= \sum_a \int_{\mathbf{y}_{ia},x_i} l'(\theta;\mathbf{y}_{ia}) \frac{\partial p(z^{(a)} \mid x_i,\gamma)/\partial\gamma}{p(z^{(a)} \mid x_i,\gamma)} p(\mathrm{d}\mathbf{y}_{ia}, \mathrm{d}x_i) \\
&= \sum_a E_{\mathbf{y}_{ia},x_i}\left[l'(\theta;\mathbf{y}_{ia}) \frac{\partial p(z^{(a)} \mid x_i,\gamma)/\partial\gamma}{p(z^{(a)} \mid x_i,\gamma)} \right]
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&E[I_i^{\theta\gamma}(\theta;\gamma)] \\
&= \sum_a \int_{\mathbf{y}_{ia},x_i} \frac{\partial \frac{l'(\theta;\mathbf{y}_{ia})}{p(z^{(a)}|x_i,\gamma)}}{\partial\gamma} p(z^{(a)} \mid x_i) p(\mathrm{d}\mathbf{y}_{ia}, \mathrm{d}x_i) \\
&= \sum_a \int_{\mathbf{y}_{ia},x_i} -l'(\theta;\mathbf{y}_{ia}) \frac{\partial p(z^{(a)} \mid x_i,\gamma)/\partial\gamma}{p(z^{(a)} \mid x_i,\gamma)^2} p(z^{(a)} \mid x_i) p(\mathrm{d}\mathbf{y}_{ia}, \mathrm{d}x_i) \\
&= E[-U_i^\theta(\theta;\gamma)U_i^\gamma(\gamma)].
\end{aligned}
$$

Thus, finally, the variance becomes

$$
\begin{aligned}
V[B_i(\theta_0,\gamma_0)] &= E[U_i^\theta(\theta_0;\gamma_0)U_i^\theta(\theta_0;\gamma_0)'] \\
&\quad - E[U_i^\theta(\theta_0;\gamma_0)U_i^\gamma(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1\prime}E[U_i^\gamma(\gamma_0)U_i^\theta(\theta_0;\gamma_0)'] \\
&\quad - E[U_i^\theta(\theta_0;\gamma_0)U_i^\gamma(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[U_i^\gamma(\gamma_0)U_i^\theta(\theta_0;\gamma_0)'] \\
&\quad + E[U_i^\theta(\theta_0;\gamma_0)U_i^\gamma(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[U_i^\gamma(\gamma_0)U_i^\theta(\theta_0;\gamma_0)'] \\
&= E[U_i^\theta(\theta_0;\gamma_0)U_i^\theta(\theta_0;\gamma_0)'] \\
&\quad - E[U_i^\theta(\theta_0;\gamma_0)U_i^\gamma(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}E[U_i^\gamma(\gamma_0)U_i^\theta(\theta_0;\gamma_0)'].
\end{aligned}
$$

The term $nV[B_i(\theta_0, \gamma_0)]$ can be approximated with

$$
\begin{aligned}
C(\hat{\theta}, \hat{\gamma}) \equiv & \sum_{i=1}^{n} U_i^{\theta}(\hat{\theta}; \hat{\gamma}) U_i^{\theta}(\hat{\theta}; \hat{\gamma})' \\
& - \left[\sum_{i=1}^{n} U_i^{\theta}(\hat{\theta}; \hat{\gamma}) U_i^{\gamma}(\hat{\gamma})' \right] [-I^{\gamma\gamma}(\hat{\gamma})]^{-1} \left[\sum_{i=1}^{n} U_i^{\gamma}(\hat{\gamma}) U_i^{\theta}(\hat{\theta}; \hat{\gamma})' \right],
\end{aligned}
$$

giving the variance estimator

$$
V[\hat{\theta}] \approx I^{\theta\theta}(\hat{\theta}; \hat{\gamma})^{-1} C(\hat{\theta}, \hat{\gamma}) I^{\theta\theta}(\hat{\theta}; \hat{\gamma})^{-1\prime}. \tag{1}
$$

It should be noted that when the true nuisance parameter values γ_0 are known, one can take the Fisher information on γ to be $E[-I_i^{\gamma\gamma}(\gamma_0)] = \infty$, in which case the asymptotic variance reduces to

$$
E[-I_i^{\theta\theta}(\theta_0; \gamma_0)]^{-1} E[U_i^{\theta}(\theta_0; \gamma_0) U_i^{\theta}(\theta_0; \gamma_0)'] E[-I^{\theta\theta}(\theta_0; \gamma_0)']^{-1} \tag{2}
$$

and the variance estimator to

$$
V[\hat{\theta}] \approx I^{\theta\theta}(\hat{\theta}; \hat{\gamma})^{-1} \left[\sum_{i=1}^{n} U_i^{\theta}(\hat{\theta}; \hat{\gamma}) U_i^{\theta}(\hat{\theta}; \hat{\gamma})' \right] I^{\theta\theta}(\hat{\theta}; \hat{\gamma})^{-1\prime}, \tag{3}
$$

which is the usual robust/sandwich variance expression.

2 Stabilized weights

With stabilized weights the estimating function is of the form

$$
q(\theta; x, y, z, \gamma, \alpha) \equiv \sum_{i=1}^{n} \sum_{a} \mathbf{1}_{\{z_i = z^{(a)}\}} \frac{p(z^{(a)} \mid \alpha)}{p(z^{(a)} \mid x_i, \gamma)} l(\theta; \mathbf{y}_{ia}),
$$

where α is an additional vector of nuisance parameters needed to be estimated. Analogously to previous section, let

$$
s(\alpha; z) \equiv \sum_{i=1}^{n} \log p(z_i \mid \alpha),
$$

$\hat{\alpha} \equiv \arg\max_\alpha s(\alpha; z)$, $\hat{\theta} \equiv \arg\max_\theta q(\theta; x, y, z, \hat{\gamma}, \hat{\alpha})$, and

$$\begin{aligned}
U^\theta(\theta;\gamma,\alpha) &\equiv \partial q(\theta;x,y,z,\gamma,\alpha)/\partial\theta,\\
I^{\theta\theta}(\theta;\gamma,\alpha) &\equiv \partial^2 q(\theta;x,y,z,\gamma,\alpha)/\partial\theta^2,\\
I^{\theta\gamma}(\theta;\gamma,\alpha) &\equiv \partial^2 q(\theta;x,y,z,\gamma,\alpha)/\partial\theta\partial\gamma,\\
I^{\theta\alpha}(\theta;\gamma,\alpha) &\equiv \partial^2 q(\theta;x,y,z,\gamma,\alpha)/\partial\theta\partial\alpha,\\
U^\gamma(\gamma) &\equiv \partial m(\gamma;x,z)/\partial\gamma,\\
I^{\gamma\gamma}(\gamma) &\equiv \partial^2 m(\gamma;x,z)/\partial\gamma^2\\
U^\alpha(\alpha) &\equiv \partial s(\alpha;z)/\partial\alpha,\\
I^{\alpha\alpha}(\alpha) &\equiv \partial^2 s(\alpha;z)/\partial\alpha^2.
\end{aligned}$$

Asymptotic variance is now obtained from first order Taylor expansions around the true parameter values $(\theta_0, \gamma_0, \alpha_0)$ and around γ_0 and α_0 separately, which give

$$\begin{aligned}
\sqrt{n}(\hat{\theta}-\theta_0) \approx E[-I_i^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)]^{-1}\Bigg[\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\big\{&U_i^\theta(\theta_0;\gamma_0,\alpha_0)\\
&+E[I_i^{\theta\gamma}(\theta_0;\gamma_0,\alpha_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^\gamma(\gamma_0)\\
&+E[I_i^{\theta\alpha}(\theta_0;\gamma_0,\alpha_0)]E[-I_i^{\alpha\alpha}(\alpha_0)]^{-1}U_i^\alpha(\alpha_0)\big\}\Bigg].
\end{aligned}$$

The asymptotic distribution is now

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{\mathrm{d}} N(0, E[-I_i^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)]^{-1}V[A_i(\theta_0,\gamma_0,\alpha_0)]E[-I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)']^{-1}),$$

where

$$\begin{aligned}
A_i(\theta_0,\gamma_0,\alpha_0) \equiv\ &U_i^\theta(\theta_0;\gamma_0,\alpha_0)\\
&+E[I_i^{\theta\gamma}(\theta_0;\gamma_0,\alpha_0)]E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1}U_i^\gamma(\gamma_0)\\
&+E[I_i^{\theta\alpha}(\theta_0;\gamma_0,\alpha_0)]E[-I_i^{\alpha\alpha}(\alpha_0)]^{-1}U_i^\alpha(\alpha_0).
\end{aligned}$$

Since $E[A_i(\theta_0,\gamma_0,\alpha_0)] = 0$, the variance of this term becomes $V[A_i(\theta_0,\gamma_0,\alpha_0)] = E[A_i(\theta_0,\gamma_0,\alpha_0)A_i(\theta_0,\gamma_0,\alpha_0)']$. For simplifying this expression, we note as before that $E[I_i^{\theta\gamma}(\theta_0;\gamma_0,\alpha_0)] = E[-U_i^\theta(\theta_0;\gamma_0,\alpha_0)U_i^\gamma(\gamma_0)]$, $E[U_i^\gamma(\gamma_0)U_i^\gamma(\gamma_0)'] =$

$E[-I_i^{\gamma\gamma}(\gamma_0)]$ and $E[U_i^{\alpha}(\alpha_0)U_i^{\alpha}(\alpha_0)'] = E[-I_i^{\alpha\alpha}(\alpha_0)]$. In addition,

$$\begin{aligned}
&E[I_i^{\theta\alpha}(\theta;\gamma,\alpha)] \\
&= \sum_a \int_{\mathbf{y}_{ia},x_i} \frac{\partial p(z^{(a)} \mid \alpha)/\partial\alpha}{p(z^{(a)} \mid x_i,\gamma)} l'(\theta;\mathbf{y}_{ia}) p(z^{(a)} \mid x_i) p(\mathrm{d}\mathbf{y}_{ia},\mathrm{d}x_i) \\
&= \sum_a \frac{\partial p(z^{(a)} \mid \alpha)}{\partial\alpha} E\left[l'(\theta;\mathbf{y}_{ia})\right].
\end{aligned}$$

When the marginal structural model is correctly specified, $E\left[l'(\theta;\mathbf{y}_{ia})\right] = 0$, and we have $E[I_i^{\theta\alpha}(\theta;\gamma,\alpha)] = 0$. Hence, the variance simplifies into

$$\begin{aligned}
V[A_i(\theta_0,\gamma_0,\alpha_0)] &= E[U_i^{\theta}(\theta_0;\gamma_0,\alpha_0)U_i^{\theta}(\theta_0;\gamma_0,\alpha_0)'] \\
&\quad - E[U_i^{\theta}(\theta_0;\gamma_0,\alpha_0)U_i^{\gamma}(\gamma_0)']E[-I_i^{\gamma\gamma}(\gamma_0)]^{-1'}E[U_i^{\gamma}(\gamma_0)U_i^{\theta}(\theta_0;\gamma_0,\alpha_0)'],
\end{aligned}$$

which is of the same form as in the previous section. Thus, variance estimator of the form (1) is applicable also with stabilized weights, with $U_i^{\theta}(\hat\theta;\hat\gamma,\hat\alpha)$ substituted for $U_i^{\theta}(\hat\theta;\hat\gamma)$ and $I^{\theta\theta}(\hat\theta;\hat\gamma,\hat\alpha)$ for $I^{\theta\theta}(\hat\theta;\hat\gamma)$.

3 Connection to the jackknife variance estimator

For simplicity, consider first the situation where the true nuisance parameter values (γ_0,α_0) are known. Then we have

$$\begin{aligned}
0 = U^{\theta}(\hat\theta;\gamma_0,\alpha_0) &\approx U^{\theta}(\theta_0;\gamma_0,\alpha_0) + I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)(\hat\theta-\theta_0) \\
\Leftrightarrow \hat\theta - \theta_0 &\approx -I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)^{-1}U^{\theta}(\theta_0;\gamma_0,\alpha_0) \\
&\approx E[-I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)]^{-1}U^{\theta}(\theta_0;\gamma_0,\alpha_0)
\end{aligned}$$

and

$$\begin{aligned}
0 = U_{-i}^{\theta}(\hat\theta;\gamma_0,\alpha_0) &\approx U_{-i}^{\theta}(\theta_0;\gamma_0,\alpha_0) + I_{-i}^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)(\hat\theta_{-i}-\theta_0) \\
\Leftrightarrow \hat\theta_{-i} - \theta_0 &\approx -I_{-i}^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)^{-1}U_{-i}^{\theta}(\theta_0;\gamma_0,\alpha_0) \\
&\approx E[-I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)]^{-1}U_{-i}^{\theta}(\theta_0;\gamma_0,\alpha_0),
\end{aligned}$$

where U_{-i}^{θ}, $I_{-i}^{\theta\theta}$ and $\hat\theta_{-i}$ are the pseudo-score function, pseudo-information and pseudo-maximum likelihood estimator when observation i is removed from the data. Combining the two approximations we have

$$\begin{aligned}\hat{\theta} - \hat{\theta}_{-i} &\approx E[-I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)]^{-1} U_i^{\theta}(\theta_0; \gamma_0, \alpha_0) \\ &\approx -I^{\theta\theta}(\hat{\theta}; \gamma_0, \alpha_0)^{-1} U_i^{\theta}(\hat{\theta}; \gamma_0, \alpha_0),\end{aligned}$$

which is an approximation of the influence function on the estimator $\hat{\theta}$ when observation i is removed. The robust/sandwich variance estimator obtained earlier can then be represented in terms of the approximate influences as

$$\begin{aligned}V[\hat{\theta}] &\approx I^{\theta\theta}(\hat{\theta}; \gamma_0, \alpha_0)^{-1} \left[\sum_{i=1}^{n} U_i^{\theta}(\hat{\theta}; \gamma_0, \alpha_0) U_i^{\theta}(\hat{\theta}; \gamma_0, \alpha_0)' \right] I^{\theta\theta}(\hat{\theta}; \gamma_0, \alpha_0)^{-1'} \\ &= \sum_{i=1}^{n} (\hat{\theta} - \hat{\theta}_{-i})(\hat{\theta} - \hat{\theta}_{-i})'.\end{aligned}$$

Thus, the robust/sandwich variance estimator is equivalent to the jackknife variance estimator calculated from the large-sample approximations of the influences. This suggests that the small sample performance of the asymptotic variances could be investigated by checking the large-sample influences against the empirical influences when the observations are removed one at a time.

When the nuisance parameters need to be estimated, the same relationship between the asymptotic variance and the jackknife variance estimator holds. We obtain similarly as before

$$\begin{aligned}0 = U^{\theta}(\hat{\theta}; \hat{\gamma}, \hat{\alpha}) &\approx U^{\theta}(\theta_0; \gamma_0, \alpha_0) + I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)(\hat{\theta} - \theta_0) \\ &\quad + I^{\theta\gamma}(\theta_0; \gamma_0, \alpha_0)(\hat{\gamma} - \gamma_0) + I^{\theta\alpha}(\theta_0; \gamma_0, \alpha_0)(\hat{\alpha} - \alpha_0) \\ &\approx U^{\theta}(\theta_0; \gamma_0, \alpha_0) + I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)(\hat{\theta} - \theta_0) \\ &\quad - I^{\theta\gamma}(\theta_0; \gamma_0, \alpha_0) I^{\gamma\gamma}(\gamma_0)^{-1} U^{\gamma}(\gamma_0) \\ \Leftrightarrow \hat{\theta} - \theta_0 &\approx E[-I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)]^{-1} U^{\theta}(\theta_0; \gamma_0, \alpha_0) \\ &\quad + E[-I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)]^{-1} E[I^{\theta\gamma}(\theta_0; \gamma_0, \alpha_0)] E[-I^{\gamma\gamma}(\gamma_0)]^{-1} U^{\gamma}(\gamma_0)\end{aligned}$$

and

$$\begin{aligned}\hat{\theta}_{-i} - \theta_0 &\approx E[-I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)]^{-1} U_{-i}^{\theta}(\theta_0; \gamma_0, \alpha_0) \\ &\quad + E[-I^{\theta\theta}(\theta_0; \gamma_0, \alpha_0)]^{-1} E[I^{\theta\gamma}(\theta_0; \gamma_0, \alpha_0)] E[-I^{\gamma\gamma}(\gamma_0)]^{-1} U_{-i}^{\gamma}(\gamma_0),\end{aligned}$$

and further,

$$
\begin{aligned}
\hat{\theta} - \hat{\theta}_{-i} &\approx E[-I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)]^{-1}U_i^{\theta}(\theta_0;\gamma_0,\alpha_0) \\
&\qquad + E[-I^{\theta\theta}(\theta_0;\gamma_0,\alpha_0)]^{-1}E[I^{\theta\gamma}(\theta_0;\gamma_0,\alpha_0)]E[-I^{\gamma\gamma}(\gamma_0)]^{-1}U_i^{\gamma}(\gamma_0) \\
&\approx -I^{\theta\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})^{-1}U_i^{\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha}) \\
&\qquad - [-I^{\theta\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})]^{-1}\left[\sum_{i=1}^{n} U_i^{\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})U_i^{\gamma}(\hat{\gamma})'\right][-I^{\gamma\gamma}(\hat{\gamma})]^{-1}U_i^{\gamma}(\hat{\gamma}),
\end{aligned}
$$

since $E[I^{\theta\gamma}(\theta;\gamma,\alpha)] = E\left[\sum_{i=1}^{n} I_i^{\theta\gamma}(\theta;\gamma,\alpha)\right] = E\left[-\sum_{i=1}^{n} U_i^{\theta}(\theta;\gamma,\alpha)U_i^{\gamma}(\gamma)'\right] = E[-U^{\theta}(\theta;\gamma,\alpha)U^{\gamma}(\gamma)']$. Now

$$
\begin{aligned}
V[\hat{\theta}] &\approx I^{\theta\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})^{-1}\left[\sum_{i=1}^{n} U_i^{\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})U_i^{\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})'\right] I^{\theta\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})^{-1\prime} \\
&\qquad - I^{\theta\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})^{-1}\left[\sum_{i=1}^{n} U_i^{\theta}(\hat{\theta};\hat{\gamma})U_i^{\gamma}(\hat{\gamma})'\right][-I^{\gamma\gamma}(\hat{\gamma})]^{-1}\left[\sum_{i=1}^{n} U_i^{\gamma}(\hat{\gamma})U_i^{\theta}(\hat{\theta};\hat{\gamma})'\right] I^{\theta\theta}(\hat{\theta};\hat{\gamma},\hat{\alpha})^{-1\prime} \\
&\approx \sum_{i=1}^{n}(\hat{\theta}-\hat{\theta}_{-i})(\hat{\theta}-\hat{\theta}_{-i})',
\end{aligned}
$$

since $\sum_{i=1}^{n} U_i^{\gamma}(\hat{\gamma})U_i^{\gamma}(\hat{\gamma})' \approx -I^{\gamma\gamma}(\hat{\gamma})$.

4 Adjustments to correct for undercoverage

Going back to expressions (2) and (3), rather than approximate $\mathrm{cov}[U_i^{\theta}(\theta_0;\gamma_0)]$ directly with the empirical score covariance, following Fay and Graubard (2001), we get with approximations

$$U_i^{\theta}(\hat{\theta};\hat{\gamma}) \approx U_i^{\theta}(\theta_0;\gamma_0) + I_i^{\theta\theta}(\hat{\theta};\hat{\gamma})(\hat{\theta}-\theta_0)$$

and

$$\hat{\theta} - \theta_0 \approx -I^{\theta\theta}(\hat{\theta};\hat{\gamma})^{-1}U^{\theta}(\theta_0;\gamma_0)$$

that

$$
\begin{aligned}
U_i^{\theta}(\hat{\theta};\hat{\gamma}) &\approx U_i^{\theta}(\theta_0;\gamma_0) - I_i^{\theta\theta}(\hat{\theta};\hat{\gamma})I^{\theta\theta}(\hat{\theta};\hat{\gamma})^{-1}[U_i^{\theta}(\theta_0;\gamma_0) + U_{-i}^{\theta}(\theta_0;\gamma_0)] \\
&= [I - I_i^{\theta\theta}(\hat{\theta};\hat{\gamma})I^{\theta\theta}(\hat{\theta};\hat{\gamma})^{-1}]U_i^{\theta}(\theta_0;\gamma_0) - I_i^{\theta\theta}(\hat{\theta};\hat{\gamma})I^{\theta\theta}(\hat{\theta};\hat{\gamma})^{-1}U_{-i}^{\theta}(\theta_0;\gamma_0)
\end{aligned}
$$

and

$$
\begin{aligned}
&E[U_i^\theta(\hat\theta;\hat\gamma)U_i^\theta(\hat\theta;\hat\gamma)'] \\
&\approx [I - I_i^{\theta\theta}(\hat\theta;\hat\gamma)I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1}]\mathrm{cov}[U_i^\theta(\theta_0;\gamma_0)][I - I_i^{\theta\theta}(\hat\theta;\hat\gamma)I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1}]' \\
&\quad + I_i^{\theta\theta}(\hat\theta;\hat\gamma)I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1}\sum_{j\neq i}\mathrm{cov}[U_j^\theta(\theta_0;\gamma_0)]I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1}I_i^{\theta\theta}(\hat\theta;\hat\gamma).
\end{aligned}
$$

Assuming that $\mathrm{cov}[U_i^\theta(\theta_0;\gamma_0)] \approx -cI_i^{\theta\theta}(\hat\theta;\hat\gamma)$ for all i this simplifies into

$$E[U_i^\theta(\hat\theta;\hat\gamma)U_i^\theta(\hat\theta;\hat\gamma)'] \approx [I - I_i^{\theta\theta}(\hat\theta;\hat\gamma)I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1}]\mathrm{cov}[U_i^\theta(\theta_0;\gamma_0)],$$

showing why the bias arises in the sandwich estimator. The bias correction proposed by Fay and Graubard (2001) approximates the score covariance with

$$\mathrm{cov}[U_i^\theta(\theta_0;\gamma_0)] \approx H_iU_i^\theta(\hat\theta;\hat\gamma)U_i^\theta(\hat\theta;\hat\gamma)'H_i',$$

where H_i is a diagonal matrix with elements $[1-\min\{b, (I_i^{\theta\theta}(\hat\theta;\hat\gamma)I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1})_{jj}\}]^{-1/2}$. Here b is a constant chosen by the user. The bias-corrected sandwich estimator then becomes

$$V[\hat\theta] \approx I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1}\left[\sum_{i=1}^n H_iU_i^\theta(\hat\theta;\hat\gamma)U_i^\theta(\hat\theta;\hat\gamma)'H_i'\right]I^{\theta\theta}(\hat\theta;\hat\gamma)^{-1'}. \tag{4}$$

References

Fay, M. P. and Graubard, B. I. (2001). Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics*, 57:1198–1206.

Robins, J. M., Mark, S. D., and Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48:479–495.