Manual variance correction for 2SDD

John Gardner

Note: These variances assume that the group and period effects can be estimated consistently. If the outcome variable is measured at the group level, convert it to a group-level rate, bootstrap the standard errors, or use one of the many excellent alternative estimators that are available.

The baseline case

Let

$$g(\gamma,\theta) = (Y - X_1'\gamma - \theta D)D = \varepsilon_1 D$$

and

$$m(\theta) = (Y - X_1'\gamma)X_1'(1-D) = \varepsilon_2 X_1'(1-D)$$

denote the second- and first-stage moments, and also define

$$\begin{aligned}
G_\theta &= -E(D^2) = -E(D) \\
G_\gamma &= -E(X_1'D) \\
\psi &= E[X_1X_1'(1-D)]^{-1}(Y - X_1'\theta)X'(1-D).
\end{aligned}$$

From Theorem 6.1 of Newey and McFadden (1994), the asymptotic variance of the two-stage estimator is

$$V = G_\theta^{-1}E\left[(g + G_\gamma\psi)(g + G_\gamma\psi)'\right]G_\theta^{-1'}.$$

In this case,

$$\begin{aligned}
V = E(D)^{-1}E\Big\{ &\left[\varepsilon_1 D - E(X_1'D)E[X_1X_1'(1-D)]^{-1}\varepsilon_1 X_1'(1-D)\right] \\
&\left[\varepsilon_1 D - E(X_1'D)E[X_1X_1'(1-D)]^{-1}\varepsilon_1 X_1'(1-D)\right]' \Big\}E(D)^{-1}.
\end{aligned}$$

Since the moment conditions are uncorrelated, so that $E(mg) = 0$, this simplifies to

$$\begin{aligned}
&E(D)^{-1}E\left[\varepsilon_1^2 D\right]E(D)^{-1'} + \\
&\quad E(D)^{-1}E(X_1'D)E[X_1X_1'(1-D)]^{-1}E\left\{\varepsilon_1^2 X_1X_1'(1-D)\right\}E[X_1X_1'(1-D)]^{-1'}E(X_1'D)'E(D)^{-1} \\
&\qquad = V_2 + CVC',
\end{aligned}$$

where $V_2 = E(\varepsilon^2 D)/E(D)^2$ is the (uncorrected) second-stage variance, $V_1 = E\left\{\varepsilon_1^2 X_1X_1'(1-D)\right\}E[X_1X'(1-D)]^{-1}$ is the first-stage variance, and $C = E(X_1'D)/E(D) = E(X_1'D)/E(D^2)$ is the matrix of coefficients from regressions of the elements of X_1 on D.

This can be estimated by replacing V_1 and V_2 with estimated (uncorrected) variance matrices.

Note that since X_1 and D only vary at the cluster level, sample means of expressions involving those variables can already be interpreted as cross-cluster expectations). This approach is also valid when the second-stage includes treatment-duration indicators (in which case the elements of C are coefficients from regressions of X_1 onto all of the second-stage covariates).

The general case

When the first-stage uses all observations, or when the second stage includes leads of treatment status, the moment conditions may be correlated. For this case, define

$$\begin{aligned} g(\gamma,\theta) &= (Y - X_{12}'\gamma - X_2'\theta)X_2 \\ m(\gamma) &= (Y - X_{11}'\gamma)X_{11} \\ G_\theta &= -E(X_2X_2') \\ G_\gamma &= -E(X_2X_{12}') \\ \psi &= E(X_{11}X_{11}')^{-1}\varepsilon_1 X_{11}, \end{aligned}$$

where X_{11} are the regressors in the first stage, X_2 are the regressors in the second stage, and X_{12} are the regressors used to form the adjusted dependent variable. In the simple case where the first stage is estimated using only untreated observations, we would have $X_{11} = X(1-D)$ and $X_{12} = X$, where X is the vector of group/period dummies (excluding whatever variable is omitted).

The asymptotic variance of $\hat{\theta}$ is

$$V = E(X_2X_2')^{-1}E\left\{\left[\underbrace{\varepsilon_2X_2}_{K_2\times 1} - \underbrace{E(X_2X_{12}')}_{K_2\times K_1}\underbrace{E(X_{11}X_{11}')^{-1}}_{K_1\times K_1}\underbrace{\varepsilon_1X_{11}}_{K_1\times 1}\right]\left[\varepsilon_2X_2 - E(X_2X_{12}')E(X_{11}X_{11}')^{-1}\varepsilon_1X_{11}\right]'\right\}E(X_2X_2')^{-1}$$

which can be estimated using

$$(\mathbf{X}_2'\mathbf{X}_2)^{-1}\left(\sum_g \mathbf{W}_g\mathbf{W}_g'\right)(\mathbf{X}_2'\mathbf{X}_2)^{-1}$$

where

$$\mathbf{W}_g = \underbrace{\mathbf{X}_{2g}'\hat{\varepsilon}_{2g}}_{K_2\times 1} - \underbrace{(\mathbf{X}_2'\mathbf{X}_{12})}_{K_2\times K_1}\underbrace{(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}}_{K_1\times K_1}\underbrace{\mathbf{X}_{11g}'\hat{\varepsilon}_{1g}}_{K_1\times 1}.$$

Doing this for the Autor example replicates Stata's GMM output (although it is common to apply a finite-sample adjustment to clustered estimates, see Cameron and Miller, 2015).

Corrected variances can also be estimated by noting that, in general,

$$V = V_2 + CV_1C' - 2E(X_2X_2')^{-1}E(X_2\varepsilon_2\varepsilon_1X_{11}')E(X_{11}X_{11}')C',$$

where $C = E(X_2X_2')^{-1}E(X_2X_{12}')$ is the $K_2 \times K_1$ matrix of coefficients from regressions of the elements of X_1 onto those of X_2, and V_1 and V_2. are defined as above. This is the same as in the baseline case, but with a third term corresponding to the correlation between the moments, and which can be estimated using

$$2\underbrace{(\mathbf{X}_2'\mathbf{X}_2)^{-1}}_{K_2\times K_2}\left(\sum_g \underbrace{\mathbf{X}_{2g}'\hat{\varepsilon}_{2g}}_{K_2\times 1}\underbrace{\hat{\varepsilon}_{1g}'\mathbf{X}_{11g}}_{1\times K_1}\right)\underbrace{(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}}_{K_1\times K_1}\underbrace{\mathbf{C}'}_{K_1\times K_2},$$

with $\mathbf{C} = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_{12}$.