Cluster-robust standard errors

Class Notes

Manuel Arellano

October 7, 2011

Let θ be a parameter vector identified from the moment condition

$$E\psi\left(w,\theta\right)=0$$

such that $\dim\left(\theta\right)=\dim\left(\psi\right)$. Consider a sample $\mathcal{W}=\{w_1,...,w_N\}$ and a consistent estimator $\widehat{\theta}$ that solves the sample moment conditions (up to a small order term).

The function $\psi\left(w,\theta\right)$ may denote orthogonality conditions in a just-identified GMM estimation problem or the first-order conditions in an m-estimation problem. Alternatively $\psi\left(w,\theta\right)$ may be regarded as a linear combination of a larger moment vector $\varphi\left(w,\theta\right)$ in an overidentified GMM estimation problem, so that $\psi\left(w,\theta\right)=B_0\varphi\left(w,\theta\right)$.

Assume that conditions for the following asymptotically linear representation of the scaled estimation error and asymptotic normality hold:

$$\sqrt{N}\left(\widehat{\theta}-\theta\right)=-D_0^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi\left(w_i,\theta\right)+o_p\left(1\right)\xrightarrow{d}\mathcal{N}\left(0,D_0^{-1}V_0D_0^{-1\prime}\right)$$

where $D_0=\partial E\psi\left(w,\theta\right)/\partial c'$ is the Jacobian of the population moments and V_0 is the limiting variance:

$$V_0=\lim_{N\to\infty}Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi\left(w_i,\theta\right)\right)=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\psi\left(w_i,\theta\right)\psi\left(w_j,\theta\right)'\right].$$

To get standard errors of $\widehat{\theta}$ we need estimates of D_0 and V_0. A natural estimate of D_0 is $\widehat{D}=\frac{1}{N}\sum_{i=1}^{N}\widehat{D}_i$ where $\widehat{d}_{ki}=\frac{1}{2\epsilon_N}\left[\psi\left(w_i,\widehat{\theta}+\epsilon_N\delta_k\right)-\psi\left(w_i,\widehat{\theta}-\epsilon_N\delta_k\right)\right]$ is the kth column of $\widehat{D}_i$, for some small $\epsilon_N>0$, and δ_k is the kth unit vector. As for V_0 the situation is different under independent or cluster sampling.

Independent sampling Denote $\psi_i=\psi\left(w_i,\theta\right)$ and $\widehat{\psi}_i=\psi\left(w_i,\widehat{\theta}\right)$. If the observations are independent $E\left(\psi_i\psi_j'\right)=0$ for $i\neq j$ so that

$$V_0=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(\psi_i\psi_i'\right).$$

If they are also identically distributed $V_0=E\left(\psi_i\psi_i'\right)$. Regardless of the latter a consistent estimate of V_0 under independence is

$$\widehat{V}=\frac{1}{N}\sum_{i=1}^{N}\widehat{\psi}_i\widehat{\psi}_i'.$$

A consistent estimate of the asymptotic variance of $\widehat{\theta}$ under independent sampling is:

$$\widehat{Var}\left(\widehat{\theta}\right)=\frac{1}{N}\widehat{D}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\psi}_i\widehat{\psi}_i'\right)\widehat{D}^{-1}. \quad (1)$$

Cluster sampling We now obtain a consistent estimate of the asymptotic variance of $\widehat{\theta}$ when the sample consists of H groups (or clusters) of M_h observations each ($N = M_1 + ... + M_H$) such that observations are independent across groups but dependent within groups, $H \to \infty$ and M_h is fixed for all h. For convenience we order observations by groups and use the double-index notation w_{hm} so that $\mathcal{W} = \{w_{11}, ..., w_{1M_1} \mid ... \mid w_{H1}, ..., w_{HM_H}\}$.

Under cluster sampling, letting $\overline{\psi}_h = \sum_{m=1}^{M_h} \psi_{hm}$ and $\widetilde{\psi}_h = \sum_{m=1}^{M_h} \widehat{\psi}_{hm}$ we have

$$V_0 = \lim_{N\to\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{h=1}^{H}\overline{\psi}_h\right) = \lim_{N\to\infty}\frac{1}{N}\sum_{h=1}^{H} E\left(\overline{\psi}_h\overline{\psi}_h'\right),$$

so that $E\left(\psi_i\psi_j'\right) = 0$ only if i and j belong to different clusters, or $E\left(\psi_{hm}\psi_{h'm'}'\right) = 0$ for $h \neq h'$. Thus, a consistent estimate of V_0 is

$$\widetilde{V} = \frac{1}{N}\sum_{h=1}^{H}\widetilde{\psi}_h\widetilde{\psi}_h'. \tag{2}$$

The estimated asymptotic variance of $\widehat{\theta}$ allowing unrestricted within-cluster correlation is therefore

$$\widetilde{Var}\left(\widehat{\theta}\right) = \frac{1}{N}\widehat{D}^{-1}\left(\frac{1}{N}\sum_{h=1}^{H}\widetilde{\psi}_h\widetilde{\psi}_h'\right)\widehat{D}^{-1}. \tag{3}$$

Note that (3) is of order $1/H$ whereas (1) is of order $1/N$.

Examples A panel data example is

$$\psi\left(w_{hm}, \theta\right) = x_{hm}\left(y_{hm} - x_{hm}'\theta\right)$$

where h denotes units and m time periods. If the variables are in deviations from means $\widehat{\theta}$ is the within-group estimator. In this case $\widehat{D}_{hm} = -x_{hm}x_{hm}'$ and $\widetilde{\psi}_h = \sum_{m=1}^{M_h} x_{hm}\widehat{u}_{hm}$ where $\widehat{u}_{hm}$ are within-group residuals. The result is the (large H, fixed M_h) formula for within-group standard errors that are robust to heteroskedasticity and serial correlation of arbitrary form in Arellano (1987):

$$\widetilde{Var}\left(\widehat{\theta}\right) = \left(\sum_{h=1}^{H}\sum_{m=1}^{M_h} x_{hm}x_{hm}'\right)^{-1}\sum_{h=1}^{H}\sum_{m=1}^{M_h}\sum_{s=1}^{M_h}\widehat{u}_{hm}\widehat{u}_{hs}x_{hm}x_{hs}'\left(\sum_{h=1}^{H}\sum_{m=1}^{M_h} x_{hm}x_{hm}'\right)^{-1}. \tag{4}$$

Another example is a τ-quantile regression with moments

$$\psi\left(w_{hm}, \theta_\tau\right) = x_{hm}\left[1\left(y_{hm} \leq x_{hm}'\theta_\tau\right) - \tau\right]$$

where $\widehat{D}_{hm} = \widehat{\omega}_{hm}x_{hm}x_{hm}'$, $\widehat{\omega}_{hm} = \mathbf{1}\left(\left|y_{hm} - x_{hm}'\widehat{\theta}\right| \leq \xi_N\right) / \left(2\xi_N\right)$,[1] $\widetilde{\psi}_h = \sum_{m=1}^{M_h} x_{hm}\widehat{u}_{hm}$ and $\widehat{u}_{hm} = 1\left(y_{hm} \leq x_{hm}'\widehat{\theta}\right) - \tau$. Cluster-robust standard errors for QR coefficients are obtained from:

$$\widetilde{Var}\left(\widehat{\theta}\right) = \left(\sum_{h=1}^{H}\sum_{m=1}^{M_h}\widehat{\omega}_{hm}x_{hm}x_{hm}'\right)^{-1}\sum_{h=1}^{H}\sum_{m=1}^{M_h}\sum_{s=1}^{M_h}\widehat{u}_{hm}\widehat{u}_{hs}x_{hm}x_{hs}'\left(\sum_{h=1}^{H}\sum_{m=1}^{M_h}\widehat{\omega}_{hm}x_{hm}x_{hm}'\right)^{-1}. \tag{5}$$

[1] This choice of $\widehat{D}$ corresponds to selecting an (i,k)-specific scaled ϵ_N given by ξ_N/x_{ik}.

Equicorrelated moments within-cluster The variance formula (3) is valid for arbitrary correlation patterns among cluster members. Here we explore the consequences for variance estimation of assuming that the dependence between any pair of members of a cluster is the same. Under the error-component structure

$$E\left(\psi_{hm}\psi_{hs}'\right) = \begin{cases} \Omega_\eta + \Omega_v & \text{if } m = s \\ \Omega_\eta & \text{if } m \neq s \end{cases}$$

we have $E\left(\overline{\psi}_h \overline{\psi}_h'\right) = M_h\Omega_v + M_h^2\Omega_\eta$. Letting $\overline{M}_\ell = \frac{1}{N}\sum_{h=1}^{H} M_h^\ell$, the limiting variance V_0 becomes

$$V_0 = \lim_{H\to\infty}\left(\Omega_v + \frac{\overline{M}_2}{\overline{M}_1}\Omega_\eta\right).$$

The natural estimator of Ω_v is the within-group variance $\widehat{\Omega}_v$. When all clusters are of the same size ($M_h = M$ for all h), the natural estimator of Ω_η is the between-group variance minus the within variance times $1/M$:[2]

$$\widehat{\Omega}_\eta = \frac{1}{H}\sum_{h=1}^{H}\frac{1}{M^2}\widetilde{\psi}_h\widetilde{\psi}_h' - \frac{1}{M}\widehat{\Omega}_v.$$

Noting that $\overline{M}_2/\overline{M}_1 = M$, it turns out that the plug-in estimate of V_0 is identical to $\widetilde{V}$ in (2):

$$\widehat{\Omega}_v + M\widehat{\Omega}_\eta = \frac{1}{HM}\sum_{h=1}^{H}\widetilde{\psi}_h\widetilde{\psi}_h'.$$

When cluster sizes differ, a method-of-moments estimator of Ω_η is a weighted between-group variance minus the within variance times a weighted average of $1/M_h$:

$$\widehat{\Omega}_\eta = \frac{1}{H}\sum_{h=1}^{H} w_h\left(\frac{1}{M_h^2}\widetilde{\psi}_h\widetilde{\psi}_h' - \frac{1}{M_h}\widehat{\Omega}_v\right)$$

for some weights w_h such that $\sum_{h=1}^{H} w_h = H$.[3] Once again, using $w_h = M_h^2/\overline{M}_2$ as weights, the plug-in estimate of V_0 is identical to $\widetilde{V}$:

$$\widehat{\Omega}_v + \frac{\overline{M}_2}{\overline{M}_1}\widehat{\Omega}_\eta = \frac{1}{H\overline{M}_1}\sum_{h=1}^{H}\widetilde{\psi}_h\widetilde{\psi}_h'$$

The conclusion is that imposing within-cluster equicorrelation is essentially innocuous for the purpose of calculating cluster-robust standard errors.

References

Arellano, M. (1987): "Computing Robust Standard Errors for Within-Group Estimators", *Oxford Bulletin of Economics and Statistics*, 49, 431-434.

Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press.

[2] The within-group sample variance is $\widehat{\Omega}_v = \frac{1}{H\left(\overline{M}_1 - 1\right)}\sum_{h=1}^{H}\sum_{m=1}^{M_h}\left(\psi_{hm} - \frac{\overline{\psi}_h}{M_h}\right)\left(\psi_{hm} - \frac{\overline{\psi}_h}{M_h}\right)'$. As for the between-group variance note that overall means are equal to zero at $\widehat{\theta}$ (Arellano, 2003, section 3.1 on error-component estimation).

[3] For example, $w_h = 1$, $w_h = M_h/\overline{M}_1$ or $w_h = M_h^2/\overline{M}_2$.