I assume you know, or are willing to read up on, the following core ideas:

1. Taylor's theorem
2. Law of large numbers
3. Central limit theorem
4. Slutsky's theorem

I will sketch each idea briefly verbally. This will be a good review for those of you who know these results and will also help those who do not yet know these results to get a sense of the underlying issues and how they relate to what we are covering today.

1 Data with Independence

Define a function which is continuous in the parameters θ, $\widetilde{h}(\theta, Z_i) \equiv h_i(\theta)$, and the average

$$h(\theta) = \frac{1}{n}\sum_i h_i(\theta) \tag{1}$$

For consistency we assume regularity conditions and the identifying assumption that

$$E\left[h_i(\theta)\right] \quad = \quad 0 \tag{2}$$

if and only if $\theta = \theta^*$. Motivated by the analogy principle, we define the estimator $\widehat{\theta}$ by

$$0 = h(\widehat{\theta}) \tag{3}$$

Consistency follows from the identifying assumption and continuity of probability limits.

For asymptotic normality, we strengthen continuity of $h(\cdot)$ to differentiability and assume regularity conditions and the existence of

$$V\left[\sqrt{n}h(\theta^*)\right] \quad = \quad \Sigma \tag{4}$$

By virtue of the iid assumption, we can put some structure on Σ. In particular, we know that

$$\Sigma \quad = \quad V\left[\sqrt{n}h(\theta^*)\right] = nV\left[h(\theta^*)\right] = n\frac{1}{n^2}\sum_i V\left[h_i(\theta^*)\right] = V\left[h_i(\theta^*)\right] \tag{5}$$

and we traditionally estimate

$$\widehat{\Sigma} \quad = \quad V\left[\sqrt{n}h(\widehat{\theta})\right] = \frac{1}{n-1}\sum_i \left(h_i(\widehat{\theta}) - h(\widehat{\theta})\right)\left(h_i(\widehat{\theta}) - h(\widehat{\theta})\right)' = \frac{1}{n-1}\sum_i h_i(\widehat{\theta})h_i(\widehat{\theta})' \tag{6}$$

Taking a first order Taylor approximation centering $h(\widehat{\theta})$ around θ^*, we have

$$0 \quad = \quad h(\widehat{\theta}) = h(\theta^*) + H(\theta^*)\left(\widehat{\theta} - \theta^*\right) + o_p(1) \tag{7}$$

where

$$H(\theta) \quad = \quad D_\theta h(\theta) = \frac{1}{n}\sum_i D_\theta h_i(\theta) \tag{8}$$

From the first order Taylor approximation and invertibility of $H(\theta^*)$, we have

$$\sqrt{n}\left(\widehat{\theta} - \theta^*\right) \quad = \quad -H(\theta^*)^{-1}\sqrt{n}h(\theta^*) + o_p(1) \tag{9}$$

By a weak law of large numbers, $H(\theta)$ converges in probability under regularity conditions to $H^*(\theta)$ for every value of θ. By a central limit theorem, $\sqrt{n}h(\theta^*)$ converges in distribution under regularity conditions to a normal distribution with mean zero and variance Σ. Consequently, under regularity conditions and by Slutsky's theorem, $\sqrt{n}\left(\widehat{\theta} - \theta^*\right)$ converges in distribution to a normal distribution with mean zero and variance

$$\Omega \quad = \quad H^*(\theta^*)^{-1}\Sigma H^*(\theta^*)'^{-1} \tag{10}$$

under regularity conditions. We then approximate

$$\widehat{\Omega} \quad = \quad H(\widehat{\theta})^{-1}\widehat{\Sigma}H(\widehat{\theta})'^{-1} \tag{11}$$

For example, if we are analyzing a weighted least squares problem with $h_i(\theta) = (Y_i - X_i'\theta)X_iW_i$, then

$$H(\theta) = D_\theta h(\theta) = -\frac{1}{n}\sum_i W_iX_iX_i' \quad \text{and} \quad \widehat{\Sigma} = \frac{1}{n-1}\sum_i X_iW_i^2\left(Y_i - X_i'\widehat{\theta}\right)^2 X_i' \tag{12}$$

and we approximate

$$\widehat{\Omega} = \left\{\frac{1}{n}\sum_i W_iX_iX_i'\right\}^{-1} \frac{1}{n-1}\sum_i X_iW_i^2\left(Y_i - X_i'\widehat{\theta}\right)^2 X_i' \left\{\frac{1}{n}\sum_i W_iX_iX_i'\right\}^{-1} \tag{13}$$

A similar approach can be used to approximate covariances. Maintaining the previous example, suppose that we want to approximate the covariance between weighted and unweighted least squares. Let $\widehat{\theta}$ continue to denote the weighted least squares estimate and let $\widehat{\theta}_{OLS}$ denote unweighted least squares. Then approximate

$$\sqrt{n}\widehat{C}\left[\widehat{\theta}\ ,\ \widehat{\theta}_{OLS}\right] = \left\{\frac{1}{n}\sum_i W_iX_iX_i'\right\}^{-1} \frac{1}{n-1}\sum_i X_iW_i\left(Y_i - X_i'\widehat{\theta}\right)\left(Y_i - X_i'\widehat{\theta}_{OLS}\right)X_i' \left\{\frac{1}{n}\sum_i X_iX_i'\right\}^{-1} \tag{14}$$

2 Data with a Grouping Structure

Now suppose

$$h(\theta) = \frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j} h_{ji}(\theta) \tag{15}$$

and define $H(\theta) = D_\theta h(\theta) = \frac{1}{n}\sum_j\sum_i h_{ji}(\theta)$ as before. We now replace the iid assumption with the less restrictive assumption that $h_{ji}(\theta)$ is independent of $h_{j'i'}(\theta)$ for $j \neq j'$. Note that this means $h_{ji}(\theta)$ is not assumed to be independent of $h_{ji'}(\theta)$ for $i \neq i'$. This leads to a different approximation for Σ. Note that

$$\Sigma = V\left[\sqrt{n}h(\theta^*)\right] = nV\left[h(\theta^*)\right] = n\frac{1}{n^2}\sum_{j=1}^{J}V\left[\sum_{i=1}^{n_j}h_{ji}(\theta^*)\right] = \frac{1}{n}\sum_{j=1}^{J}V\left[\sum_{i=1}^{n_j}h_{ji}(\theta^*)\right] \tag{16}$$

Since we have independence across groups j, we can approximate

$$\widehat{V}\left[\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right] = \frac{1}{J-1}\sum_{j=1}^{J}\left\{\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right) - \frac{1}{J}\sum_{j=1}^{J}\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right)\right\}\left\{\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right) - \frac{1}{J}\sum_{j=1}^{J}\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right)\right\}' \tag{17}$$

$$= \frac{1}{J-1}\sum_{j=1}^{J}\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right)\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right)' \tag{18}$$

which leads to the estimate

$$\widehat{\Sigma} = \frac{1}{n}\frac{J}{J-1}\sum_{j=1}^{J}\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right)\left(\sum_{i=1}^{n_j}h_{ji}(\widehat{\theta})\right)' \tag{19}$$

and the limiting variance of $\sqrt{n}\left(\widehat{\theta} - \theta^*\right)$ can then be approximated as

$$\widehat{\Omega} = H(\widehat{\theta})^{-1}\widehat{\Sigma}H(\widehat{\theta})'^{-1} \tag{20}$$

It is instructive to rewrite this as follows. Define

$$g_j(\theta) = \sum_{i=1}^{n_j}h_{ji}(\theta) \quad \text{and} \quad G_j(\theta) = \sum_{i=1}^{n_j}H_{ji}(\theta) \tag{21}$$

This then implies

$$h(\theta) = \frac{1}{n}\sum_j g_j(\theta) \quad , \quad H(\theta) = \frac{1}{n}\sum_j G_j(\theta) \quad \text{and} \quad \widehat{\Sigma} = \frac{1}{n}\frac{J}{J-1}\sum_j g_j(\widehat{\theta})g_j(\widehat{\theta})' \tag{22}$$

and

$$\widehat{\Omega} = \left\{ \frac{1}{n} \sum_j G_j(\widehat{\theta}) \right\}^{-1} \frac{1}{n} \frac{J}{J-1} \sum_j g_j(\widehat{\theta}) g_j(\widehat{\theta})' \left\{ \frac{1}{n} \sum_j G_j(\widehat{\theta})' \right\}^{-1} \tag{23}$$

Note that the rank of the matrix $\widehat{\Omega}$ is $J-1$. This follows from the fact that each matrix $g_j(\widehat{\theta})g_j(\widehat{\theta})'$ is rank 1 and from the fact that the moments evaluated at $\widehat{\theta}$ by construction sum to zero overall in the sample. This means that it is impossible to test any hypothesis involving more than $J-1$ restrictions using this approach.

Specializing again to the weighted least squares case, we thus have

$$\begin{aligned} \widehat{\Omega} &= \left\{ \frac{1}{n} \sum_j \sum_i W_{ji} X_{ji} X_{ji}' \right\}^{-1} \frac{1}{n} \frac{J}{J-1} \sum_{j=1}^{J} \left(\sum_{i=1}^{n_j} W_{ji} (Y_{ji} - X_{ji}'\widehat{\theta}) X_{ji} \right) \left(\sum_{i=1}^{n_j} W_{ji} (Y_{ji} - X_{ji}'\widehat{\theta}) X_{ji} \right)' \\ &\times \left\{ \frac{1}{n} \sum_j \sum_i W_{ji} X_{ji} X_{ji}' \right\}^{-1} \end{aligned}$$

and the same approach can be used to approximate covariances. For weighted and unweighted least squares, approximate

$$\begin{aligned} \sqrt{n}\widehat{C}\left[\widehat{\theta}\ ,\ \widehat{\theta}_{OLS}\right] &= \left\{ \frac{1}{n} \sum_j \sum_i W_{ji} X_{ji} X_{ji}' \right\}^{-1} \frac{1}{n} \frac{J}{J-1} \sum_{j=1}^{J} \left(\sum_{i=1}^{n_j} W_{ji} (Y_{ji} - X_{ji}'\widehat{\theta}) X_{ji} \right) \left(\sum_{i=1}^{n_j} (Y_{ji} - X_{ji}'\widehat{\theta}_{OLS}) X_{ji} \right)' \\ &\times \left\{ \frac{1}{n} \sum_j \sum_i X_{ji} X_{ji}' \right\}^{-1} \end{aligned}$$

3 Panel Data

Now suppose we have data on groups over time with individuals observed within groups at any point in time. Time is indexed by t. We believe in independence across groups, but there may not be very many groups, i.e., J may be small. Redefine

$$h(\theta) = \frac{1}{n} \sum_{j=1}^{J} \sum_{t=1}^{T_j} \sum_{i=1}^{n_{jt}} h_{jti}(\theta) \equiv \frac{1}{n} \sum_{j=1}^{J} \sum_{t=1}^{T_j} g_{jt}(\theta) \tag{24}$$

where $g_{jt}(\theta) = \sum_{i=1}^{n_{jt}} h_{jti}(\theta)$ by analogy with the previous notation. If J is large, then the strategy from the previous section can be used; this allows for arbitrary serial correlation patterns. However, if J is small (below, say, 10), that strategy performs poorly in finite samples. When J is small it may thus be worth considering imposing time series assumptions as well in order to improve finite sample coverage. Intuitively, when the restrictions are correct, this replaces noisy estimates of zero with true zeros and results in improved finite sample performance.

Recall that for any variable ξ_{jt}, $V[\sum_t \xi_{jt}] = \sum_t \sum_\tau C[\xi_t, \xi_\tau]$. Then note that

$$\Sigma = V\left[\sqrt{n} h(\theta^*)\right] = \frac{1}{n} \sum_{j=1}^{J} V\left[\sum_{t=1}^{T_j} \sum_{i=1}^{n_{jt}} h_{jti}(\theta^*) \right] = \frac{1}{n} \sum_{j=1}^{J} V\left[\sum_{t=1}^{T_j} g_{jt}(\theta^*) \right] \tag{25}$$

$$= \frac{1}{n} \sum_{j=1}^{J} \sum_{t=1}^{T_j} \sum_{\tau=1}^{T_j} C\left[g_{jt}(\theta^*), g_{j\tau}(\theta^*)\right] \equiv \frac{1}{n} \sum_{j=1}^{J} T_j S_j \tag{26}$$

where S_j is defined as

$$S_j = \frac{1}{T_j} \sum_{t=1}^{T_j} \sum_{\tau=1}^{T_j} C\left[g_{jt}(\theta^*), g_{j\tau}(\theta^*)\right] \tag{27}$$

Following Newey and West (1987), define

$$\widehat{D}_{jm} = \frac{1}{T_j} \sum_{t=m+1}^{T_j} g_{jt}(\widehat{\theta}) g_{j,t-m}(\widehat{\theta})' \tag{28}$$

$$\widehat{S}_j = \widehat{D}_{j0} + \sum_{m=1}^{M} \left(1 - \frac{m}{M+1} \right) \left(\widehat{D}_{jm} + \widehat{D}_{jm}' \right) \tag{29}$$

where M is either fixed (Theorem 1) or increases slowly with the number of time series observations (Theorem 2). This approach approximates

$$\widehat{\Sigma} = \frac{1}{n}\sum_{j=1}^{J} T_j \widehat{S}_j = \frac{1}{n}\sum_{j=1}^{J} T_j \left\{ \widehat{D}_{j0} + \sum_{m=1}^{M} \left(1 - \frac{m}{M+1}\right)\left(\widehat{D}_{jm} + \widehat{D}'_{jm}\right)\right\} \tag{30}$$

or a linear combination of Newey-West variance matrices for the various groups.

Specializing this result to weighted least squares, we have

$$\widehat{\Omega} = \left\{\frac{1}{n}\sum_j\sum_t\sum_i W_{jti}X_{jti}X'_{jti}\right\}^{-1} \frac{1}{n}\sum_{j=1}^{J} T_j\widehat{S}_j \left\{\frac{1}{n}\sum_j\sum_t\sum_i W_{jti}X_{jti}X'_{jti}\right\}^{-1}$$

and as before, a similar approach works for covariances. For weighted and unweighted least squares, redefine $\widehat{D}_{jm}$ as

$$\widehat{D}_{jm} = \frac{1}{T_j}\sum_{t=m+1}^{T_j}\left(\sum_{i=1}^{n_{jt}} W_{jti}(Y_{jti} - X'_{jti}\widehat{\theta})X_{jti}\right)\left(\sum_{i=1}^{n_{jt}}(Y_{j,t-m,i} - X'_{j,t-m,i}\widehat{\theta}_{OLS})X_{j,t-m,i}\right)' \tag{31}$$

Then define $\widehat{S}_j$ as before and approximate

$$\sqrt{n}\widehat{C}\left[\widehat{\theta}\ ,\ \widehat{\theta}_{OLS}\right] = \left\{\frac{1}{n}\sum_j\sum_t\sum_i W_{jti}X_{jti}X'_{jti}\right\}^{-1} \frac{1}{n}\sum_{j=1}^{J} T_j\widehat{S}_j \left\{\frac{1}{n}\sum_j\sum_t\sum_i X_{jti}X'_{jti}\right\}^{-1}$$

Note that when $M = 0$ and setting aside degrees of freedom corrections, this approach reduces to the "cluster" approach from the preceding section, where independence is assumed across j and t. This is useful as a check on computing.