Optimal Weighting Matrix

Suppose $\{\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t)\}_{t=1}^{T}$ is strictly stationary and define

$\Gamma_\nu = E[\boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_t)\boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_{t-\nu})]$

and

$$S = \sum_{\nu=-\infty}^{\infty} \Gamma_\nu = \Gamma_0 + \sum_{\nu=1}^{\infty}(\Gamma_\nu + \Gamma_\nu^{'})$$

Asymptotic theory dictates

$\sqrt{T}(\boldsymbol{g}_T(\boldsymbol{\theta}) - E[\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t)]) \xrightarrow{d} N(0, S)$

or that

$$\sum_{t=1}^{T} \boldsymbol{g}_T(\boldsymbol{\theta})\boldsymbol{g}_T(\boldsymbol{\theta})^{'} \xrightarrow{p} S$$

Another way to say this (intuitively)

$\boldsymbol{g}_T(\boldsymbol{\theta}) \overset{approx}{\sim} N\left(E[\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{y}_t)], \frac{S}{T}\right)$

The optimal GMM weighting matrix is S^{-1}:

$Q_T(\boldsymbol{\theta}) = \boldsymbol{g}_T(\boldsymbol{\theta})^{'} S^{-1} \boldsymbol{g}_T(\boldsymbol{\theta})$

If $\{\boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_t)\}_{t=-\infty}^{\infty}$ is serially uncorrelated, S is consistently estimated by

$$S_T^* = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_t)\boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_t)^{'}$$

If it is serially correlated,

$$S_T^* = \Gamma_{0,T}^* + \sum_{\nu=1}^{q}\left(1 - \frac{\nu}{q+1}\right)\left(\Gamma_{\nu,T}^* + \Gamma_{\nu,T}^{*\,'}\right)$$

where

$$\Gamma_{\nu,T}^* = \frac{1}{T}\sum_{t=\nu+1}^{T} \boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_t)\boldsymbol{h}(\boldsymbol{\theta_0}, \boldsymbol{y}_{t-\nu})^{'}$$

Notice that S^* depends on $\boldsymbol{\theta}_0$, which is unknown

We substitute an estimate $\hat{\theta}$ for $\boldsymbol{\theta}_0$ in S^* and denote the estimated value as $\hat{S}$(where $\hat{S}$ may make use of appropriate definitions of $\hat{\Gamma}_{\nu,T}$ if there is serial correlation)

Under certain regularity conditions

$\hat{S} \xrightarrow{p} S$

Note that we want to use $\hat{S}^{-1}$ as the optimal weighting matrix to compute $\hat{\theta}$, but that $\hat{S}^{-1}$ depends on $\hat{\theta}$.

To compute optimal $\hat{\theta}_{gmm}$, first estimate $\hat{\theta}_{gmm}$ with $w_T = I_r$

Use the initial $\hat{\theta}_{gmm}$ to compute $\hat{S}_T(\hat{\theta}_{gmm})$ and set $w_T = \hat{S}_T(\hat{\theta}_{gmm})^{-1}$

Compute $\hat{\theta}_{gmm}$ again.

How is the two-stage procedure better?

That is, why is S^{-1} optimal?

Using S^{-1} or a consistent estimate $\hat{S}^{-1}$ results in $\hat{\theta}_{gmm}$ with less estimation error.

<u>Asymptotic distribution of GMM estimator</u>

A central limit theorem exists for $\hat{\theta}_{gmm}$:

$\sqrt{T}(\boldsymbol{\theta}_{gmm} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, V)$

where: $V = (DS^{-1}D')^{-1} \dfrac{\partial \boldsymbol{g}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{p} D$

That is,

$\hat{\theta}_{gmm} \overset{approx}{\sim} N(\boldsymbol{\theta}_0, \dfrac{\hat{V}_T}{T})$ for large T.

where,

$\hat{V}_T = (\hat{D}_T \hat{S}_T^{-1} \hat{D}_T')^{-1}$

$\hat{D}_T = \dfrac{\partial \boldsymbol{g}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Bigg|_{\boldsymbol{\theta}=\hat{\theta}}$