EC824A
Extremum Estimators - Covariance Matrix Estimation

Department of Economics
Michigan State University

Extremum Estimators – Covariance Matrix Estimation

- Consistent estimation of asymptotic covariance matrix $B_0^{-1}\Omega_0 B_0^{-1}$ for $\widehat{\theta}_n$
- Lemma 1 from last lecture with Assumption EE(i) and CF(i) yield

$$\widehat{B}_n^{-1} = \left(\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\widehat{\theta}_n) \right)^{-1} \xrightarrow{p} B_0^{-1}$$

  so focus on consistent estimation of Ω_0
- Do the usual:
  1. replace expectations with sample averages
  2. replace unknown parameters with consistent estimators

1. Consistent Covariance Estimation

2. Optimal Weights (GMM, MD, TS)

Consistent Covariance Estimation - ML

- Let
$$\begin{aligned} \widehat{B}_n &= -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta\partial\theta'} \log f(W_i, \widehat{\theta}_n) \text{ and} \\ \widehat{\Omega}_n &= \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f(W_i, \widehat{\theta}_n)\frac{\partial}{\partial\theta'} \log f(W_i, \widehat{\theta}_n). \end{aligned}$$
- Obtain $\widehat{\Omega}_n \xrightarrow{p} \Omega_0$ by verifying conditions (i), (ii), and (iii) of Lemma 1 in last lecture
- Condition (i) holds by consistency of $\widehat{\theta}_n$
- Conditions (ii) and (iii) hold by the U-WCON of Theorem 3 in Lecture Note 1 provided $\frac{\partial}{\partial\theta} \log f(w,\theta)$ (or equivalently $f(w,\theta)$ and $\frac{\partial}{\partial\theta} f(w,\theta)$) is continuous in θ on Θ_0 $\forall w \in \mathcal{W}$ and
$$E\left[\sup_{\theta\in\Theta_0}\left\|\frac{\partial}{\partial\theta}\log f(w,\theta)\right\|^2\right] < \infty,$$
where Θ_0 is a compact neighborhood of θ_0.

Consistent Covariance Estimation - ML

- If the model is correctly specified, then $B_0 = \Omega_0$ and the covariance matrix $B_0^{-1}\Omega_0 B_0^{-1}$ can be estimated by $\widehat{B}_n^{-1}\widehat{\Omega}_n\widehat{B}_n^{-1}$, $\widehat{B}_n^{-1}$, or $\widehat{\Omega}_n^{-1}$
- $\widehat{\Omega}_n$ requires calculation of the first derivative of $f(w, \theta)$, whereas $\widehat{B}_n$ requires calculation of the second derivatives - for computational reasons, it is better to use $\widehat{\Omega}_n$

Consistent Covariance Estimation - LS

- Let
$$\widehat{B}_n = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(\frac{\partial}{\partial\theta}g(X_i,\widehat{\theta}_n)\frac{\partial}{\partial\theta'}g(X_i,\widehat{\theta}_n) - (Y_i - g(X_i,\widehat{\theta}_n))\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta'}g(X_i,\widehat{\theta}_n)\right) \text{ and}$$
$$\widehat{\Omega}_n = \frac{1}{n}\sum\nolimits_{i=1}^{n} U_i^2\frac{\partial}{\partial\theta}g(X_i,\widehat{\theta}_n)\frac{\partial}{\partial\theta'}g(X_i,\widehat{\theta}_n), \text{ where } \widehat{U}_i = Y_i - g(X_i,\widehat{\theta}_n).$$
- Obtain $\widehat{\Omega}_n \xrightarrow{p} \Omega_0$ by verifying three conditions (i), (ii), and (iii) of Lemma 1 (Lecture 2)
- Condition (i) holds by consistency of $\widehat{\theta}_n$
- Conditions (ii) and (iii) hold by Theorem 3 (Lecture 1) provided $g(x,\theta)$ and $\frac{\partial}{\partial\theta}g(x,\theta_0)$ are continuous in θ on Θ_0 $\forall x \in \mathcal{X}$ and
$$E\left[\sup_{\theta\in\Theta_0}\left\|(Y_i - g(X_i,\theta))\frac{\partial}{\partial\theta}g(X_i,\theta)\right\|^2\right] < \infty,$$
where Θ_0 is a compact neighborhood of θ_0.

Consistent Covariance Estimation - LS

- Correctly specified (i.e., $E[Y|X] = g(X, \theta_0)$ a.s.), then instead of $\widehat{B}_n$ use,

$$\widetilde{B}_n = \frac{1}{n} \sum\nolimits_{i=1}^{n} \frac{\partial}{\partial \theta} g(X_i, \widehat{\theta}_n) \frac{\partial}{\partial \theta'} g(X_i, \widehat{\theta}_n)$$

- In correctly specified case, $\widetilde{B}_n \xrightarrow{p} B_0$ when CF(iv) holds.
- Hence, consistent covariance matrix estimator (correct model) is,

$$\widetilde{B}_n^{-1} \widehat{\Omega}_n \widetilde{B}_n^{-1}.$$

  Note that by design this estimator allows for conditional heteroskedasticity
- If model correctly specified and errors conditionally homoskedastic, then $\Omega_0 = \sigma^2 B_0$ and $\widehat{\Omega}_n$ can be replaced by

$$\widehat{\sigma}^2 \widetilde{B}_n^{-1}, \text{ where } \widehat{\sigma}^2 = \frac{1}{n} \sum\nolimits_{i=1}^{n} (Y_i - g(X_i, \widehat{\theta}_n))^2.$$

  For linear regression, $\widehat{\sigma}^2 \widetilde{B}_n^{-1}$ equals to $\widehat{\sigma}^2 \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1}$.

Consistent Covariance Estimation - GMM

- Let

$$
\begin{aligned}
\widehat{\Gamma}_n &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial m(W_i,\widehat{\theta}_n)}{\partial \theta'}, \widehat{V}_n = \frac{1}{n}\sum_{i=1}^{n} m(W_i,\widehat{\theta}_n)m(W_i,\widehat{\theta}_n)' \\
\widehat{B}_n &= \widehat{\Gamma}'_n A'_n A_n \widehat{\Gamma}_n \text{ and } \widehat{\Omega}_n = \widehat{\Gamma}'_n A'_n A_n \widehat{V}_n A'_n A_n \widehat{\Gamma}_n.
\end{aligned}
$$

- Note $\widehat{B}_n$ does not include $\frac{\partial^2 Q_n(\widehat{\theta}_n)}{\partial\theta\partial\theta'}$ – this is under correct specification (the second term converges in probability to zero since $E[m(W_i,\theta_0)] = 0$)
- Each component of $\widehat{\Omega}_n$ can be shown (as in Lecture 2) to converge in probability to the corresponding component of Ω_0
- $\widehat{V}_n$ converges to $V_0 = E[m(W_i,\theta_0)m(W_i,\theta_0)']$ by Lemma 1 (Lecture 2) and Theorem 3 (Lecture 1) provided that $\widehat{\theta}_n \xrightarrow{p} \theta_0$, $m(w,\theta)$ is continuous in θ on Θ_0 $\forall w \in \mathcal{W}$ and $E\left[\sup_{\theta\in\Theta_0} \|m(W_i,\theta)\|^2\right] < \infty$.

Consistent Covariance Estimation - MD

- Let
$$\begin{aligned}\widehat{B}_n &= \left(\frac{\partial}{\partial\theta'}g(\widehat{\theta}_n)\right)' A_n' A_n \frac{\partial}{\partial\theta'}g(\widehat{\theta}_n) \text{ and} \\ \widehat{\Omega}_n &= \left(\frac{\partial}{\partial\theta'}g(\widehat{\theta}_n)\right)' A_n' A_n \widehat{V}_n A_n' A_n \frac{\partial}{\partial\theta'}g(\widehat{\theta}_n),\end{aligned}$$
  where $\widehat{V}_n$ consistent estimator of V_0, asymptotic covariance matrix of $\sqrt{n}\,(\widehat{\pi}_n - \pi_0)$
- The definition of $\widehat{B}_n$ does not include $\frac{\partial^2 Q_n(\widehat{\theta}_n)}{\partial\theta\partial\theta'}$, because the latter converges in probability to zero given that $\pi_0 = g(\theta_0)$
- Each component of $\widehat{\Omega}_n$ can be shown (as in Lecture 2) to converge in probability to the corresponding component of Ω_0. $\widehat{V}_n \xrightarrow{p} V_0$ is directly assumed.

Consistent Covariance Estimation - TS

- Let

$$
\begin{aligned}
\widehat{B}_n &= \left(\frac{\partial}{\partial\theta'}G_n(\widehat{\theta}_n,\widehat{\tau}_n)\right)' A_n' A_n \frac{\partial}{\partial\theta'}G_n(\widehat{\theta}_n,\widehat{\tau}_n) \text{ and} \\
\widehat{\Omega}_n &= \left(\frac{\partial}{\partial\theta'}G_n(\widehat{\theta}_n,\widehat{\tau}_n)\right)' A_n' A_n \left(\widehat{V}_{1n} + \widehat{\Lambda}_n \widehat{V}_{2n}' + \widehat{V}_{2n}\widehat{\Lambda}_n' + \widehat{\Lambda}_n \widehat{V}_{3n}\widehat{\Lambda}_n'\right) \\
&\quad \times A_n' A_n \frac{\partial}{\partial\theta'}G_n(\widehat{\theta}_n,\widehat{\tau}_n), \text{ where} \\
\widehat{\Lambda}_n &= \frac{\partial}{\partial\tau'}G_n(\widehat{\theta}_n,\widehat{\tau}_n)
\end{aligned}
$$

  and $\widehat{V}_{jn}$ is some consistent estimator of V_{j0} for $j = 1, 2, 3$
- If Λ_0 is zero, then we can take $\widehat{\Lambda}_n = 0$ and the estimators $\widehat{V}_{2n}$ and $\widehat{V}_{3n}$ are not required.

Optimal Weights

- GMM, MD, TS asymptotic variance of form,

$$\left(\Gamma_0' C \Gamma_0\right)^{-1} \Gamma_0' C \Sigma_0 C \Gamma_0 \left(\Gamma_0' C \Gamma_0\right)^{-1},$$

- $C = A'A$ and Σ_0 symmetric positive semi-definite, depends on the estimator
- Formally show optimal weight matrix A_n is such that

$$A'A = \Sigma_0^{-1}, \text{ where } A_n \xrightarrow{p} A. \tag{1}$$

  Optimal choice means asymptotic covariance matrix of $\widehat{\theta}_n$ minimised
- When (1) holds, asymptotic covariance matrix simplifies to $\left(\Gamma_0' \Sigma_0^{-1} \Gamma_0\right)^{-1}$. Then,

$$\left(\Gamma_0' C \Gamma_0\right)^{-1} \Gamma_0' C \Sigma_0 C \Gamma_0 \left(\Gamma_0' C \Gamma_0\right)^{-1} - \left(\Gamma_0' \Sigma_0^{-1} \Gamma_0\right)^{-1} \geqslant 0, \tag{2}$$

  where "$\geqslant 0$" denotes "is psd."

Optimal Weights

- First note that $F^{-1} - G^{-1} \geqslant 0$ if and only if $G - F \geqslant 0$. Thus, (2) is equivalent to

$$\Gamma_0' \Sigma_0^{-1} \Gamma_0 - \Gamma_0' C \Gamma_0 \left(\Gamma_0' C \Sigma_0 C \Gamma_0 \right)^{-1} \Gamma_0' C \Gamma_0 \geqslant 0. \tag{3}$$

- Define $H := \Gamma_0' \Sigma_0^{-1/2}$, $M := I_k - \Sigma_0^{1/2} C \Gamma_0 \left(\Gamma_0' C \Sigma_0 C \Gamma_0 \right)^{-1} \Gamma_0' C \Sigma_0^{1/2}$
- The left-hand side of (3):

$$\begin{aligned} & \Gamma_0' \Sigma_0^{-1/2} \left[I_k - \Sigma_0^{1/2} C \Gamma_0 \left(\Gamma_0' C \Sigma_0 C \Gamma_0 \right)^{-1} \Gamma_0' C \Sigma_0^{1/2} \right] \Sigma_0^{-1/2} \Gamma_0 \\ = \quad & HMH' \\ = \quad & HM(HM)' \geqslant 0, \end{aligned}$$

- Second equality from M being a projection matrix (symmetric and idempotent, $M^2 = M$)
- Matrix of form $HM(HM)'$ is psd, since $z'HM(HM)'z = \|(MH')z\|^2 \geqslant 0 \ \forall z \in \mathbb{R}^d$.

Optimal Weights

- For GMM and MD estimators, $\Sigma_0 = V_0$ and optimal weight matrix A_n such that

$$A_n' A_n \xrightarrow{p} A' A = V_0^{-1}.$$

- For TS estimator, optimal weight matrix A_n such that

$$A_n' A_n \xrightarrow{p} A' A = \left(V_{10} + \Lambda_0 V_{20}' + V_{20} \Lambda_0' + \Lambda_0 V_{30} \Lambda_0' \right)^{-1}.$$