TA section 2

April 10th, 2009 by Yang

1. Big Picture

We look for the conditions that guarantee the asymptotic normality of extremum estimators. For example, consider a class of m-estimators that minimize some kind of sample mean

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(W_i, \theta)$$

Then,

Asymptotic Normality

↑ **Assumption CF** ← (i)–(iv) below

- (i) $\theta_0 : int(\Theta)$
- (ii) $Q_n(\theta) \in C^2(\Theta_0)$
- (iii) $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) \xrightarrow{d} N(0, \Omega_0)$
- (iv) $\sup_{\theta\in\Theta_0}\|\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta) - B(\theta)\| \xrightarrow{p} 0$ where $B(\theta)$ is continous at θ_0 and $B(\theta_0)$ is nonsingular.

↑ $(\star)$

- (1) $\theta_0 : int(\Theta)$
- (2) $m(w, \theta) \in C^2(\Theta_0)$
- (3) $m(w, \theta)$ is integrable in w.
- (4) Data (W_i) are iid.
- (5) $E\sup_{\theta\in\Theta_0}\|\frac{\partial}{\partial\theta}m(W_i, \theta)\| < \infty$
- (6) $E\|\frac{\partial}{\partial\theta}m(W_i, \theta_0)\|^2 < \infty$
- (7) $E\sup_{\theta\in\Theta_0}\|\frac{\partial^2}{\partial\theta\partial\theta'}m(W_i, \theta)\| < \infty$
- (8) $E\frac{\partial^2}{\partial\theta\partial\theta'}m(W_i, \theta_0)$ is nonsingular.

↑ **Assumption EE2** ← (i)–(ii) below

- (i) $\widehat{\theta}_n \xrightarrow{p} \theta_0$
- (ii) $\frac{\partial}{\partial\theta}Q_n(\widehat{\theta}_n) = o_p(n^{-1/2})$

You may skip $(\star)$ if you are not interested in detailed primitive assumptions of m-estimators case. (1)-(8) are the primitive assumptions that guarantee Assumption CF. More specifically, (2) implies (ii). (2)-(5) together imply $E\frac{\partial}{\partial\theta}Q_n(\theta_0) = 0$. This along with (4) and (6) implies (iii) by CLT. Note that choice of Θ_0 guarantees compactness of Θ_0. (2)-(4) and (7)-(8) along with compactness of Θ_0 imply that (iv) holds. ML estimators and NLS estimators fall in the class of m-estimators, but GMM estimators do not.

2. Binary choice example

Consider the following model.

$$Y_i = \begin{cases} 1 & \text{if } X_i'\theta_0 - \varepsilon_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$\Pr(\varepsilon_i < t) = \frac{e^t}{1+e^t}$$

and the data $W_i = (Y_i, X_i)$ are iid, $i = 1, \cdots, n$. This is so-called Logit model. We can estimate θ using MLE or NLS.

2.1 MLE

We make use of the fact that the joint probability of Y_i is given by

$$f(Y_i, X_i; \theta) = \left(\frac{e^{X_i'\theta}}{1+e^{X_i'\theta}}\right)^{Y_i} \left(\frac{1}{1+e^{X_i'\theta}}\right)^{1-Y_i}$$

This follows from

$$\Pr(Y_i = 1|X_i) = \Pr(\varepsilon_i < X_i'\theta|X_i) = \frac{e^{X_i'\theta}}{1+e^{X_i'\theta}}$$

and the fact that Y_i follows a Bernoulli distribution with succes probability $\Pr(Y_i = 1|X_i)$. So MLE would solve

$$\max_\theta l(W_i; \theta) = \frac{1}{n}\sum_{i=1}^n \log f(Y_i, X_i; \theta)$$

or equivalently,

$$\min_\theta Q_n(\theta) := -\frac{1}{n}\sum_{i=1}^n \log f(Y_i, X_i; \theta) = -\frac{1}{n}\sum_{i=1}^n \left[Y_i X_i'\theta - \log(1+e^{X_i'\theta})\right]$$

where

$$m(W_i, \theta) := -\log f(Y_i, X_i; \theta) = -Y_i X_i'\theta + \log(1+e^{X_i'\theta})$$

Now let us investigate whether the assumptions (1)-(8) are satisfied. (1) has to be assumed. (2) is satisfied, and thus (ii) is satisfied. (3) is true. (4) is assumed in the model. (5) requires $E||X_i|| < \infty$. (6) and (7) are satisfied if $E||X_i||^2 < \infty$. (8) is guaranteed if EX_iX_i' is nonsingular. To summarize, under the generous assumptions that $\theta_0 \in int(\Theta)$, that $E||X_i||^2 < \infty$ and that EX_iX_i' is nonsingular, Assumption CF is satisfied.

Remark. For the ML estimator of this model to be asymptotically normal, Assumption EE, Assumption EE2 (ii), iid data, compact Θ, $\theta_0 \in int(\Theta)$, $E||X_i||^2 < \infty$, and invertible EX_iX_i' are enough. To see this, note first that compact Θ and $E||X_i|| < \infty$ guarantee $\widehat{\theta}_n \xrightarrow{p} \theta_0$, and thus Assumption EE2 and Assumption CF are all satisfied, which imply that ML estimator $\widehat{\theta}_n$ would be asymptotically normal.

2.2 NLS

Here we make use of the fact that Y_i is a Bernoulli random variable, so

$$E[Y_i|X_i] = \Pr(Y_i = 1|X_i) = \frac{e^{X_i'\theta_0}}{1 + e^{X_i'\theta_0}}$$

As we know that conditional expectation minimizes the mean square error, it is natural to think of the following estimation method

$$\min_\theta \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \frac{e^{X_i'\theta}}{1 + e^{X_i'\theta}} \right)^2$$

Define

$$m(W_i, \theta) = \frac{1}{2} \left(Y_i - \frac{e^{X_i'\theta}}{1 + e^{X_i'\theta}} \right)^2$$

then the stochastic criterion function would be $Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} m(W_i, \theta)$. Now let us find the conditions that support the assumptions (1)-(8). (1) should be assumed. (2) and (3) are satisfied. (4) is assumed in the model. (5) is satisfied if $E||X_i|| < \infty$. (6) and (7) require $E||X_i||^2 < \infty$. (8) would be true if EX_iX_i' is invertible. So Assumption CF is guaranteed by the same assumptions that we made in the MLE case.

2.3 Covariance Matrix

We have

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, B_0^{-1}\Omega_0 B_0^{-1}\right)$$

In MLE case,

$$\begin{aligned}
\Omega_0 &= E \frac{\partial}{\partial \theta} m(W_i, \theta_0) \frac{\partial}{\partial \theta'} m(W_i, \theta_0) \\
&= E \left[\left(-Y_i X_i + \frac{e^{X_i'\theta_0} X_i}{1 + e^{X_i'\theta_0}} \right) \left(-Y_i X_i' + \frac{e^{X_i'\theta_0} X_i'}{1 + e^{X_i'\theta_0}} \right) \right] \\
&= E \left[\left(Y_i - \frac{e^{X_i'\theta_0}}{1 + e^{X_i'\theta_0}} \right)^2 X_i X_i' \right] \\
&= E \left[E \left[\left(Y_i - \frac{e^{X_i'\theta_0}}{1 + e^{X_i'\theta_0}} \right)^2 \middle| X_i \right] X_i X_i' \right] \\
&= E \left[Var(Y_i|X_i) X_i X_i' \right] \\
&= E \left[\frac{e^{X_i'\theta_0} X_i X_i'}{(1 + e^{X_i'\theta_0})^2} \right]
\end{aligned}$$

$$B_0 = E \frac{\partial^2}{\partial \theta \partial \theta'} m(W_i, \theta_0) = E \left[\frac{e^{X_i'\theta_0} X_i X_i'}{(1 + e^{X_i'\theta_0})^2} \right]$$

We can verify that $B_0 = \Omega_0$, so

$$\sqrt{n}(\widehat{\theta}_n^{\text{ML}} - \theta_0) \xrightarrow{d} N\left(0, E\left[\frac{e^{X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^2}\right]^{-1}\right)$$

The estimation of covariance matrix can be done by replacing expectation with sample mean, and θ_0 with $\widehat{\theta}_n$.

$$n\widehat{Var}(\widehat{\theta}_n) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{e^{X_i'\widehat{\theta}_n}X_iX_i'}{(1+e^{X_i'\widehat{\theta}_n})^2}\right)^{-1}$$

In NLS case,

$$\begin{aligned}
\Omega_0 &= E\frac{\partial}{\partial\theta}m(W_i,\theta_0)\frac{\partial}{\partial\theta'}m(W_i,\theta_0)\\
&= E\left[\left(-\left[Y_i - \frac{e^{X_i'\theta_0}}{1+e^{X_i'\theta_0}}\right]\frac{e^{X_i'\theta_0}X_i}{(1+e^{X_i'\theta_0})^2}\right)\left(-\left[Y_i - \frac{e^{X_i'\theta_0}}{1+e^{X_i'\theta_0}}\right]\frac{e^{X_i'\theta_0}X_i'}{(1+e^{X_i'\theta_0})^2}\right)\right]\\
&= E\left[\left(Y_i - \frac{e^{X_i'\theta_0}}{1+e^{X_i'\theta_0}}\right)^2\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4}\right]\\
&= E\left[Var(Y_i|X_i)\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4}\right]\\
&= E\left[\frac{e^{3X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^6}\right]\\
B_0 &= E\frac{\partial^2}{\partial\theta\partial\theta'}m(W_i,\theta_0)\\
&= E\left[\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4} - \left(Y_i - \frac{e^{X_i'\theta_0}}{1+e^{X_i'\theta_0}}\right)\frac{e^{X_i'\theta_0}(1-e^{X_i'\theta_0})X_iX_i'}{(1+e^{X_i'\theta_0})^3}\right]\\
&= E\left[\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4} - E\left[Y_i - \frac{e^{X_i'\theta_0}}{1+e^{X_i'\theta_0}}\middle| X_i\right]\frac{e^{X_i'\theta_0}(1-e^{X_i'\theta_0})X_iX_i'}{(1+e^{X_i'\theta_0})^3}\right]\\
&= E\left[\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4}\right]
\end{aligned}$$

Now we can see $\Omega_0 \neq B_0$. We have

$$\sqrt{n}(\widehat{\theta}_n^{\text{NLS}} - \theta_0) \xrightarrow{d} N\left(0, E\left[\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4}\right]^{-1} E\left[\frac{e^{3X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^6}\right] E\left[\frac{e^{2X_i'\theta_0}X_iX_i'}{(1+e^{X_i'\theta_0})^4}\right]^{-1}\right)$$

Note that the covariance matrix cannot be simplified further as long as X_i is stochastic. The covariance matrix would be estimated by

$$n\widehat{Var}(\widehat{\theta}_n) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{e^{2X_i'\widehat{\theta}_n}X_iX_i'}{(1+e^{X_i'\widehat{\theta}_n})^4}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{e^{3X_i'\widehat{\theta}_n}X_iX_i'}{(1+e^{X_i'\widehat{\theta}_n})^6}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{e^{2X_i'\widehat{\theta}_n}X_iX_i'}{(1+e^{X_i'\widehat{\theta}_n})^4}\right)^{-1}$$

3. Linear IV example

Consider linear IV model as follows

$$y_i = x_i'\beta_0 + \varepsilon_i$$
$$x_i = z_i'\pi_0 + u_i$$

where

$$Ex_i\varepsilon_i \neq 0$$
$$Ez_i\varepsilon_i = 0$$

and the data are iid. As we have seen in the last section, using $A_n'A_n = \left(\frac{1}{n}z_i z_i'\right)^{-1}$ yields 2SLS estimator as

$$\widehat{\beta}_n = (X'P_Z X)^{-1}X'P_Z Y$$

(a) Let us begin with standard method of asymptotics.

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = (X'P_Z X)^{-1}\sqrt{n}X'P_Z(Y - X\beta_0)$$

Rewriting the RHS as

$$\left[\left(\frac{1}{n}\sum_{i=1}^n x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_i x_i'\right)\right]^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i\varepsilon_i\right)$$

would let us conclude that it converges in distribution to

$$\left[Ex_i z_i'(Ez_i z_i')^{-1}Ez_i x_i'\right]^{-1}Ex_i z_i'(Ez_i z_i')^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i\varepsilon_i\right)$$

under the assumption that $E||x_i z_i'|| < \infty$, $E||z_i||^2 < \infty$, and invertible $Ex_i z_i'(Ez_i z_i')^{-1}Ez_i x_i'$. By CLT,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i\varepsilon_i \xrightarrow{d} N\left(0, E\varepsilon_i^2 z_i z_i'\right)$$

and under conditional homoskedasticity assumption that $Var(\varepsilon_i|z_i) = \sigma^2$, the variance will be $\sigma^2 Ez_i z_i'$. So we can write as

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} N\left(0, \sigma^2\left[Ex_i z_i'(Ez_i z_i')^{-1}Ez_i x_i'\right]^{-1}\right)$$

(b) Now let us take GMM approach. Recall that $g(y_i, x_i, z_i, \beta) = z_i(y_i - x_i'\beta)$. Assumption CF (i) has to be assumed. CF (ii) is satisfied immediately. CF (iii) is implied by $E||x_i z_i'|| < \infty$, $E||z_i||^2 < \infty$, and conditional homoskedasticity. CF (iv) is implied by full column rank $Ez_i x_i'$ and $E||x_i z_i'|| < \infty$. (You may skip how to derive CF iii and iv from those assumptions.)

The statement that Ez_ix_i' has full column rank is equivalent to saying that $Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'$ is invertible. So the set of assumptions required for Assumption CF is the same as that in (a).

Now calculate Ω_0 and B_0 as follows. Note first that

$$\begin{aligned}
\Gamma_0 &= E\frac{\partial}{\partial\beta'}g(w_i,\beta_0) = Ez_ix_i' \\
A'A &= (Ez_iz_i')^{-1} \\
V_0 &= Eg(w_i,\beta_0)g(w_i,\beta_0)' = \sigma^2 Ez_iz_i' \\
[B(\beta)]_{m,j} &= E\frac{\partial}{\partial\beta_m}g(w_i,\beta)'A'AE\frac{\partial}{\partial\beta_j}g(w_i,\beta) + E\frac{\partial^2}{\partial\beta_m\partial\beta_j}g(w_i,\theta)'A'AEg(w_i,\beta)
\end{aligned}$$

and thus that

$$\begin{aligned}
\Omega_0 &= \Gamma_0'A'AV_0A'A\Gamma_0 = \sigma^2 Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i' \\
B_0 &= B(\beta_0) = \Gamma_0'A'A\Gamma_0 = Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'
\end{aligned}$$

By the theorem,

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} N\left(0, \sigma^2\left[Ex_iz_i'(Ez_iz_i')^{-1}Ez_ix_i'\right]^{-1}\right)$$

which is the same as the result in (a).

(c) The covariance matrix can be estimated by

$$\widehat{\sigma}^2\left[\frac{1}{n}\sum_{i=1}^n x_iz_i'\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^n z_ix_i'\right]^{-1}$$

where $\widehat{\sigma}^2$ is a consistent estimate of σ^2 with

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - x_i'\widehat{\beta}_n)^2$$

Remark. As we proved in the class, choice of $A_n'A_n = \left(\frac{1}{n}z_iz_i'\right)^{-1}$ is optimal under the conditional homoskedasticity, since $V_0 = \sigma^2 Ez_iz_i'$, and $A_n'A_n \xrightarrow{p} \sigma^2 V_0^{-1}$.

Remark. If the conditional homoskedasticity does not hold, the optimal choice of A_n would be different. It should be such that $A_n'A_n = (\frac{1}{n}\widehat{\varepsilon}_i^2 z_iz_i')^{-1}$, where $\widehat{\varepsilon}_i$ is a consistent estimate of ε_i. Also we need an additional assumption $E||z_i\varepsilon_i||^2 < \infty$ to guarantee asymptotic normality of the estimator.

Appendix. (Central Limit Theorem) Assume iid data and $E||s_i||^2 < \infty$. Then,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n s_i - Es_i\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (s_i - Es_i) \xrightarrow{d} N(0, Var(s_i))$$

where $Var(s_i) = E[(s_i - Es_i)(s_i - Es_i)']$.