# TA section 9

June 5th, 2009 by Yang

## 1. Final Exam 2005

**Q1**. **(i)** FALSE.
In the linear IV model, the OLS estimator is biased and not even consistent. We would care about the finite sample bias, which is why we are reluctant to use the 2SLS estimator, but using OLS instead of 2SLS is worse. 2SLS is at least consistent. We may estimate by 2SLS method and employ other test procedures such as LMcue. Other ways are using LIML estimation or Fuller-type estimation.

**(ii)** FALSE.
Recall PS5, Q1. As the number of instruments goes up, the asymptotic variance decreases in the positive definite sense.

**(iii)** QUESTIONABLE.
An MA($q$) process is weakly stationary, since its mean and autocovariance function do not depend on $t$. It is ergodic too, since its autocovariance is 0 if the lag is greater than $q$. But the weak stationarity does not imply ergodicity. 'Therefore' part is misleading.

**(iv)** TRUE.
See PS5, Q3 (v). A simpler proof is as follows. Suppose $1/X_n = o_p(1)$ and $X_n = O_p(1)$. By the theorem, $1/X_n \cdot X_n = o_p(1)O_p(1) = o_p(1)$. But this is equivalent to $1 = o_p(1)$, which is a contradiction.

**(v)** FALSE.
When the data are iid, HAC estimation is consistent even with $S_T = 0$. This is because the product of terms with lags more than 1 has expectation 0 when the data are iid.

**(vi)** TRUE.
The Wald test is known to sometimes overreject the true null. This means that the asymptotic distribution of the Wald statistic does not approximate the finite sample distribution very well, and that the critical value obtained by the asymptotic distribution is far less than it should be when the correct distribution is used. So we may obtain too narrow confidence interval when we use the Wald statistic. In other words, if we do inference by inverting a Wald statistic, the probability that the true null is contained in the confidence interval would be less than the nominal coverage probability.

**(vii)** FALSE.
The correct reason that a large value of the J-statistic is not reliable is that its finite sample distribution is not approximated by the $\chi^2$ distribution very well. In fact, inconsistency of the J-test against certain local alternatives makes us observe a too low J-statistic even when the null is false, which is not a

problem with rejecting the null when we observe a big J-statistic.

**(viii)** FALSE.
The optimal linear forecast of $Y_t$ based on $X_t$ is $\alpha' X_t = EY_tX_t'(EX_tX_t')^{-1}X_t$ which depends only on the first and the second moments regardless of the distribution of the process. 'Only if' part is incorrect.

**(ix)** QUESTIONABLE.
It depends on the situation. If we want a consistent estimator, we can still use the 2SLS estimator. When we look for an efficient estimator, the 2SLS estimator is not appropriate to use since it is not asymptotically efficient when the conditional homoskedasticity does not hold.

**(x)** TRUE.
The asymptotic variance of the sample median is $\frac{1}{4f(0)^2}$. This is actually $\frac{\pi\sigma^2}{2}$, which depends on $\sigma^2$.

**Q2**. **(i)** See PS8, Q2 (c). Try also the case where an AR(1) process is observed every third time period.

**(ii)** To prove that there is no stationary causal solution, see PS8, Q2 (a). Prove nonexistence of a stationary noncausl solution as follows. Suppose there is a stationary noncausal solution. From $y_t = y_{t-1} + \varepsilon_t$, we have $y_{t-1} = y_t - \varepsilon_t$. Square both sides and take expectation, then

$$Ey_{t-1}^2 = Ey_t^2 - 2Ey_t\varepsilon_t + E\varepsilon_t^2$$

By stationarity, $Ey_{t-1}^2 = Ey_t^2$ and by a noncausal solution, $Ey_t\varepsilon_t = 0$. So the above equation reads as $E\varepsilon_t^2 = 0$, which contradicts to $\sigma^2 > 0$.

**Q3**. See PS5, Q1.

**Q4**. For the asymptotic normality of extremum estimators, we need two sets of assumptions.

**Assumption CF**

(i) $\theta_0 \in int(\Theta)$
(ii) $Q_n(\theta) \in C^2(\Theta_0)$ for some neighborhood $\Theta_0 \subset \Theta$ of $\theta_0$ (wp1)
(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 a nonstochastic function $B(\theta)$ is continuous at $\theta_0$, and $B_0 := B(\theta_0)$ is nonsingular.

**Assumption EE2**

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

If the above assumptions hold, we have

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

How do those assumptions specialize in the MLE case?

- CF(i) has to be assumed.
- CF(ii): Set up $Q_n(\theta)$ as follows.

$$Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log f(W_i;\theta)$$

The condition is satisfied if $f(W_i;\theta)$ is twice continuously differentiable in $\theta$, or equivalently, if $f(W_i;\theta) \in C^2(\Theta_0)$.

- CF(iii): Obtain

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(W_i;\theta_0)$$

Note first that

$$E\frac{\partial}{\partial\theta}Q_n(\theta_0) = -E\frac{\partial}{\partial\theta}\log f(W_i;\theta_0) = -\frac{\partial}{\partial\theta}E\log f(W_i;\theta_0) = \frac{\partial}{\partial\theta}Q(\theta_0) = 0$$

where we should assume integrability of $\log f(W_i;\theta)$ in $W_i$, and $E\sup_{\theta\in\Theta_0}\|\frac{\partial}{\partial\theta}\log f(W_i;\theta)\| < \infty$ for interchangibility of integral and derivative signs. Then,

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(W_i;\theta_0) \xrightarrow{d} N\left(0, E\frac{\partial}{\partial\theta}\log f(W_i;\theta_0)\frac{\partial}{\partial\theta'}\log f(W_i;\theta_0)\right)$$

by CLT where we should assume iid data, and $E\|\frac{\partial}{\partial\theta}\log f(W_i;\theta_0)\|^2 < \infty$. Then,

$$\Omega_0 := E\frac{\partial}{\partial\theta}\log f(W_i;\theta_0)\frac{\partial}{\partial\theta'}\log f(W_i;\theta_0)$$

- CF(iv): Obtain

$$\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta'}\log f(W_i;\theta)$$

Define $B(\theta) := -E\frac{\partial^2}{\partial\theta\partial\theta'}\log f(W_i;\theta_0)$, then we have

$$\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta) - B(\theta)\right\| \xrightarrow{p} 0$$

by uniform WLLN under the additional assumption that $E\sup_{\theta\in\Theta_0}\|\frac{\partial^2}{\partial\theta\partial\theta'}\log f(W_i;\theta)\| < \infty$. Finally, continuity of $B(\theta)$ at $\theta_0$ is guaranteed by the assumptions made already, and nonsingularity of $B_0 := B(\theta_0)$ has to be assumed.

- EE(i) has to be assumed here, although we can find more primitive conditions implying this.
- EE(ii) has to be assumed.

In fact, in the MLE case, $B_0 = \Omega_0$ by the information equality theorem under another condition guaranteeing interchangibility of integral and derivative signs in its proof. Then, we do not need to assume nonsingularity of $B_0$ separately. This leads us to the simpler result

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

**Q5.** Consider a VAR($p$) model.

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where $\varepsilon \sim N(0, \Omega)$. When some of the parameters are assumed to be 0, we may want to focus on estimating nonzero parameters. Write the above model as

$$\begin{aligned} y_{1t} =& x'_{1t}\beta_1 + \varepsilon_{1t} \\ &\vdots \\ y_{nt} =& x'_{nt}\beta_n + \varepsilon_{nt} \end{aligned}$$

where each equation corresponds to the respective element of the vectors in the original model, and the restricted parameters are dropped out along with the corresponding data. Write the system of equations as one equation as follows.

$$\underbrace{\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix}}_{=y_t} = \underbrace{\begin{pmatrix} x'_{1t} & 0 & \cdots & 0 \\ 0 & x'_{2t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & x'_{nt} \end{pmatrix}}_{=:\mathcal{X}'_t} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}}_{=:\beta} + \underbrace{\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix}}_{=\varepsilon_t}$$

We maximize

$$\mathcal{L}(\beta, \Omega) = -\frac{Tn}{2}\log 2\pi + \frac{T}{2}\log|\Omega^{-1}| - \frac{1}{2}\sum_{t=1}^{T}(y_t - \mathcal{X}'_t\beta)'\Omega^{-1}(y_t - \mathcal{X}'_t\beta)$$

The first order condition with respect to $\beta$ is

$$\sum_{t=1}^{T}\mathcal{X}_t\Omega^{-1}(y_t - \mathcal{X}'_t\beta) = 0$$

Assuming that we know $\Omega$, we obtain the following estimator.

$$\widehat{\beta} = \left(\sum_{t=1}^{T}\mathcal{X}_t\Omega^{-1}\mathcal{X}'_t\right)^{-1}\sum_{t=1}^{T}\mathcal{X}_t\Omega^{-1}y_t$$

This can also be written as

$$\widehat{\beta} = \left( \sum_{t=1}^{T} \sum_{i=1}^{n} \widetilde{x}_{it} \widetilde{x}_{it}' \right)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n} \widetilde{x}_{it} \widetilde{y}_{it}$$

where $L$ such that $L'L = \Omega^{-1}$ is used to define

$$\widetilde{x}_t' := \begin{pmatrix} \widetilde{x}_{1t}' \\ \vdots \\ \widetilde{x}_{nt}' \end{pmatrix} = L\mathcal{X}_t' \text{ and } \widetilde{y}_t := \begin{pmatrix} y_{1t} \\ \cdots \\ y_{nt} \end{pmatrix} = Ly_t$$

This is infeasible when we do not know $\Omega$. Another infeasible way is to solve the simultaneous equations of the FOCs. An easier and feasible way is using a two-step method.

1. Estimate the model using equation-by-equation OLS, and denote it by $\widehat{\beta}_{(0)}$.

$$\widehat{\beta}_{(0)} = \begin{pmatrix} \left( \sum_{t=1}^{T} x_{1t} x_{1t}' \right)^{-1} \sum_{t=1}^{T} x_{1t} y_{1t} \\ \vdots \\ \left( \sum_{t=1}^{T} x_{nt} x_{nt}' \right)^{-1} \sum_{t=1}^{T} x_{nt} y_{nt} \end{pmatrix}$$

2. Obtain $\widehat{\Omega}_{(0)}$ using $\widehat{\beta}_{(0)}$.

$$\widehat{\Omega}_{(0)} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \mathcal{X}_t' \widehat{\beta}_{(0)})(y_t - \mathcal{X}_t' \widehat{\beta}_{(0)})'$$

3. Use $\widehat{\Omega}_{(0)}$ to get $\widehat{\beta}_{(1)}$.

$$\widehat{\beta}_{(1)} = \left( \sum_{t=1}^{T} \sum_{i=1}^{n} \widetilde{x}_{it} \widetilde{x}_{it}' \right)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n} \widetilde{x}_{it} \widetilde{y}_{it}$$

We can iterate more, but this is enough in the sense that $\widehat{\beta}_{(1)}$ has the same asymptotic distribution as the infeasible MLE does (Magnus, 1978). Thus it is asymptotically efficient.

**Q6**. **(i)** (Only the bootstrap part) Suppose $X_1, \cdots, X_n$ are iid with mean $\mu$ and $\sigma^2$. By CLT, we have

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

where $\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$ is a sample mean. We estimate $\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$, then the $t$-statistic is given by

$$T_n = \frac{\sqrt{n}(\overline{X} - \mu)}{\widehat{\sigma}}$$

**A symmetric two-sided confidence interval using the bootstrap**

Obtain the bootstrap samples $(X_{bi}^*)_{i=1}^n$ by randomly drawing from $(X_i)_{i=1}^n$ with replacement. Estimate $\overline{X}_b^* := \frac{1}{n} \sum_{i=1}^{n} X_{bi}^*$ and $\widehat{\sigma}_b^{*2} = \frac{1}{n} \sum_{i=1}^{n} (X_{bi}^* - \overline{X}^*)^2$. The bootstrap $t$-statistic is given by

$$T_{nb}^* = \frac{\sqrt{n}(\overline{X}_b^* - \overline{X})}{\widehat{\sigma}_b^*}$$

Do this for $b = 1, \cdots, B$, then we have $B$ $t$-statistics which can be used to find a critical value. Specifically, order $|T^*_{nb}|$ from smallest to larget, and denote them by $|T^*_n|_{(b)}$. Define $\widehat{k}_{1-\alpha} := |T^*_n|_{(\llcorner(1-\alpha)(B+1)\lrcorner)}$. Since $\Pr(|T_n| \leq \widehat{k}_{1-\alpha}) \to 1-\alpha$,

$$\Pr\left(\overline{X} - \frac{\widehat{k}_{1-\alpha}\widehat{\sigma}}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{\widehat{k}_{1-\alpha}\widehat{\sigma}}{\sqrt{n}}\right) \to 1-\alpha$$

The symmetric two-sided confidence interval for $\mu$ is thus

$$\left[\overline{X} - \frac{\widehat{k}_{1-\alpha}\widehat{\sigma}}{\sqrt{n}}, \overline{X} + \frac{\widehat{k}_{1-\alpha}\widehat{\sigma}}{\sqrt{n}}\right]$$

**An equal-tailed two-sided confidence interval using the bootstrap**

Order $T^*_{nb}$ from smallest to largest, and denote them by $T^*_{n(b)}$. Let $\widehat{k}_{1-\alpha/2} := T^*_{n(\llcorner(1-\alpha/2)(B+1)\lrcorner)}$, and $\widehat{k}_{\alpha/2} := T^*_{n(B+1-\llcorner(1-\alpha/2)(B+1)\lrcorner)}$. Since $\Pr(\widehat{k}_{\alpha/2} \leq T_n \leq \widehat{k}_{1-\alpha/2}) \to 1-\alpha$,

$$\Pr\left(\overline{X} - \frac{\widehat{k}_{1-\alpha/2}\widehat{\sigma}}{\sqrt{n}} \leq \mu \leq \overline{X} - \frac{\widehat{k}_{\alpha/2}\widehat{\sigma}}{\sqrt{n}}\right) \to 1-\alpha$$

or the equal-tailed two-sided confidence interval for $\mu$ is

$$\left[\overline{X} - \frac{\widehat{k}_{1-\alpha/2}\widehat{\sigma}}{\sqrt{n}}, \leq \overline{X} - \frac{\widehat{k}_{\alpha/2}\widehat{\sigma}}{\sqrt{n}}\right]$$

**(ii)** Note that the null rejection property of the test has been discussed in PS7, Q2 (i). For the power properties, suppose that $H_0$ is false, or in other words, $\theta > 0$. Since $\widehat{\theta} \xrightarrow{p} \theta > 0$,

$$T^*_n = \sqrt{n}\frac{\widehat{\theta}^* - \widehat{\theta}}{s(\widehat{\theta}^*)} + \sqrt{n}\frac{\widehat{\theta} - \theta}{s(\widehat{\theta}^*)} + \sqrt{n}\frac{\theta}{s(\widehat{\theta}^*)}$$

would diverge to infinity, where the first summand converges in distribution to $N(0,1)$, the second summand is bounded in probability (converging to $N(0,1)$ under the assumption that $s(\widehat{\theta}^*)/s(\widehat{\theta}) \xrightarrow{p} 1$), and the last summand diverges to infinity (under the assumption that $s(\widehat{\theta}^*) \xrightarrow{p} s$ for some $s$). This implies that $q^*_n(1-\alpha)$ also diverges to infinity. At the same time,

$$T_n = \sqrt{n}\frac{\widehat{\theta} - \theta}{s(\widehat{\theta})} + \sqrt{n}\frac{\theta}{s(\widehat{\theta})}$$

would diverge to infinity at the same speed as $T^*_n$ does (under the same assumption). So the probability that $T_n > q^*_n(1-\alpha)$ may stay at a level between 0 and 1 even when $n \to \infty$. To see this,

$$\Pr(T_n > q^*_n(1-\alpha)) = \Pr\left(\frac{T_n}{\sqrt{n}} > \frac{q^*_n(1-\alpha)}{\sqrt{n}}\right) = \Pr\left(o_p(1) + \frac{T^*_n}{\sqrt{n}}\frac{s(\widehat{\theta}^*)}{s(\widehat{\theta})} > \frac{q^*_n(1-\alpha)}{\sqrt{n}}\right)$$

would converge to some number which is not far away from $\alpha$.