Econometrics Practice

Date: 07/02/2007

OUTLINE

1.Delta Method+CLT

2.Adding one observation to the information set

3. Consistency of s^2

4. Asymtotics for trend stationary linear model

1. **Review Question 5 of Section 2.1: Delta Method+CLT**

Let $\{z_i\}$ be a sequence of i.i.d random variables. with E(z_i)=$\mu \neq 0$ and var(z_i)=σ^2 and let also be $\bar{z}_n$ the sample mean, i.e. $\bar{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_i$. Show that

$$\sqrt{n}(\frac{1}{\bar{z}_n} - \frac{1}{\mu}) \rightarrow_d N \sim (0, \frac{\sigma^2}{\mu^4})$$

Proof. We know from *Lindeberg-Levy CLT* that given i.i.d $\{z_i\}$, E(z_i)=μ and var(z_i)=Σ,

$$CLT : \sqrt{n}(\bar{z}_n - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i - \mu) \rightarrow_d N \sim (0, \Sigma)$$

We also know from *Lemma 2.5 (Delta Method)* that given a continuous function $a(\cdot)$

$$\sqrt{n}(a(\bar{z}_n) - a(z)) \rightarrow_d A(z)z$$

where

$$\begin{aligned} \bar{z}_n &\rightarrow \ _p z \\ A(z) &= \frac{\partial a(z)}{\partial z^T} \end{aligned}$$

Exploting these two theorems we can see that, since

$$\begin{aligned} a(\bar{z}_n) &= \frac{1}{\bar{z}_n} \\ a(\mu) &= \frac{1}{\mu} \\ A(\mu) &= \frac{\partial a(\mu)}{\mu} = -\frac{1}{\mu^2} \end{aligned}$$

Hence

$$\sqrt{n}(a(\bar{z}_n) - a(z)) \rightarrow_d -\frac{1}{\mu^2} N \sim (0, \sigma^2)$$

bringing $A(\mu)$ into the variance term we complete the proof

$$\sqrt{n}(\frac{1}{\bar{z}_n} - \frac{1}{\mu}) \rightarrow_d N \sim (0, \frac{\sigma^2}{\mu^4})$$

2. **Review Question 8 of Section 2.2: Revision of expectations is m.d.s**

Let $\{y_i\}$ be a process such that $E(y_i|y_{i-1}, y_{i-2}.....y_1) < \infty$, i.e. exists and finite. Define $r_i^1 = E(y_i|y_{i-1}, y_{i-2}.....y_1) - E(y_i|y_{i-2}, y_{i-3}.....y_1)$,i.e. r_i^1 is the change in expectation as one more observation is added to the information set. Show that $\{r_i^1\}$ $(i \geq 2)$ is an martingale difference sequence w.r.t $\{y_i\}$.

Proof. So we have to show that $E(r_i^1|r_{i-1}^1, r_{i-2}^1.....r_2^1) = 0$, $i \geq 3$. First we define the information sets $(y_{i-1}, y_{i-2}.....y_1) := Y_{-1}$ and $(y_{i-2}, y_{i-3}.....y_1) := Y_{-2}$.Note that

$$(r_{i-1}^1, r_{i-2}^1.....r_2^1) \subset Y_{-2}$$

By Law of iterated expectations(L.i.E)

$$E(r_i^1|r_{i-1}^1, r_{i-2}^1.....r_2^1) = E(E(r_i^1|Y_{-2})|r_{i-1}^1, r_{i-2}^1.....r_2^1)$$

We need to show that $E(r_i^1|Y_{-2}) = 0$.
Since

$$E(r_i^1|Y_{-2}) = E(E(y_i|Y_{-1}) - E(y_i|Y_{-2})|Y_{-2})$$

Since

$$Y_{-2} \subset Y_{-1}$$

Again by L.i.E and linearity of conditional expectation

$$E(y_i|Y_{-2}) - E(y_i|Y_{-2}) = 0.$$

3. **Consistency of** s^2(**Variance of the residuals**)

Let $e_i = y_i - x_i^T\hat{\beta}$ be the OLS residuals for the ith observation. Under the assumptions 2.1-2.4 and provided that $E(\varepsilon_i^2) < \infty$,

$$\begin{aligned} s^2 &= \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 \rightarrow_p E(\varepsilon_i^2) \\ s^2 &= \frac{n}{n-K}\sum_{i=1}^{n}(\frac{1}{n}e_i^2) \end{aligned}$$

To show that s^2 is a consistent estimate of $E(\varepsilon_i^2)$ it suffices to show that the sample mean of e_i^2, ie. $\sum_{i=1}^n e_i^2$ converges in probability to $E(\varepsilon_i^2)$, since $\lim_{n\to\infty}\frac{n}{n-K} = 1$. To show consistency it is a good practice the express the residual i.t.o sampling error and ε_i, because then we can exploit the consistency of the coefficient estimator to show our claim. So, using add/substract trick

$$\begin{aligned} e_i &= y_i - x_i^T\hat{\beta} = y_i - x_i^T\hat{\beta} + x_i^T\beta - x_i^T\beta = \\ &= y_i - x_i^T\beta - x_i^T(\hat{\beta}-\beta) = \\ &= \varepsilon_i - x_i^T(\hat{\beta}-\beta) \end{aligned}$$

then we take the square of the residual

$$e_i^T e_i = \varepsilon_i^2 - 2(\hat{\beta}-\beta)^T x_i\varepsilon_i + (\hat{\beta}-\beta)^T x_i x_i^T(\hat{\beta}-\beta)$$

summing over i and multiplying with $\frac{1}{n}$ we obtain

$$\frac{1}{n}\sum_{i=1}^n e_i^2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - 2(\hat{\beta}-\beta)^T\frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i + \frac{1}{n}(\hat{\beta}-\beta)^T\sum_{i=1}^n x_i x_i^T(\hat{\beta}-\beta)$$

We want to show that the second and the third term in the above equation converge to 0, which will complete the proof. Recall the notation that

$$\begin{aligned} \bar{g} &= \frac{1}{n}\sum_{i=1}^n g_i = \frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i \\ S_{xx} &= \frac{1}{n}\sum_{i=1}^n x_i x_i^T \end{aligned}$$

then we can express it as

$$\frac{1}{n}\sum_{i=1}^n e_i^2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - 2(\hat{\beta}-\beta)^T\bar{g} + (\hat{\beta}-\beta)^T S_{xx}(\hat{\beta}-\beta)$$

Recall from the assumption 2.2, that $\{y_i, x_i\}$ is jointly stationary and ergodic. (Also any measurable function $f\ (y_i, x_i)$ is ergodic and stationary). So we will exloit the Ergodic Theorem and the assumption 2.3, predetermined regressors, that $E(g_i) = 0$, we have

$$\begin{aligned} \text{plim } \bar{g} &= E(g_i) = 0 \\ \frac{1}{n}\sum_{i=1}^n x_i x_i^T &\to {}_p\Sigma_{xx} = E(x_i x_i^T) \Rightarrow \text{plim } S_{xx} = \Sigma_{xx} \end{aligned}$$

since

$$\hat{\beta} - \beta = S_{xx}^{-1}\bar{g}$$

by lemma 2.3(a)

$$\begin{aligned}
\text{plim}(\hat{\beta}-\beta) &= \Sigma_{xx}^{-1}0 = 0 \\
\text{plim }(\hat{\beta}-\beta)^T\bar{g} &= \text{plim }(\hat{\beta}-\beta)^T\text{plim }\bar{g} = 0 \\
\text{plim }(\hat{\beta}-\beta)^T S_{xx}(\hat{\beta}-\beta) &= \text{plim }(\hat{\beta}-\beta)^T\text{plim }S_{xx}\text{plim }(\hat{\beta}-\beta) = 0
\end{aligned}$$

4. **Asymtotics for trend stationary linear model**
Assume we have the following model

$$y_t = \alpha + \delta t + \varepsilon_t$$

where by assumption $\{\varepsilon_t\}$ is independent white noise. The model can also be written as

$$\begin{aligned}
y_t &= x_t^T\beta + \varepsilon_t \\
x_t^T &= \begin{bmatrix} 1 \\ t \end{bmatrix}, \beta = \begin{bmatrix} \alpha \\ \delta \end{bmatrix}
\end{aligned}$$

then the OLS estimate of β based on the sample size T will be

$$\hat{\beta}_T = (\sum_{t=1}^{T} x_t x_t^{'})^{-1}(\sum_{t=1}^{T} x_t y_t)$$

(in order to avoid confusion " ' " is used instead of superscript "T" to indicate transpose.) Then i.t.o sampling error

$$\hat{\beta}_T - \beta = (\sum_{t=1}^{T} x_t x_t^{'})^{-1}(\sum_{t=1}^{T} x_t \varepsilon_t)$$

explitictly written

$$\begin{bmatrix} \hat{\alpha}_T - \alpha \\ \hat{\delta}_T - \delta \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} 1 & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t.\varepsilon_t \end{bmatrix}$$

since

$$\begin{aligned}
\sum_{t=1}^{T} t &= \frac{T(T+1)}{2} \\
\sum_{t=1}^{T} t^2 &= \frac{T(T+1)(2T+1)}{6}
\end{aligned}$$

we can write S$_{xx}$ as follows

$$S_{xx} = \begin{bmatrix} T & \frac{T(T+1)}{2} \\ \frac{T(T+1)}{2} & \frac{T^3}{3} + \frac{T^2}{2} + \frac{T}{6} \end{bmatrix}$$

Since x_t is not a stationary process, $\frac{S_{xx}}{T}$ does not converge in probability to a nonsingular matrix, it diverges. Consider the general convergence formula

$$\frac{1}{T^{\mu+1}}\sum_{t=1}^{T} t^{\mu} \rightarrow_p \frac{1}{\mu+1}$$

then

$$\frac{1}{T^3}S_{xx} \rightarrow \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

which is a singular matrix.

It turns out that the OLS estimates are consistent but have different rates of convergence; the rate of convergence is $\sqrt{T}$ for $\hat{\alpha}_T$ and $\sqrt{T^3}$ for $\hat{\delta}_T$. To assign the different rates of convergence to the elements of $\hat{\beta}_T - \beta$ we define the following matrix

$$\Phi_T = \begin{bmatrix} \sqrt{T} & 0 \\ 0 & \sqrt{T^3} \end{bmatrix}$$

and postmultiply

$$\begin{aligned}
\Phi_T(\hat{\beta}_T - \beta) &= \begin{bmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ \sqrt{T^3}(\hat{\delta}_T - \delta) \end{bmatrix} = \Phi_T(S_{xx})^{-1}S_{x\varepsilon} \\
\text{since } \mathbf{I} &= \Phi_T\Phi_T^{-1} \\
&= \Phi_T(S_{xx})^{-1}\Phi_T\Phi_T^{-1}S_{x\varepsilon} \\
&= (\Phi_T^{-1}S_{xx}\Phi_T^{-1})^{-1}(\Phi_T^{-1}S_{x\varepsilon}) \\
&= Q_n^{-1}v_n \\
Q_n &= (\Phi_T^{-1}S_{xx}\Phi_T^{-1}) \text{ and } v_n = (\Phi_T^{-1}S_{x\varepsilon})
\end{aligned}$$

if we substitute the values above using the definitions

$$\begin{aligned}
Q_n &= \begin{bmatrix} \frac{1}{\sqrt{T}} & 0 \\ 0 & \frac{1}{\sqrt{T^3}} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^{T} 1 & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{T}} & 0 \\ 0 & \frac{1}{\sqrt{T^3}} \end{bmatrix} \\
&= \begin{bmatrix} T^{-1}\sum_{t=1}^{T} 1 & T^{-2}\sum_{t=1}^{T} t \\ T^{-2}\sum_{t=1}^{T} t & T^{-3}\sum_{t=1}^{T} t^2 \end{bmatrix} \rightarrow_{T\rightarrow\infty} \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} = Q
\end{aligned}$$

so Q_n converges to Q which not zero, not infinite and nonsingular.

On the other hand

$$v_n = (\Phi_T^{-1}S_{x\varepsilon}) = \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \\ \frac{1}{\sqrt{T^3}}\sum_{t=1}^{T} t.\varepsilon_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} t.\frac{\varepsilon_t}{T} \end{bmatrix}$$

We have by assumption that $E(\varepsilon_t) = 0$ (independent white noise), we will show that $t.\frac{\varepsilon_t}{T}$ is a m.d.s as follows

$$\begin{aligned}
E(t.\frac{\varepsilon_t}{T}|\frac{t-1}{T}\varepsilon_{t-1}, ..., \frac{\varepsilon_1}{T}) &= \frac{t}{T}E(\varepsilon_t|\varepsilon_{t-1}, ...\varepsilon_1) = 0 \\
var(t.\frac{\varepsilon_t}{T}) &= \frac{t^2}{T^2}\sigma^2 = E[(\frac{t}{T}\varepsilon_t)]^2 = \sigma_t^2
\end{aligned}$$

We would like to show that the LLN on σ_t^2 implies (b$\rightarrow_p c$ and a$\rightarrow_{m.s} b \Rightarrow a \rightarrow_p c$, msc:z$_n \rightarrow_{m.s} \alpha = \lim_{n\rightarrow\infty} E(z_n - \alpha)^2 = 0$)

$$\frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 = \sigma^2(\frac{1}{T^3})\sum_{t=1}^{T}t^2 \rightarrow \frac{\sigma^2}{3}$$

To prove this claim

$$\begin{aligned}
&= E(\frac{1}{T}\sum_{t=1}^{T}[(\frac{t}{T}\varepsilon_t)]^2 - \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2)^2 = \\
&= E(\frac{1}{T}\sum_{t=1}^{T}[(\frac{t}{T}\varepsilon_t)]^2 - \frac{1}{T}\sum_{t=1}^{T}\frac{t^2}{T^2}\sigma^2 = \\
&= E(\frac{1}{T}\sum_{t=1}^{T}\frac{t^2}{T^2}(\varepsilon_t^2 - \sigma^2)^2 = \\
&= (\frac{1}{T})^2\sum_{t=1}^{T}\frac{t^4}{T^4}E(\varepsilon_t^2 - \sigma^2)^2 =
\end{aligned}$$

T times the above formula converges to (by convergence rate formula)

$$(\frac{1}{T})\sum_{t=1}^{T}\frac{t^4}{T^4}E(\varepsilon_t^2 - \sigma^2)^2 \rightarrow \frac{1}{6}E(\varepsilon_t^2 - \sigma^2)^2$$

implying that the term itself converges to 0

$$\frac{1}{T^6}\sum_{t=1}^{T}t^4 E(\varepsilon_t^2 - \sigma^2)^2 \rightarrow 0$$

Then

$$\begin{aligned}
[(\frac{t}{T}\varepsilon_t)]^2 &\rightarrow {}_{m.s}(\frac{t^2}{T^2}\sigma^2) \\
\frac{1}{T}\sum_{t=1}^{T}[(\frac{t}{T}\varepsilon_t)]^2 &\rightarrow {}_{P}\frac{\sigma^2}{3} \\
&\Rightarrow \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{t}{T}\varepsilon_t \rightarrow_d N \sim (0, \frac{\sigma^2}{3})
\end{aligned}$$

Finally,

$$v_n = (\Phi_T^{-1}S_{x\varepsilon}) = \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t \\ \frac{1}{\sqrt{T^3}}\sum_{t=1}^{T}t.\varepsilon_t \end{bmatrix} \rightarrow_d N \sim (\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 Q)$$

Thus the asymtotic distribution of $\Phi_T(\hat{\beta}_T - \beta)$ is normal with mean 0 and variance $Q^{-1}\sigma^2 Q.Q^{-1} = \sigma^2 Q^{-1}$. So, we have proved the proposition 2.11 (OLS estimation of time regression).