Chapter 6 Serial Correlation

6.5 Asymptotics for Sample Means of Serially Correlated Processes

(1) LLN for Covariance-Stationary

Prop 6.8

Let $\{y_t\}$ be covariance-stationary with mean μ and $\{\gamma_j\}$ be the auto-covariances of $\{y_t\}$, then:

(a) $\bar{y} \xrightarrow{m.s.} \mu$ as $n \to \infty$ if $\lim_{j\to\infty} \gamma_j = 0$

Proof ☆ (b) $\lim_{n\to\infty} Var(\sqrt{n}\,\bar{y}) = \sum_{j=-\infty}^{\infty} \gamma_j < \infty$ if $\{\gamma_j\}$ is summable [long-run variance]

(2) CLT for covariance-stationary processes

Prop 6.10

Suppose Gordin's condition holds for vector ergodic stationary process $\{g_t\}$. Then $E(g_t)=0$, $\{\Gamma_j\}$ is absolutely summable and $\sqrt{n}\,\bar{g} \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty}\Gamma_j)$

Tips. CLT In Chapter 2 requires ① Ergodic ② Stationary ③ m.d.s. We relax m.d.s. condition to permit serial correlation.

supplement

补充: Gordin's condition on ergodic stationary processes

(a) $E(g_t g_t')$ exists and is finite (b) $E(g_t \mid g_{t-j}, g_{t-j-1}, \cdots) \xrightarrow{m.s.} 0$ as $j\to\infty$

(c) $\sum_{j=0}^{\infty} [E(r_{tj}' r_{tj})]^{\frac{1}{2}}$ is finite where $r_{tj} \equiv E(g_t \mid g_{t-j}, g_{t-j-1}, \cdots) - E(g_t \mid g_{t-j-1}, g_{t-j-2}, \cdots)$

6.6 Serial Correlation in GMM

(1) Relaxation of Assumption 3.5 $[g_t = x_t \varepsilon_t]$

Assumption 3.5': $\{g_t\}$ satisfies Gordin's condition. Its long-run covariance matrix is non-singular.

(2) Asymptotic Distribution

Under A3.5', we have: $\sqrt{n}\,\bar{g} \xrightarrow{d} N(0, S)$

$$S = \sum_{j=-\infty}^{\infty} \Gamma_j = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j') \quad \text{where } \Gamma_j = E(g_t g_{t-j}')$$

Tips. Under A3.5, $S = \Gamma_0 = E(g_t g_t')$

With $\hat{S}$ be a proper consistent estimator of S, all results carry over.

(3) Estimation of Long-run Variance S

Natural Estimator for Γ_j: $\hat{\Gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} \hat{g}_t \hat{g}_{t-j}'$ Under some forth-moment condition, we have consistency of $\hat{\Gamma}_j$.

① Case 1: We know a priori $\Gamma_j = 0$ for $j > q$

$$\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{q} (\hat{\Gamma}_j + \hat{\Gamma}_j') = \sum_{j=-q}^{q} \hat{\Gamma}_j \qquad \hat{S} \text{ is consistent.}$$

② Case 2. No information of q

Naive Estimator: $\hat{S}_N = \sum_{j=1-n}^{n-1} \hat{\Gamma}_j \quad \hat{S} \not\to S$

(4) Kernel Estimator for Case 2

$$\hat{S} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{q(n)}\right) \cdot \hat{\Gamma}_j$$ ☆

k: kernel function ~~with $k(x) = \begin{cases} 1 & \text{if } |x| \le 1 \\ 0 & \text{if } |x| > 1 \end{cases}$~~

$q(n)$: bandwidth

① Truncated Kernel: $k(x) = \begin{cases} 1 & \text{for } |x| \le 1 \\ 0 & \text{for } |x| > 1 \end{cases}$ (Not guaranteed to be p.s.d. for $\hat{S}$)

② Bartlett Kernel: $k(x) = \begin{cases} 1-|x| & \text{for } |x| \le 1 \\ 0 & \text{for } |x| > 1 \end{cases}$ Newey-West Estimator

(Can get nonnegative definite $\hat{S}$)

③ Quadratic Spectral Kernel: $k(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$

**(5) VARHAC

Basic idea: Fit a finite-order VAR to the k-dimensional series $\{\hat{g}_t\}$ and then construct long-run covariance matrix implied by the estimated VAR.

Step 1: Lag length selection for VAR

$$\hat{g}_{kt} = \phi_1^{(k)} \hat{g}_{t-1} + \dots + \phi_p^{(k)} \hat{g}_{t-p} + e_{kt} \quad \text{where } \underset{(1\times k)}{\phi_j^{(k)}} = (\phi_{1j}^{(k)}, \dots, \phi_{kj}^{(k)})$$

Let $P_{max}(n)$ to be some integer known to be greater or equal to the true lag length. [Known ex ante] Let $SSR_p^{(k)}$ be the sum of squared ~~sum~~ residuals from OLS estimation above for $t = P_{max}(n)+1, P_{max}(n)+2, \dots, n$.

BIC criterion: $\hat{p} = \arg\min_p BIC = \ln(SSR_p^{(k)}/n) + p \cdot K \cdot \ln(n)/n$

$P\uparrow \Rightarrow \downarrow \qquad P\uparrow \Rightarrow \uparrow$

$\hat{p} \xrightarrow{P} p$ as $n \to \infty$.

Step 2. Calculating the implied long-run variance

Let $P^* = \max\{p(1), p(2), \dots, p(K)\}$

$$\hat{g}_t = \hat{\phi}_1 \hat{g}_{t-1} + \dots + \hat{\phi}_p \hat{g}_{t-p} + \hat{e}_t \quad \text{where k-th row of } \underset{(k\times k)}{\hat{\phi}_p} = \begin{cases} \hat{\phi}_p^{(k)} & \text{for } 1 \le p \le p(k) \\ 0 & \text{for } p(k) < p \le P^* \end{cases}$$

$$\underset{(k\times k)}{\hat{\Omega}} \equiv \frac{1}{n} \sum_{t=P_{max}(n)+1}^{n} \hat{e}_t \hat{e}_t' \qquad \hat{S} = [I_k - \sum_{p=1}^{P} \hat{\phi}_p]^{-1} \hat{\Omega} [I_k - \sum_{p=1}^{P} \hat{\phi}_p']^{-1}$$

Chapter 7 Extremum Estimator

7.1 Definition and Examples

(1) Def. An estimator $\hat{\theta}$ is called an extremum estimator if there is a scalar objective function $Q_n(\theta)$ such that $\hat{\theta}$ max $Q_n(\theta)$ subject to $\theta \in \Theta \subset R^p$, where Θ is a parameter space. (*)

(2) Existence

Lemma 7.1 Suppose (i) Θ is compact subset of R^p (ii) $Q_n(\theta)$ is cont. (iii) $Q_n(\theta)$ is a measurable function of the data $\forall \theta \in \Theta$ then there exists a measurable function $\hat{\theta}$ of the data that solves (*).

(3) Two important classes

① M-Estimators

Def. An extremum estimator is M-estimator if the objective function is a sample average: $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} m(w_t;\theta)$ where m is a real-valued function of (w_t, θ).

Ex. ML, NLS

② GMM

Def. An extremum estimator is a GMM estimator if the objective function can be written as $Q_n(\theta) = -\frac{1}{2} g_n(\theta)' \hat{W} g_n(\theta)$ with $\underset{(K\times 1)}{g_n(\theta)} \equiv \frac{1}{n}\sum g(w_t;\theta)$, where $\hat{W}$ is a $K\times K$ symmetric and P.D. matrix.

③ Classical Minimum Distance Estimators

$Q_n(\theta) = -\frac{1}{2} g_n(\theta)' \hat{W} g_n(\theta)$ but $g_n(\theta)$ is not necessary a sample mean.

(4) Maximum Likelihood (ML)

① Settings: $\{w_t\}$ is i.i.d., density indexed by finite-dimentional vector θ: $f(w_t;\theta)$, $\theta \in \Theta$. Assume $f(w_t;\theta_0)$ is true density.

② ML estimator maximizes log likelihood function.

$$\ln\left[\prod_{t=1}^{n} f(w_t;\theta)\right] = \sum_{t=1}^{n} \ln f(w_t;\theta)$$

Based on Def of M-Estimator: $m(w_t;\theta) = \ln f(w_t;\theta)$, $Q_n(\theta) = \frac{1}{n}\sum \ln f(w_t;\theta)$.

③ Efficiency: Efficient in quite general classes of asymptotically normal estimators.

(5) Conditional ML

Let $f(y_t|x_t;\theta_0)$ be the conditional density of y_t given x_t, $f(x_t;\psi_0)$ be the marginal density of x_t. Then $f(y_t, x_t;\theta_0,\psi_0) = f(y_t|x_t;\theta_0) f(x_t;\psi_0)$. Suppose θ_0 and ψ_0 are not functionally related, $\therefore \frac{1}{n}\sum \ln f(w_t;\theta,\psi) = \underbrace{\frac{1}{n}\sum \ln f(y_t|x_t;\theta)}_{\text{avg log conditional likelihood}} + \frac{1}{n}\sum \ln f(x_t;\psi)$

We max first term and ignore second term.

(6) Invariance of ML

① Reparameterizing

Consider a mapping $\lambda = \tau(\theta)$ on Θ. Let $\Lambda \equiv \tau(\Theta) \equiv \{\lambda \mid \lambda = \tau(\theta) \text{ for some } \theta \in \Theta\}$. $\tau: \Theta \to \Lambda$ is a reparameterization if it is one-to-one. ∴ We have $\tau^{-1}: \Lambda \to \Theta$.

② Invariance

An extremum estimator $\hat{\theta}$ is invariant to reparameterization τ if the extremum estimator for the reparameterized model is $\tau(\hat{\theta})$.

Let $\tilde{Q}_n(\lambda)$ be the objective function associated with the reparameterized model. An extremum

* ☆ Proof

estimator is invariant iff $\tilde{Q}_n(\lambda) = Q_n(\tau^{-1}(\lambda))$ for all $\lambda \in \Lambda$.

ML is invariant since $\tilde{f}(\cdot;\lambda) = f(\cdot;\tau^{-1}(\lambda))$.

* (7) NLS

Let θ_0 is true. $E(y_t|x_t) = \varphi(x_t;\theta_0)$. $\varepsilon_t \equiv y_t - E(y_t|x_t)$

$\Rightarrow y_t = \varphi(x_t;\theta_0) + \varepsilon_t$, $E(\varepsilon_t|x_t) = 0$, $\theta_0 \in \Theta$. Using least square method.

∴ $m(w_t;\theta) = -[y_t - \varphi(x_t;\theta)]^2$, $Q_n(\theta) = -\frac{1}{n}\sum[y_t - \varphi(x_t;\theta)]^2$

(8) GMM

General case: $a(y_t, z_t;\theta_0) = \varepsilon_t$ $\quad g(w_t;\theta) = x_t \cdot a(y_t, z_t;\theta)$

7.2 Consistency

(1) General Consistency Theorems

① Basic idea: We want [if $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$, $\theta_0 = \arg\max Q_0(\theta)$, then $\hat{\theta} \xrightarrow{p} \theta_0$.]

Uniform Convergence in Probability, $\sup_{\theta \in \Theta} \| Q_n(\theta) - Q_0(\theta) \| \xrightarrow{p} 0$ as $n \to \infty$.

② Thm with compactness

* ☆ Proof

Prop 7.1 Suppose (i) Θ is compact in R^p (ii) $Q_n(\theta)$ is cont. (iii) $Q_n(\theta)$ is measurable

ompact Θ + Cont. $Q_n(\theta)$ $\Rightarrow \hat{\theta}$ is well-defined. If (a) $Q_0(\theta)$ is uniquely maximized on Θ at $\theta_0 \in \Theta$ [Identification];

·Uniform Cov.

(b) $Q_n(\cdot)$ converges uniformly to $Q_0(\cdot)$ [Uniform Convergence]

Then $\hat{\theta} \xrightarrow{p} \theta_0$.

Identification + Convergence

③ Thm without compactness

nterior θ_0 + Concave Qn(θ) + Pointwise Cov.

Prop 7.2 Suppose (i) True θ_0 is an element of the interior of a convex space Θ (ii) $Q_n(\theta)$ is concave (iii) $Q_n(\theta)$ is measurable $\Rightarrow \hat{\theta}$. If (a) $Q_0(\theta)$ is uniquely maximized on Θ at $\theta_0 \in \Theta$ (b) $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$ pointwisely [Pointwise Convergence]

Then $\hat{\theta} \xrightarrow{p} \theta_0$.

(2) Consistency of M-Estimators

① If $\{w_t\}$ is ergodic stationary, then based on ergodic theorem, $Q_n(\theta) \xrightarrow{P} Q_0(\theta) = E[m(w_t;$ (pointwise)

To convert it to uniform convergence, we have:

Lemma 7.2 [Uniform Law of Large Numbers]

Let $\{w_t\}$ be ergodic stationary. Suppose (i) Θ is compact; (ii) $m(w_t;\theta)$ is cont in θ for all w_t; (iii) $m(w_t;\theta)$ is measurable (iv) [Dominance Condition] $\exists$ function $d(w_t)$ such that $|m(w_t;\theta)| \le d(w_t)$ $\forall \theta \in \Theta$ and $E[d(w_t)] < \infty$.

Then $\frac{1}{n}\sum m(w_t;\cdot)$ converges uniformly to $E[m(w_t;\cdot)]$ over Θ and $E[m(w_t;\theta)]$ is a cont. function of θ.

In vector form we rewrite Dominance Condition: $E[\sup_{\theta\in\Theta} \|h(w_t;\theta)\|] < \infty$. **

∴ Combining, we have two thms for M-Estimators

Prop 7.3 Let $\{w_t\}$ be ergodic stationary. Suppose (i) Θ compact (ii) $m(w_t;\theta)$ cont. (iii) $m(w_t;\theta)$ measurable $\Rightarrow \hat\theta$ If (a) $E[m(w_t;\theta)]$ uniquely max on Θ at $\theta_0 \in \Theta$; (b) $E[\sup_{\theta\in\Theta} |m(w_t;\theta)|] < \infty$ 区别(7.2)

Then $\hat\theta \xrightarrow{P} \theta_0$.

Prop 7.4 Let $\{w_t\}$ be ergodic stationary. Suppose (i) θ_0 interior of convex Θ (ii) $m(w_t;\theta)$ concave over θ $\forall w_t$ (iii) $m(w_t;\theta)$ measurable $\Rightarrow \hat\theta$ If (a) $E[m(w_t;\theta)]$ is uniquely max on Θ at $\theta_0 \in \Theta$; (b) $E[|m(w_t;\theta)|] < \infty$ $\forall \theta\in\Theta$

different from (7.2)

区别(7.2) $\Rightarrow$ pointwise Cov by ergodic Th

Then $\hat\theta \xrightarrow{P} \theta_0$.

(3) Concavity after Reparameterization

Suppose $m(\cdot)$ is not concave in θ, $\exists$ one-to-one mapping $\tau(\theta): \Theta \to \Lambda \equiv \tau(\Theta)$ such that $\tilde m(w_t;\lambda) \equiv m(w_t;\tau^{-1}(\lambda))$ is concave in λ and $\Lambda = \tau(\Theta)$ is convex. Let $\tilde Q_n(\lambda) \equiv \frac{1}{n}\sum \tilde m(w_t;\lambda)$ be the objective function after this reparameterization.

We can achieve concavity using reparameterization.

(4) Identification in NLS and ML

① NLS

First, let $m(w_t;\theta) = -[y_t - \varphi(x_t;\theta)]^2$, $\underbrace{E[y_t|x_t] = \varphi(x_t;\theta_0)}_{CEF}$, $\varepsilon_t = y_t - E[y_t|x_t]$.

We know CEF minimizes MSE. (Proof) Let $X \neq Y$ mean $Prob(X \neq Y) > 0$.

∴ Identification Condition: $\varphi(x_t;\theta) \neq \varphi(x_t;\theta_0)$ $\forall \theta \neq \theta_0$. [Min of MSE is unique]

② ML

a) Kullback-Leibler information inequality

$E[\ln f(y_t|x_t;\theta)] = \int \ln f(y_t|x_t;\theta) f(y_t, x_t; \theta_0, \psi_0)\, dy_t\, dx_t$

Kullback-Leibler:

$$\begin{cases} E[\ln f(y_t|x_t;\theta)] < E[\ln f(y_t|x_t;\theta_0)] & \text{if } f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0) \\ E[\ln f(y_t|x_t;\theta)] = E[\ln f(y_t|x_t;\theta_0)] & \text{if } f(y_t|x_t;\theta) = f(y_t|x_t;\theta_0). \end{cases}$$

Proof

$\therefore$ Identification Condition: $f(y_t|x_t;\theta) \neq f(y_t|x_t;\theta_0)\ \forall \theta \neq \theta_0$.

Remark: a) K-L Inequality also holds for unconditional densities.

b) Consistency is assured in ergodic stationary but not only i.i.d.

(5) Consistency of GMM

$$Q_0(\theta) = -\frac{1}{2} E[g(w_t;\theta)]' W E[g(w_t;\theta)]$$

$\because E[g(w_t;\theta_0)] = 0$, $\therefore Q_0(\theta_0) = 0$, $\therefore$ Identification Condition: $E[g(w_t;\theta)] \neq 0\ \forall \theta \neq \theta_0$.

Prop 7.7 Let $\{w_t\}$ be ergodic stationary, $\hat\theta = \underset{\theta\in\Theta}{\arg\min}[\frac{1}{n}\sum g(w_t;\theta)]'\hat{W}[\frac{1}{n}\sum g(w_t;\theta)]$, $\hat{W} \xrightarrow{P} W$, $E[g(w_t;\theta_0)] = 0$. Suppose (i) Θ compact (ii) $g(w_t;\theta)$ cont. (iii) $g(w_t;\theta)$ measurable. If a) $E[g(w_t;\theta)] \neq 0\ \forall \theta \neq \theta_0$ in Θ b) $E[\sup_{\theta\in\Theta} \| g(w_t;\theta)\|] < \infty$

Then, $\hat\theta \xrightarrow{P} \theta_0$.

* Identification for GMM

7.3 Normality

(1) Asymptotic Normality of M-Estimators

expand the derivative of Qn

$$Q_n(\theta) = \frac{1}{n}\sum m(w_t;\theta)$$

Settings: $\underset{(P\times 1)}{S(w_t;\theta)} = \frac{\partial m(w_t;\theta)}{\partial\theta}$ Gradient, $S_n = \frac{\partial Q_n(\theta)}{\partial\theta}$

Hessian $\underset{(P\times P)}{H(w_t;\theta)} = \frac{\partial S(w_t;\theta)}{\partial\theta'} = \frac{\partial^2 m(w_t;\theta)}{\partial\theta\partial\theta'}$, $H_n = \frac{\partial^2 Q_n(\theta)}{\partial\theta^2}$

Proof ② Prop 7.8

Using Mean-Value Theorem

对Q_n导数展开

Suppose $\{w_t\}$ is ergodic stationary and M-Estimator $\hat\theta$ is consistent. If (1) θ_0 is interior of Θ; (2) $m(w_t;\theta)$ is twice differentiable in θ $\forall w_t$, (3) $\frac{1}{\sqrt{n}}\sum S(w_t;\theta_0) \xrightarrow{d} N(0,\Sigma)$, Σ is P.D, (4) $E[\sup_{\theta\in N}\| H(w_t;\theta)\|] < \infty$ for some neighborhood N of θ_0, so that $\frac{1}{n}\sum H(w_t;\tilde\theta) \xrightarrow{P} E[H(w_t;\theta_0)]$; (5) $E[H(w_t;\theta_0)]$ is nonsingular.

Then $\sqrt{n}(\hat\theta - \theta_0) \xrightarrow{d} N(0, \underbrace{(E[H(w_t;\theta_0)])^{-1}\Sigma(E[H(w_t;\theta_0)])^{-1}}_{Avar(\hat\theta)})$

$$☆\ \underbrace{\frac{\partial Q_n(\hat\theta)}{\partial\theta}}_{=0\ [F.O.C]} = \frac{\partial Q_n(\theta_0)}{\partial\theta} + \frac{\partial^2 Q_n(\bar\theta)}{\partial\theta\partial\theta'}(\hat\theta - \theta_0) = \frac{1}{n}\sum S(w_t;\theta_0) + [\frac{1}{n}\sum H(w_t;\bar\theta)](\hat\theta - \theta_0)$$

Remark: If w_t is ergodic stationary, then so is $S(w_t;\theta_0)$, $\therefore \Sigma$ is the long-run variance matrix of $\{S(w_t;\theta_0)\}$.

③ Consistent Estimation of $Avar(\hat\theta)$

$$Avar(\hat\theta) = (E[H(w_t;\theta_0)])^{-1}\Sigma(E[H(w_t;\theta_0)])^{-1}$$

$$\widehat{Avar}(\hat\theta) = \{\frac{1}{n}\sum H(w_t;\hat\theta)\}^{-1}\hat\Sigma\{\frac{1}{n}\sum H(w_t;\hat\theta)\}^{-1}$$

change the order of derivative and integration

(2) Asymptotic Normality of Conditional ML

Assume i.i.d.

① Information Matrix Equality

~~$E[S(w_t;\theta_0)] = 0$~~ and $E[S(w_t;\theta_0)S(w_t;\theta_0)'] = -E[H(w_t;\theta_0)]$ ☆ Proof 交换积分求导顺序

② Prop 7.9 $E[S(w_t;\theta_0)|x_t]=0$

Let w_t be i.i.d. $\hat{\theta} \xrightarrow{P} \theta_0$

(1) θ_0 interior of Θ ; (2) $f(y_t|x_t;\theta)$ twice differentiable in θ ; (3) Information Matrix Equality holds [Under some technical condition] ; (4) $E[\sup_{\theta\in N} \|H(w_t;\theta)\|] < \infty$ $\frac{1}{n}\sum H(w_t;\tilde{\theta}) \xrightarrow{P} E[H(w_t;\theta_0)]$; (5) $E[H(w_t;\theta_0)]$ nonsingular

Then $Avar(\hat{\theta}) = -\{E[H(w_t;\theta_0)]\}^{-1} = \{E[S(w_t;\theta_0)S(w_t;\theta_0)']\}^{-1}$.

Remark: (3) requires interchange of integration and differentiation.

This proposition can be adapted to unconditional ML, replacing $f(y_t|x_t;\theta)$ by $f(w_t;\theta)$

expand gn

(3) Asymptotic Normality of GMM

对g_n[redacted]展开

$Q_n(\theta) = -\frac{1}{2} g_n(\theta)'\hat{W} g_n(\theta)$, $g_n(\theta) \equiv \frac{1}{n}\sum g(w_t;\theta)$

记☆ $0 = \frac{\partial Q_n(\hat{\theta})}{\partial \theta} = -G_n(\hat{\theta})'\hat{W} g_n(\hat{\theta})$ $G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'}$

① Prop 7.10

Suppose $\{w_t\}$ is ergodic stationary, $\hat{W} \xrightarrow{P} W$, W ~~P.S.~~ P.D. $\hat{\theta}$ is consistent.

$G \equiv E\left[\frac{\partial g(w_t;\theta_0)}{\partial \theta'}\right]$

(1) θ_0 interior of Θ ; (2) $g(w_t;\theta)$ dif in θ ; (3) $\frac{1}{\sqrt{n}}\sum g(w_t;\theta_0) \xrightarrow{d} N(0,S)$, S P.D. ; (4) $E\left[\sup_{\theta\in N}\left\|\frac{\partial g(w_t;\theta)}{\partial \theta'}\right\|\right] < \infty$; (5) $E\left[\frac{\partial g(w_t;\theta_0)}{\partial \theta'}\right]$ is of full column rank.

Then $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'WSWG(G'WG)^{-1})$.

☆ $g_n(\hat{\theta}) = g_n(\theta_0) + G_n(\bar{\theta})(\hat{\theta}-\theta_0) \Rightarrow$ F.O.C.

Consistent Estimator: $\widehat{Avar}(\hat{\theta}) = (\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{S}\hat{W}\hat{G}(\hat{G}'\hat{W}\hat{G})^{-1}$ $\hat{G} \equiv G_n(\hat{\theta}) = \frac{1}{n}\sum\frac{\partial g(w_t;\hat{\theta})}{\partial \theta'}$

Remark: a) Efficient : $\hat{W} = \hat{S}^{-1}$

b) J-test : $J \equiv n g_n(\hat{\theta})'\hat{S}^{-1} g_n(\hat{\theta}) \xrightarrow{d} \chi^2_{(k-p)}$

c) S is the long-run variance of $\{g(w_t;\theta_0)\}$. In i.i.d. case, $S = E[g(w_t;\theta_0)g(w_t;\theta_0)']$

(4) GMM vs. ML

$Avar(\hat{\theta}_{GMM}) \geq E[S(w_t;\theta_0)S(w_t;\theta_0)']^{-1} = Avar(\hat{\theta}_{ML})$

When $g(w_t;\theta) = S(w_t;\theta) \equiv \frac{\partial \ln f(w_t;\theta)}{\partial \theta}$, they are equivalent (asymptotically).

(5) GMM and ML in a Common Format

① M-Estimator

By prop 7.8 . $\sqrt{n}(\hat{\theta}-\theta_0) = -\left[\frac{1}{n}\sum H(w_t;\bar{\theta})\right]^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta}$

Let $\Psi = E[H(w_t;\theta_0)]$, then :

$\xrightarrow{p} 0$

$$\sqrt{n}(\hat{\theta}-\theta_0) = -\psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} - \underbrace{\{[\frac{1}{n}\Sigma H(W_t;\bar{\theta})]^{-1}-\psi^{-1}\}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta}}$$

$\therefore$ We can express $\sqrt{n}(\hat{\theta}-\theta_0) = -\psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p \xrightarrow{d} N(0, \psi^{-1}\Sigma\psi^{-1})$

$\Sigma \equiv Avar(\frac{\partial Q_n(\theta_0)}{\partial\theta})$. Here: $\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{n}}\Sigma S(W_t;\theta_0)$

② GMM

$\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} = -[G_n(\theta_0)]'\hat{W}\frac{1}{\sqrt{n}}\Sigma g(W_t;\theta_0) \xrightarrow{d} N(0,\Sigma)$

$\Sigma \equiv Avar(\frac{\partial Q_n(\theta_0)}{\partial\theta}) = G'WSWG$

Based on Prop 7.10: $\sqrt{n}(\hat{\theta}-\theta_0) = -B^{-1}C$, $B \equiv -G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})$, $C \equiv -G_n(\hat{\theta})'\hat{W}\frac{1}{\sqrt{n}}\Sigma g(W_t;\theta_0)$

We can write as: $\sqrt{n}(\hat{\theta}-\theta_0) = -\psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p$

where $\psi = -G'WG$

Derive ☆☆ See Table 7.1

Tips: When i.i.d., $\Sigma = -\psi$ for ML; When efficient GMM, $\Sigma = G'S^{-1}G = -\psi$

7.4 Hypothesis Testing

(1) Settings

The null hypothesis: $(H_0: \underset{(r\times 1)}{a(\theta_0)} = 0)$ $\quad \underset{(r\times p)}{A(\theta)} \equiv \frac{\partial a(\theta)}{\partial\theta'}$

$\therefore$ Constrained Estimator: $\tilde{\theta}$ ~~$\in$~~ $\max_{\theta\in\Theta} Q_n(\theta)$ s.t. $a(\theta)=0$

So we have assumptions

(A) $\sqrt{n}(\hat{\theta}-\theta_0) = -\psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} + o_p$ $\quad$ (B) $\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta} \xrightarrow{d} N(0,\Sigma)$ $\quad \frac{\partial Q_n(\hat{\theta})}{\partial\theta} \to 0$

(C) $\sqrt{n}(\tilde{\theta}-\theta_0) \xrightarrow{d} N(0,V)$ $\quad$ (D) $\Sigma = -\psi$

Remark: For (D), Wald and LM can be simplified with it and still $\chi^2(r)$ even without it; But LR will not be $\chi^2(r)$ without it.

(2) Wald Statistic

Prof $\sqrt{n}a(\hat{\theta}) = A(\bar{\theta})\sqrt{n}(\hat{\theta}-\theta_0) \Rightarrow \sqrt{n}a(\hat{\theta}) = A_0\sqrt{n}(\hat{\theta}-\theta_0) + o_p$ $\quad A_0 \equiv A(\theta_0)$

$\Rightarrow \sqrt{n}a(\hat{\theta}) \xrightarrow{d} N(0, A_0\Sigma^{-1}A_0) \Rightarrow W = n\,a(\hat{\theta})'[A(\hat{\theta})\hat{\Sigma}^{-1}A(\hat{\theta})']^{-1}a(\hat{\theta}) \xrightarrow{d} \chi^2(r)$

(3) LM Statistic

expand

F.O.C. of constraint problem $\begin{cases} \sqrt{n}\frac{\partial Q_n(\tilde{\theta})}{\partial\theta} + A(\tilde{\theta})'\sqrt{n}\gamma_n = 0 & ① \\ \sqrt{n}\,a(\tilde{\theta}) = 0 & ② \end{cases}$

拆 $\frac{\partial Q_n(\tilde{\theta})}{\partial\theta}$

① $\Rightarrow \frac{\partial Q_n(\theta_0)}{\partial\theta} + \frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}(\tilde{\theta}-\theta_0) + A'(\tilde{\theta})\gamma_n = 0$

$[\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}A'(\tilde{\theta})\gamma_n = -[\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}\frac{\partial Q_n(\theta_0)}{\partial\theta} - (\tilde{\theta}-\theta_0)$ $\quad (*)$

拆 $a(\tilde{\theta})$

② $\Rightarrow \overset{=0}{a(\theta_0)} + A(\bar{\theta}_1)(\tilde{\theta}-\theta_0) = 0$

$\sqrt{n}A(\bar{\theta}_1)\{[\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}\frac{\partial Q_n(\theta_0)}{\partial\theta} - [\frac{\partial Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}A'(\tilde{\theta})\gamma_n\} = 0$

$\xrightarrow{P}\psi$ for GMM and ML [Review Question 2]

$$\underbrace{\sqrt{n}A(\bar{\theta}_1)[\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}}_{\xrightarrow{P}A_0}\underbrace{A'(\tilde{\theta})}_{\xrightarrow{P}A_0}\gamma_n=-\sqrt{n}\underbrace{A(\bar{\theta}_1)}_{\xrightarrow{P}A_0}\underbrace{[\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}}_{\xrightarrow{P}\psi}\underbrace{\frac{\partial Q_n(\theta_0)}{\partial\theta}}_{\xrightarrow{d}N(0,\Sigma)}$$

$$\therefore \sqrt{n}\gamma_n\xrightarrow{d}N(0,\underbrace{[A_0\psi^{-1}A_0']^{-1}A_0\psi^{-1}\Sigma\psi^{-1}A_0'[A_0\psi^{-1}A_0']^{-1}}_{\equiv Avar(\gamma_n)})$$

expand... different from the one for LM

$$\therefore LM\equiv n\gamma_n'[\widehat{Avar}\,\gamma_n]^{-1}\gamma_n\xrightarrow{d}\chi^2_r$$

$$\therefore LM=n\gamma_n'[A(\tilde{\theta})\tilde{\Sigma}^{-1}A(\tilde{\theta})']\gamma_n=n(\frac{\partial Q_n(\tilde{\theta})}{\partial\theta})'\tilde{\Sigma}^{-1}(\frac{\partial Q_n(\tilde{\theta})}{\partial\theta})$$ by ① and $\Sigma=-$

拆 $\frac{\partial Q_n(\hat{\theta})}{\partial\theta}$ (与LM略不同)

(4) LR Statistic

$$LR\equiv 2n[Q_n(\hat{\theta})-Q_n(\tilde{\theta})]$$

From ①: $\underbrace{\frac{\partial Q_n(\hat{\theta})}{\partial\theta}}_{0}+\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}(\tilde{\theta}-\hat{\theta})+A(\tilde{\theta})'\gamma_n=0$

$$\Downarrow$$

$$\sqrt{n}(\tilde{\theta}-\hat{\theta})=-[\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}A'(\tilde{\theta})\sqrt{n}\gamma_n\xrightarrow{d}N(0,Avar(\tilde{\theta}-\hat{\theta}))$$

When $\Sigma=-\psi$, $Avar(\tilde{\theta}-\hat{\theta})=\psi^{-1}A_0'(A_0\Sigma^{-1}A_0)^{-1}A_0\psi^{-1}=\Sigma^{-1}A_0'(A_0\Sigma^{-1}A_0')^{-1}A_0\Sigma^{-1}$

$$\therefore LR=2n[Q_n(\hat{\theta})-[Q_n(\hat{\theta})+\underbrace{\frac{\partial Q_n(\hat{\theta})}{\partial\theta}'}_{=0}(\tilde{\theta}-\hat{\theta})+\frac{1}{2}(\tilde{\theta}-\hat{\theta})'\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}(\tilde{\theta}-\hat{\theta})]]$$

$$=-n(\tilde{\theta}-\hat{\theta})'\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}(\tilde{\theta}-\hat{\theta})$$

Lemma: $Z\sim N(0,I)$, M is idempotent with rank r, then: $Z'MZ\sim\chi^2_r$

Define $\pi=\Sigma^{-\frac{1}{2}}A_0'(A_0\Sigma^{-1}A_0')^{-1}A_0\Sigma^{\frac{1}{2}}$

$\sqrt{n}\pi^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}(\tilde{\theta}-\hat{\theta})\xrightarrow{d}N(0,I)$

$$\Rightarrow n(\tilde{\theta}-\hat{\theta})'\underbrace{\Sigma^{\frac{1}{2}}\underbrace{\pi^{-\frac{1}{2}}\pi\pi^{-\frac{1}{2}}}_{=I}\Sigma^{\frac{1}{2}}}_{=\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta\partial\theta'}}(\tilde{\theta}-\hat{\theta})\xrightarrow{d}\chi^2_r.$$

7.5 Numerical Optimization

Chapter 8 Examples of Maximum Likelihood

8.1 Qualitative Response (QR) Models

(1) Definition

Def. Regression models in which the dependent variable takes on discrete values are called qualitative response models.

(2) Logit Model

$$\begin{cases} f(y_t=1|x_t;\theta_0)=\Lambda(x_t'\theta_0) \\ f(y_t=0|x_t;\theta_0)=1-\Lambda(x_t'\theta_0)\end{cases} \qquad \Lambda(v)\equiv\frac{\exp(v)}{1+\exp(v)} \qquad \Lambda'(v)=\Lambda(v)[1-\Lambda(v)]$$ 记

$$\Rightarrow f(y_t|x_t;\theta_0)=\Lambda(x_t'\theta_0)^{y_t}\times[1-\Lambda(x_t'\theta_0)]^{1-y_t}$$

$$Q_n(\theta)=\frac{1}{n}\sum\{y_t\ln\Lambda(x_t'\theta)+(1-y_t)\ln(1-\Lambda(x_t'\theta))\}$$

$E(y_i|x_i)=\Lambda(x_i'\theta_0)$ $\therefore$ Marginal Effect: $\frac{\partial E(y_i|x_i)}{\partial x_{ij}}=\Lambda(x_i'\theta)[1-\Lambda(x_i'\theta)]\theta_j$

Multinomial Logit: $P(y_i=j|x_i;\theta)=\frac{\exp(x_i'\theta_j)}{\sum_{j=0}^{J}\exp(x_i'\theta_j)}=\frac{\exp(x_i'\theta_j)}{1+\sum_{j=1}^{J}\exp(x_i'\theta_j)}$

$i=1,\cdots,n$

$j=0,\cdots,J$

8.2 Truncated Regression Models (截取)

(1) Definition

Def. Observations for which y_t does not meet some prespecified criterion are excluded from the sample.

(2) Model

$\{y_t, x_t\}$ i.i.d. $\quad y_t = x_t'\beta_0 + \varepsilon_t, \ \varepsilon_t | x_t \sim N(0, \sigma_0^2)$

Truncation from below: Only $y > c$ are observed.

$$f(y|y>c) = \frac{f(y)}{\text{Prob}(y>c)}$$

记 ☆Thm. If $y \sim N(\mu_0, \sigma_0^2)$, c is constant, then:

$$\begin{cases} E(y|y>c) = \mu_0 + \sigma_0 \lambda(v) \\ Var(y|y>c) = \sigma_0^2\{1 - \lambda(v)[\lambda(v) - v]\} \end{cases} \quad \text{where } v = \frac{c-\mu_0}{\sigma_0}, \ \lambda(v) \equiv \frac{\phi(v)}{1-\Phi(v)}$$

$\therefore E(y|y>c) \neq \mu_0$, we have sample selection bias. $\sigma_0\lambda(v)$: bias

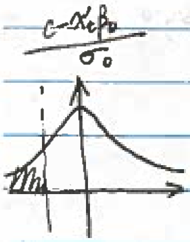
$\lambda(v)$: inverse Mill's ratio. $\lambda'(v) = \lambda(v)(\lambda(v) - v)$ $\begin{cases} v \to \infty \Rightarrow \lambda(v) \to v \\ v \to -\infty \Rightarrow \lambda(v) \to 0 \end{cases}$.

(3) ML Estimation

Before truncation: $y_t|x_t \sim N(x_t'\beta_0, \sigma_0^2)$ $\quad f(y_t|x_t) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{1}{2}\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right)^2\right] = \frac{1}{\sigma_0}\phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right)$

$\therefore E(y_t|x_t, t \text{ in sample}) = x_t'\beta_0 + \sigma_0\lambda\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)$ $\Rightarrow$ OLS is inconsistent.

$Var(y_t|x_t, t \text{ in sample}) = \sigma_0^2\left\{1 - \lambda\left(\frac{c-x_t'\beta_0}{\sigma_0}\right)\left[\lambda\left(\frac{c-x_t'\beta_0}{\sigma_0}\right) - \frac{c-x_t'\beta_0}{\sigma_0}\right]\right\}$



$\therefore \text{Prob}(y_t > c|x_t) = 1 - \text{Prob}(y_t \le c|x_t)$

$= 1 - \text{Prob}\left(\frac{y_t - x_t'\beta_0}{\sigma_0} \le \frac{c - x_t'\beta_0}{\sigma_0} \Big| x_t\right) = 1 - \Phi\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)$

$\therefore f(y_t|x_t, y_t>c) = \dfrac{\frac{1}{\sigma_0}\phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)}$ ☆推导 先算 $P(y_t>c|x_t)$，再算 $f(y_t|x_t, y_t>c)$

calculate...first, then calculate...

8.3 Censored Regression Models

(1) $y_t^* = x_t'\beta_0 + \varepsilon_t$ $\quad y_t = \begin{cases} y_t^* & 若 \ y_t^* > c \\ c & 若 \ y_t^* \le c \end{cases}$ $\quad$ y_t^*: 真值 $\quad y_t$: 观测值

(2) Tobit Likelihood Function

For those data unchanged: $f(y_t|x_t) = \frac{1}{\sigma_0}\phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right)$

For those data being censored:

$\text{Prob}(y_t^* \le c|x_t) = \text{Prob}\left(\frac{y_t^* - x_t'\beta_0}{\sigma_0} \le \frac{c - x_t'\beta_0}{\sigma_0} \Big| x_t\right)$

$= \Phi\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)$

Combine together:

$f(y_t|x_t; \beta, \sigma^2) = \left[\frac{1}{\sigma_0}\phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right)\right]^{1-D_t} \times \left[\Phi\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)\right]^{D_t}$ $\quad D_t = \begin{cases} 0 & 若 \ y_t > c \\ 1 & 若 \ y_t = c \end{cases}$

$\Rightarrow \ln f = (1 - D_t)\ln\left[\frac{1}{\sigma}\phi\left(\frac{y_t - x_t'\beta}{\sigma}\right)\right] + D_t \ln\Phi\left(\frac{c - x_t'\beta}{\sigma}\right)$

8.4 Sample Selection (Hansen 18.4)

(1) Model

$y_i = x_i'\beta + e_{1i}$ $\quad$ y_i is observed iff $T_i = 1$.

$T_i = \mathbb{1}(z_i'\gamma + e_{0i} > 0)$

Assume $\begin{pmatrix} e_{0i} \\ e_{1i} \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & \sigma^2 \end{pmatrix}\right)$. $\quad e_{1i} = \rho e_{0i} + v_i$, $v_i \sim N(0,1)$ $\quad v_i \perp\!\!\!\perp e_{0i}$

$E(e_{0i} \mid e_{0i} > -x) = \lambda(x) = \frac{\phi(x)}{\Phi(x)}$ [See Thm in Truncated Model]

(2) Error component

$$E(e_{1i} \mid T_i = 1, z_i) = E(e_{1i} \mid \{e_{0i} > -z_i'\gamma\}, z_i)$$
$$= \rho E(e_{0i} \mid \{e_{0i} > -z_i'\gamma\}, z_i)$$
$$= \rho\lambda(z_i'\gamma)$$

$\therefore e_{1i} = \rho\lambda(z_i'\gamma) + u_i$ where $E(u_i \mid T_i = 1, z_i) = 0$

$y_i = x_i'\beta + \underbrace{\rho\lambda(z_i'\gamma)}_{} + u_i$

$\hookrightarrow$ not included in Naive OLS.

(3) Method to have consistent estimation

① Heckit

Step 1: Estimate $\hat{\gamma}$ from a Probit model

It's valid since under assumptions above, we have:

$$\begin{cases} f(T_i = 1 \mid z_i, \gamma) = \phi(z_i'\gamma) \\ f(T_i = 0 \mid z_i, \gamma) = 1 - \phi(z_i'\gamma) \end{cases} \Rightarrow \text{Probit}$$

Step 2: Reg y_i on x_i, $\lambda(z_i'\hat{\gamma})$

Remarks on Heckit:

a) Conventional Variance will be incorrect;

b) Assumption of normality is too strong;

c) Work poorly if $\lambda(z_i'\hat{\gamma})$ does not have much in-sample variation.

② MLE Approach

For $T_i = 1$:

$$f(y_i, T_i = 1 \mid x_i, z_i) = f(y_i \mid x_i, z_i) \cdot f(T_i = 1 \mid y_i, x_i, z_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta_0}{\sigma}\right) \cdot \int_{-z_i'\gamma}^{\infty} f(e_{0i} \mid e_{1i}, x_i)\, de_{0i}$$

$f(e_{0i} \mid e_{1i}) = \phi\left(\frac{e_{0i} - \frac{\rho}{\sigma}e_{1i}}{\sqrt{1-\rho^2}}\right)$

$$= \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \cdot \int_{-z_i'\gamma}^{\infty} \phi\left(\frac{e_{0i} - \frac{\rho}{\sigma}(y_i - x_i'\beta)}{\sqrt{1-\rho^2}}\right) de_{0i}$$

$$= \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\left[1 - \Phi\left(\frac{z_i'\gamma - \frac{\rho}{\sigma}(y_i - x_i'\beta)}{\sqrt{1-\rho^2}}\right)\right] \quad (*)$$

For $T_i = 0$:

$\Pr(T_i = 0 \mid x_i, z_i) = \Pr(e_{0i} \le -z_i'\gamma) = \Phi(-z_i'\gamma)$