Hayashi Econometrics

Chapter 2 Large Sample Theory

2.1 Basic Limit Theorem

(1) Convergence in Probability

$\lim_{n\to\infty} \Pr(|z_n - \alpha| > \varepsilon) = 0 \Rightarrow$ "$z_n \xrightarrow{p} \alpha$".

Vector form: $\lim_{n\to\infty} \Pr(|z_{nk} - \alpha_k| > \varepsilon) = 0 \ \forall k$ [Every element converges in probability]

(2) Convergence Almost Surely

$\Pr(\lim_{n\to\infty} z_n = \alpha) = 1 \Rightarrow$ "$z_n \xrightarrow{a.s.} \alpha$".

*(3) Convergence in Mean Square

$\lim_{n\to\infty} E[(z_n - \alpha)^2] = 0$

(4) Convergence in Distribution

Def. $\{z_n\}$ converges in distribution to a random scalar z if the c.d.f. F_n of z_n converges to the c.d.f. ~~of~~ F of z at every continuity point of F. "$z_n \xrightarrow{d} z$"

Tips. For convergence in distribution, element-by-element convergence $\not\Rightarrow$ convergence for vector sequence.

* Multivariate Convergence in Distribution Theorem:

Let $\{z_n\}$ be a sequence of k-dimentional random vectors. Then:

"$z_n \xrightarrow{d} z$" $\Leftrightarrow$ "$\underset{1\times k}{\lambda'} \underset{k\times 1}{z_n} \xrightarrow{d} \lambda' z$" for any k-dimentional vector of real numbers."

(5) Convergence in distribution and in moments

Let α_{sn} be the s-th moment of z_n and $\lim_{n\to\infty} \alpha_{sn} = \alpha_s$ where α_s is finite. Then: "$z_n \xrightarrow{d} z$" $\Rightarrow$ "α_s is the s-th moment of z".

(6) Relationship between Convergences

See Amemiya Notes.

(7) CMT

$z_n \xrightarrow{p} \alpha \Rightarrow a(z_n) \xrightarrow{p} a(\alpha)$; $z_n \xrightarrow{p} z \Rightarrow a(z_n) \xrightarrow{p} a(z)$

$\therefore$ Particularly: $x_n \xrightarrow{p} \beta, y_n \xrightarrow{p} \gamma \Rightarrow x_n + y_n \xrightarrow{p} \beta + \gamma$; $x_n y_n \xrightarrow{p} \beta\gamma$; $\frac{x_n}{y_n} \to \frac{\beta}{\gamma}$;

$Y_n \xrightarrow{p} \Gamma \Rightarrow Y_n^{-1} \xrightarrow{p} \Gamma^{-1}$. [Counter-part of Slutzky's Theorem]

(8) Slutzky

(a) $x_n \xrightarrow{d} x$, $y_n \xrightarrow{p} \alpha \Rightarrow x_n + y_n \xrightarrow{d} x + \alpha$

(b) $x_n \xrightarrow{d} x$, $y_n \xrightarrow{p} 0 \Rightarrow y_n' x_n \xrightarrow{p} 0$

(c) $x_n \xrightarrow{d} x$, $A_n \xrightarrow{p} A \Rightarrow A_n x_n \xrightarrow{d} Ax$ In particular, if $x \sim N(0, \Sigma)$, then: $A_n x_n \xrightarrow{d} N(0, A\Sigma A')$

(d) $x_n \xrightarrow{d} x$, $A_n \xrightarrow{p} A \Rightarrow x_n' A_n^{-1} x_n \xrightarrow{d} x' A^{-1} x$

prerequisite for Delta Method is that xn is a consistent estimation of beta

(9) Asymptotically equivalent

When $z_n - x_n \xrightarrow{p} 0$, then we say z_n and x_n are asymptotically equivalent.

$$z_n \overset{a}{\sim} x_n, \quad \text{or} \quad z_n = x_n + o_p$$

Delta Method前提: x_n是β的一致估计.

(10) Delta Method (Proof)

Suppose $\{x_n\}$ is a sequence of k-dimensional random vectors such that $x_n \xrightarrow{p} \beta$ and $\sqrt{n}(x_n - \beta) \xrightarrow{d} z$. Suppose $a(\cdot): R^k \to R^r$ has continuous first derivative with $A(\beta)$ denoting the r×k matrix of first derivatives at β. $\underset{(r \times k)}{A(\beta)} = \frac{\partial a(\beta)}{\partial \beta'}$

prerequisite

Then, $\sqrt{n}[a(x_n) - a(\beta)] \xrightarrow{d} A(\beta) z$. [Another edition, see Hansen]

Asymptotic variance $\Rightarrow A(\beta) \cdot Avar(b) A(\beta)'$

Estimator $\Rightarrow A(b) Avar(b) A(b)'$

(11) CLT and LLN

WLLN: $\bar{z}_n = \frac{1}{n}\sum z_i \xrightarrow{p} \mu = E(z_i)$ [Kolmogorov] 前提: $\{z_i\}$ i.i.d, $E(z_i) = \mu$.

Another version: "$\lim_{n\to\infty} E(\bar{z}_n) = \mu$, $\lim_{n\to\infty} Var(\bar{z}_n) = 0$" ⇒ "$\bar{z}_n \xrightarrow{p} \mu$" [Chebychev] (Proof)

Lindeberg-Levy CLT: Let $\{z_i\}$ be i.i.d, $E(z_i) = \mu$, $Var(z_i) = \Sigma$. Then,

$$\sqrt{n}(\bar{z}_n - \mu) = \frac{1}{\sqrt{n}}\sum(z_i - \mu) \xrightarrow{d} N(0, \Sigma)$$

2.2 Basic Concepts in Time-Series

(1) Stationary Processes

① Strict Stationary

Def. A stochastic process $\{z_i\}$ is (strictly stationary) if:

$$f(z_{i_1}, z_{i_2}, \cdots z_{i_r}) = f(z_{i_1+k}, z_{i_2+k}, \cdots, z_{i_r+k}) \quad \forall\, i_j, k.$$

[What matters for distribution is the relative position in the sequence.]

S.S ⇒ W.S (When variance and covariances are finite.)

Particularly, all mean, variance and existing moments are the same across i.

$f(z_i) = f(z_j)$ ∀ i, j. Tips: Any transformation of stationary process is stationary.

e.g. i.i.d; constant series

Tips: Joint Stationarity $\overset{\Rightarrow}{\not\Leftarrow}$ element-wise Stationarity

② Covariance Stationarity

Def. A stochastic process $\{z_i\}$ is weakly stationary if:

(i) $E(z_i) = \mu < \infty$ (ii) $Cov(z_i, z_{i-j}) = r(j)$ depends only on j but not i.

Particularly: $Var(z_i) = \sigma^2 < \infty$.

j-th order autocovariance Γ_j: $\Gamma_j \equiv Cov(z_j, z_{i-j})$ $\Gamma_j = \Gamma'_{-j}$

j-th order autocorrelation coefficient ρ_j: $\rho_j = \frac{\sigma_j}{\sigma_0} = \frac{Cov(z_i, z_{i-j})}{Var(z_i)}$

Meaning: Asymptotically Independent.

* ③ Ergodicity

Def. A stationary process $\{z_i\}$ is ergodic if, for any two bounded functions $f: R^k \to R$ and $g: R^l \to R$,

$$\lim_{n\to\infty} |E[f(z_i, \cdots, z_{i+k}) g(z_{i+n}, \cdots, z_{i+n+l})]| = |E[f(z_i, \cdots, z_{i+k})]| \, |E[g(z_{i+n}, \cdots, z_{i+n}$$

Stationary + Ergodic ⇒ Ergodic Stationary Process

Ergodic Theorem.

Let $\{z_i\}$ be a stationary and ergodic process with $E(z_i)=\mu$. Then,

$$\bar{z}_n \equiv \frac{1}{n}\sum z_i \xrightarrow{a.s.} \mu$$

☆ Generalization of Kolmogorov's LLN [Serial dependence is allowed, but not asymptotically.]

e.g. AR(1) $z_i = c + \rho z_{i-1} + \varepsilon_i$, $|\rho|<1$

$\{\varepsilon_i\}$ is ind white noise

(2) Martingale

① Def. A scalar process $\{z_i\}$ is called a martingale w.r.t. $\{x_i\}$ if

$$E(z_i \mid \underbrace{x_{i-1}, x_{i-2}, \cdots}_{\text{information set}}) = z_{i-1} \quad \text{for } i \geq 2$$

Particularly, $\{z_i\}$ is a martingale if

$$E(z_i \mid z_{i-1}, z_{i-2}, \cdots) = z_{i-1} \quad \text{for } i \geq 2$$

② Def. $\{g_i\}$ with $E(g_i)=0$ is m.d.s if

$$E(g_i \mid g_{i-1}, g_{i-2}, \cdots, g_1) = 0 \quad \text{for } i \geq 2$$

Properties of m.ds.

a) Cumulative sum $\{z_i\}$ is a martingale.

* Proof

b) If $\{z_i\}$ is martingale, first difference of $\{z_i\}$ is m.d.s.

c) Cov$(g_i, g_{i-j})=0$ $\forall i$, $j\neq 0$.

W.N. is a covariance stationary process with some specific \mu and cov

(3) White Noise Processes

特定 μ, Cov下的 Covariance-Stationary Process.

Def. A covariance-stationary process $\{z_i\}$ is white noise if $E(z_i)=0$ and Cov$(z_i, z_{i-j})=0$ for $j\neq 0$.

White Noise + Independent ⇒ Independent White Noise

*(4) Random Walks

Def. A random walk, $\{z_i\}$, is a sequence of cumulative sums:

$$z_1 = g_1,\ z_2 = g_1 + g_2,\ \cdots,\ z_i = g_1 + g_2 + \cdots + g_i,\ \cdots$$

where $\{g_i\}$ is vector independent white noise process.

Properties: a) Sum of independent white noise (by def.)

b) martingale Proof

c) FD is white noise (independent)

(5) ARCH Processes

Autoregressive Conditional Heteroskedastic process

☆ Independent White Noise ⇒ stationary m.d.s. with finite variance
⇒ White noise

Def. A process $\{g_i\}$ is said to be an ARCH(1) if it can be written as
$g_i = \sqrt{\zeta + \alpha g_{i-1}^2} \cdot \varepsilon_i$ where $\{\varepsilon_i\}$ is i.i.d with mean zero and unit variance.

Properties:

Proof $\left\{ \begin{array}{l} \text{a) ARCH(1) is m.d.s.} \\ \text{b) Conditional second moment is function of its own history of process} \end{array} \right.$

$E(g_i^2 | g_{i-1}, g_{i-2}, \cdots, g_1) = \zeta + \alpha g_{i-1}^2$ [Own Conditional heteroskedasticity]

c) Strictly stationary and ergodic if $|\alpha| < 1$
and $E(g_i^2) = \frac{\zeta}{1-\alpha}$ (under Stationarity)

(b) CLT for Ergodic Stationary m.d.s.

Let $\{g_i\}$ be a vector m.d.s. that is stationary and ergodic with $E(g_i g_i') = \Sigma$. Let $\bar{g} \equiv \frac{1}{n} \Sigma g_i$. Then,

$$\sqrt{n}\,\bar{g} = \frac{1}{\sqrt{n}} \Sigma g_i \xrightarrow{d} N(0, \Sigma).$$

2.3 Large-Sample Distribution of OLS

(1) Model Setting

Tips:	
A 2.3 only restrict the contemporaneous relationship between the error term and the regressors.	A 2.1 (Linearity) $y_i = x_i'\beta + \varepsilon_i$
	A 2.2 $\{x_i, y_i\}$ is jointly ergodic and stationary
	A 2.3 All regressors are predetermined. $E(x_{ik}\varepsilon_i) = 0 \quad \forall i, k$ $\quad$ $E(g_i)=0 \Updownarrow E[x_i(y_i - x_i'\beta)] = 0$
	A 2.4 $E(x_i x_i')$ is nonsingular Σ_{xx}
A2.5 is stronger than A2.3.	A 2.5 $\{g_i\}$ is m.d.s., $E(g_i g_i')$ is nonsingular. $S \equiv Avar(\bar{g}) = E(g_i g_i')$

When x includes a constant:
$E(\varepsilon_i) = 0$, $Cov(x_{ik}, \varepsilon_i) = 0$, $E(\varepsilon_i | g_{i-1}, g_{i-2}, \cdots) = 0$, $E(\varepsilon_i | \varepsilon_{i-1}, \varepsilon_{i-2}, \cdots) =$

(2) OLS Estimator

① Proposition 2.1

Proof ☆ $\left\{ \begin{array}{l} \text{(a) Under A 2.1} \sim \text{2.4} \quad \operatorname{plim}_{n\to\infty} b = \beta. \\ \text{(b) Under A2.1} \sim \text{2.5} \\ \quad \sqrt{n}(b - \beta) \xrightarrow{d} N(0, Avar(b)) \quad Avar(b) = \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1} \\ \text{(c) } \widehat{Avar}(b) = S_{xx}^{-1} \hat{S} S_{xx}^{-1} \quad S_{xx}^{-1} = \frac{1}{n} X'X \quad \hat{S} = \frac{1}{n} \Sigma \hat{\varepsilon}_i^2 x_i x_i' \quad \text{consistent estimat} \end{array} \right.$

$\Sigma_{xx} = E(x_i x_i')$
$S = E(g_i g_i')$

(3) Consistent estimation of S

Proof ☆ (a) Estimation of error variance

Under A2.1~2.4, $s^2 \equiv \frac{1}{n-k} \Sigma \hat{e}_i^2 \xrightarrow{P} E(\varepsilon_i^2)$ where $\hat{e}_i \equiv y_i - x_i'b$

(b) Estimation of $S = E(\varepsilon_i^2 x_i x_i')$

Choice 1: $\hat{S} = \frac{1}{n} \Sigma \hat{\varepsilon}_i^2 x_i x_i'$ where $\hat{\varepsilon}_i \equiv y_i - x_i'\hat{\beta}$ Consistent Proof ☆ *

$$B = \begin{bmatrix} \hat{\varepsilon}_1^2 & & \\ & \ddots & \\ & & \hat{\varepsilon}_n^2 \end{bmatrix}$$

In matrix notation: $\hat{S} = \frac{\overset{k\times n\ n\times n\ n\times k}{X'BX}}{n}$ $\therefore \hat{Avar}(b) = n \cdot (X'X)^{-1}(X'BX)(X'X)^{-1}$

Choice 2: $\hat{S} = \frac{1}{n}\sum \frac{e_i^2}{(1-P_i)^d} x_i x_i'$ $d=1$ or 2 $P_i \equiv x_i'(X'X)^{-1}x_i = \frac{x_i' S_{xx}^{-1} x_i}{n}$.

2.4 Hypothesis Testing

(1) t-test

$H_0: \beta_k = \bar{\beta}_k$ $\sqrt{n}(b_k - \bar{\beta}_k) \xrightarrow{d} N(0, Avar(b_k))$

$t_k = \frac{\sqrt{n}(b_k - \bar{\beta}_k)}{\sqrt{\hat{Avar}(b_k)}} = \frac{b_k - \bar{\beta}_k}{\sqrt{\hat{Avar}(b_k)/n}} = \frac{b_k - \bar{\beta}_k}{SE(b_k)} \xrightarrow{d} N(0,1)$ $SE(b_k) \equiv \sqrt{\frac{1}{n}(S_{xx}^{-1}\hat{S}S_{xx}^{-1})_{kk}}$

(2) Wald-test

① For linear restriction

$H_0: R\beta = r$ $R: \#r \times k$

$W \equiv n \cdot (Rb - r)'[R[\hat{Avar}(b)]R']^{-1}(Rb - r) \xrightarrow{d} \chi^2_{\#r}$

② For nonlinear hypothesis

$H_0: a(\beta) = 0$ $_{=0 \text{ in } H_0}$ Let $A(\beta) = \frac{\partial a(\beta)}{\partial \beta'}$ $\#a$: restriction for H_0, dimension of $a(\beta)$

$\sqrt{n}[a(b) - \underset{\smile}{a(\beta)}] \xrightarrow{d} N(0, A(\beta)\, Avar(b)\, A(\beta)')$

$\sqrt{n}\, a(b)'[A(b)\hat{Avar}(b)\, A(b)']^{-1}\sqrt{n}\, a(b) \xrightarrow{d} \chi^2_{\#a}$

(3) Consistency of Test

Power: Prob of rejecting null when it is false.

Consistency of Test: A test is consistent against a set of [illegible] which satisfies the null, if the power against any particular memb[illegible] es unity as $n \to \infty$ for any assumed significance level.

(4) Asymptotic Local Power

Also see Hansen (8.20)

① Local Alternatives [Pitman drift]

Tips: 假定$\delta_l^{(n)}$为真值

Assume \delta_{l}^{n} is true

$\underset{\sim}{\delta_l^{(n)} = \bar{\delta}_l + \frac{\gamma}{\sqrt{n}}}$ *

$\therefore$ Rewrite t. $t_l = \frac{\sqrt{n}(\hat{\delta}_l(\hat{W}) - \delta_l^{(n)})}{\sqrt{\hat{Avar}(\hat{\delta}(\hat{W}))_{ll}}} + \underbrace{\frac{\gamma}{\sqrt{\hat{Avar}(\hat{\delta}(\hat{W}))_{ll}}}}_{}^{\mu} \xrightarrow{d} N(\mu, 1)$

$\therefore \mu$越大, Asymptotic $\underset{P(t_n > c | H_1)}{\text{power}}$越大 $\therefore$ Efficient GMM 的 power最大!

本质: Index the parameter by sample size so that the asymptotic distribution of the statistic is continuous in a localizing parameter.

So larger \mu, larger asymptotic power. So Efficient GMM has the largest power!

2.6 Implication of Homoskedasticity

A 2.7 (Conditional Homoskedasticity) $E(\varepsilon_i^2 | x_i) = \sigma^2 > 0$

(1) F-statistic

Def. $F = (Rb - r)'[R(\hat{Var}(b|X))R']^{-1}(Rb - r)/\#r = \frac{(SSR_R - SSR_U)/\#r}{SSR_U/(n-k)}$ *

(2) Implication to $\hat{S}$ estimation

$S = \cancel{\Sigma_{gx}}\ E(\varepsilon_i^2 x_i x_i') = E[E(\varepsilon_i^2 x_i x_i' | x_i)] = \sigma^2 E(x_i x_i') = \sigma^2 \Sigma_{xx}$

$\therefore \hat{S} = \hat{\sigma}^2 S_{xx} = s^2 S_{xx} \qquad \widehat{Avar}(b) = S^2 S_{xx}^{-1} = n \cdot s^2 \cdot (X'X)^{-1}$

(3) Implication to F and Wald

W is numerically identical to $\#r \cdot F$. *

2.7 Testing Conditional Homoskedasticity [White, 1980]

☆ Step 1: ~~Choose kxk symmetric matrix $x_i x_i'$~~

(Sketch Proof) Choose ψ_i be a vector collecting unique and nonconstant elements of $x_i x_i'$.

Step 2: Under some matrix $\hat{B}$ [Variance matrix estimator] *

$n \cdot C_n' \hat{B}^{-1} C_n \xrightarrow{d} \chi^2(m)$ where $C_n = \frac{1}{n}\sum(e_i^2 - s^2)\psi_i \xrightarrow{P} 0$.

m is dimension of C_n.

For certain choice $\hat{B}$, $n \cdot C_n' \hat{B}^{-1} C_n = nR^2$ where R^2 is from Reg e_i^2 on constant and ψ_i.

$\therefore nR^2 \xrightarrow{d} \chi^2(m)$. [Under homoskedasticity]

2.10 Testing for Serial Correlation

☆ Idea: When regressors include a constant, A 2.5 implies error term is a scalar m.d.s. $\therefore$ If error is serially correlated, A 2.5 should fail.

(1) Serial Correlation of univariate time-series

Setting: A sample of size n, $\{z_1, z_2, \cdots, z_n\}$ drawn from a scalar covariance-stationary process.

Def. Sample j-th order autocovariance

$\hat{\gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n}(z_t - \bar{z}_n)(z_{t-j} - \bar{z}_n) \qquad \bar{z}_n = \frac{1}{n}\sum_{t=1}^{n} z_t$

j-th order autocorrelation coefficient $\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$

$\hat{\gamma}_j \xrightarrow{P} \gamma_j = Cov(z_i, z_{i-j})$ ☆ Proof

Proposition 2.9 ☆ Proof (Under Ergodic) ☆

Suppose $\{z_t\}$ can be written as $\mu + \varepsilon_t$, ε_t stationary m.d.s. with $E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots) = \sigma^2$. Then $\sqrt{n}\,\hat{\gamma} \xrightarrow{d} N(0, \sigma^4 I_p)$, $\sqrt{n}\,\hat{\rho} \xrightarrow{d} N(0, I_p)$.

$n\hat{\rho}'\hat{\rho} = n\sum \hat{\rho}_j^2$

* Box-Pierce Q: $n\sum_{j=1}^{p}\hat{\rho}_j^2 = \sum(\sqrt{n}\hat{\rho}_j)^2 \xrightarrow{d} \chi^2(p) \qquad \sqrt{n}\hat{\rho}_j = \frac{\hat{\rho}_j}{1/\sqrt{n}} \xrightarrow{d} N(0,1)$

Ljung-Box Q: $n(n+2)\sum_{j=1}^{p}\frac{\hat{\rho}_j^2}{n-j} = \sum\frac{n+2}{n-j}(\sqrt{n}\hat{\rho}_j)^2 \xrightarrow{d} \chi^2(p)$

(2) Serial Autocorrelation Calculated from Residuals

m.d.s. $\varepsilon_t = y_t - x_t'\beta$ $\quad e_t = y_t - x_t' b$ $\quad \tilde{\rho}_j \equiv \frac{\tilde{\gamma}_j}{\tilde{\gamma}_0}$ $\quad \tilde{\gamma}_j \equiv \frac{1}{n}\sum_{t=j+1}^{n} \varepsilon_t \varepsilon_{t-j} \xrightarrow{p} E(\varepsilon_t \varepsilon_{t-j})$

Estimate $\tilde{\gamma}_j$: $\hat{\rho}_j \equiv \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$ $\quad \hat{\gamma}_j \equiv \frac{1}{n}\sum_{t=j+1}^{n} e_t e_{t-j}$

① When regressors are strictly exogenous

$$\sqrt{n}\,\hat{\gamma}_j = \sqrt{n}\,\tilde{\gamma}_j - \frac{1}{n}\sum_{t=j+1}^{n}(x_{t-j}\cdot\varepsilon_t + x_t\cdot\varepsilon_{t-j})'\sqrt{n}(b-\beta) + \sqrt{n}(b-\beta)'\Big(\frac{1}{n}\sum_{t=j+1}^{n} x_t x_{t-j}'\Big)(b-\beta)$$

If $E(x_{t-j}\cdot\varepsilon_t) + E(x_t\cdot\varepsilon_{t-j}) = 0$ $[E(x_t\cdot\varepsilon_s)=0\ \forall t,s]$, then second term $\to 0$,

$\sqrt{n}\,\hat{\gamma}_j \overset{A}{\sim} \sqrt{n}\,\tilde{\gamma}_j$, we can use $\hat{\gamma}_j$ to compute Q statistic.

② When regressors are predetermined, but not strictly exogenous

☆ We need to modify the Q statistic. $\quad$ ↗ stronger than A2.5

Additional assumptions: ① $E(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, x_t, x_{t-1}, \cdots) = 0$

② $E(\varepsilon_t^2 \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, x_t, x_{t-1}, \cdots) = \sigma^2 > 0$

Proposition 2.10 ☆ Proof

* $\sqrt{n}\,\hat{\gamma} \xrightarrow{d} N(0, \sigma^4\cdot(I_p - \phi))$ $\quad \sqrt{n}\,\hat{\rho} \xrightarrow{d} N(0, I_p - \phi)$

where $\phi_{jk} = E(x_t\varepsilon_{t-j})' E(x_t x_t')^{-1} E(x_t\cdot\varepsilon_{t-k})/\sigma^2$

modified Box-Pierce $Q \equiv n\cdot\hat{\rho}'(I_p - \hat{\phi})^{-1}\hat{\rho} \xrightarrow{d} \chi^2(p)$

where $\hat{\phi} \equiv \hat{\mu}_j' S_{xx}^{-1}\hat{\mu}_k / s^2$ $\quad s^2 \equiv \frac{1}{n-K}\sum e_t^2$ $\quad \hat{\mu}_j \equiv \frac{1}{n}\sum_{t=j+1}^{n} x_t\cdot e_{t-j}$

Modified B-P $Q \equiv n\cdot\hat{\rho}'(I_p - \hat{\phi})^{-1}\hat{\rho} \xrightarrow{d} \chi^2(p)$

** 2.12 Time Regressions

(1) Setting

Not stationary. $y_t = \alpha + \delta t + \varepsilon_t$ $\quad \varepsilon_t$. independent white noise

$y_t = x_t'\beta + \varepsilon_t$ $\quad x_t = (1, t)'$, $\beta = (\alpha, \delta)'$

(2) Asymptotics for OLS Estimator

$$b = \begin{bmatrix}\hat{\alpha}\\ \hat{\delta}\end{bmatrix} = (\sum x_t x_t')^{-1}(\sum x_t y_t) \qquad b-\beta = (\sum x_t x_t')^{-1}(\sum x_t\varepsilon_t) \xrightarrow{p} 0$$

$$\sum x_t x_t' = \begin{bmatrix} n & n\cdot(n+1)/2 \\ n\cdot(n+1)/2 & n\cdot(n+1)(2n+1)/6 \end{bmatrix}$$

Problem: $\hat{\alpha}, \hat{\delta}$ have different rates of convergence, respectively, $\sqrt{n}$ and $n^{\frac{3}{2}}$.

∴ Considering $\Upsilon_n = \begin{bmatrix}\sqrt{n} & 0 \\ 0 & n^{\frac{3}{2}}\end{bmatrix}$ $\quad \Upsilon_n(b-\beta) = \Upsilon_n(\sum x_t x_t')^{-1}(\sum x_t\varepsilon_t) = \underbrace{[\Upsilon_n^{-1}(\sum x_t x_t')\Upsilon_n^{-1}]}_{Q_n}{}^{-1}\underbrace{(\Upsilon_n^{-1}\sum x_t\varepsilon_t)}_{v_n}$

$\equiv Q_n^{-1} v_n$

Proposition 2.11 ☆ Proof

$\Upsilon_n(b-\beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$.

[Homoskedasticity]
The key assumption
for this to hold

* ☆ 补充: Auxiliary Regression-Based Test

关键假设

reg e_t on $x_t, e_{t-1}, e_{t-2}, \cdots, e_{t-p}$ $\quad$ F test for coefficients of e_{t-i} to be zero.

$Q \overset{A}{\sim} p\cdot F$ $\quad$ ∴ $p\cdot F \xrightarrow{d} \chi^2(p)$ [同差] $\quad p\cdot F = (n - \#x_t - p)\frac{R^2}{1-R^2}$

$\Rightarrow nR^2 = \frac{1}{1+\frac{p\cdot F}{n-\#x_t-p}}\cdot p\cdot F \xrightarrow{p} p\cdot F$ (Breusch-Godfrey Test)

Chapter 3 Single-Equation GMM

3.1 Endogeneity

Def. We say that a regressor is endogenous if it is not predetermined, i.e., if it does not satisfy the orthogonality condition.

① Simultaneity Bias

② Errors-in-Variables

3.3 The General Formulation

(1) A 3.1 (linearity) $\underset{1\times L}{y_i = z_i'}\underset{L\times 1}{\delta} + \varepsilon_i$

A 3.2 (ergodic stationary) x_i is vector of instruments. w_i is unique and nonconstant elements of (y_i, z_i, x_i). $\{w_i\}$ is jointly stationary and ergodic

A 3.3 (orthogonality conditions) All the K variables in x_i are predetermined in the sense that they are all orthogonal to current error term:

$E(x_{ik}\varepsilon_i)=0 \quad \forall i$ and k $\quad E[x_i(y_i - z_i'\delta)]=0 \quad E(g_i)=0$, $g_i \equiv x_i\varepsilon_i$

☆ Predetermined regressors should be included in instruments x_i.

A 3.4 (rank condition for identification) $E(x_i z_i')$ is of full column rank (L). Denote it by Σ_{xz}.

Thm.

Suppose there exists a solution to $Ax=b$. Then this linear system has only one solution iff A is of full column rank.

(2) Identification

$E(g_i)=0 \Rightarrow E[x_i(y_i - z_i'\tilde{\delta})]=0 \Rightarrow \underset{K\times L}{E(x_i z_i')}\underset{L\times 1}{\tilde{\delta}} = \underset{K\times 1}{E(x_i y_i)}$

$\tilde{\delta}=\delta$ is a solution [As assumed], $E(x_i z_i')$ is full column rank $\Rightarrow \tilde{\delta}=\delta$ is unique solution

(3) Order condition

① A necessary condition for identification:

$K \geq L$. [Order condition] # orthog... ≥ # parameters

number of instruments should be larger than number of endogenous variables

(工具变量数大于内生变量数)

② $K>L$: overidentified $\quad K=L$: just identified $\quad K<L$: unidentified

(4) Assumption for Asymptotic Normality

A 3.5 (g_i is m.d.s. with finite second moments) Let $g_i \equiv x_i\varepsilon_i$. $\{g_i\}$ is a m.d.s. $[E(g_i)=0]$. $E(g_i g_i')$ is nonsingular. Let $S = Avar(\bar{g}) = E(g_i g_i')$

Implications:

① $\{\varepsilon_i\}$ is m.d.s. ② Sufficient condition for A3.5: $E(\varepsilon_i | \varepsilon_{i-1}, \cdots \varepsilon_1, x_i, x_{i-1}, \cdots x_1)=0$ ③ $\{g_i\}$ is not serially correlated

$g_i(\delta) = g_i = x_i(y_i - z_i'\delta) = x_i\varepsilon_i$

$\hat{g}_i = x_i(y_i - z_i'\hat{\delta})$

$g_n = \frac{1}{n}\sum g_i$

$g_n(\delta) = \frac{1}{n}\sum g_i(\delta)$

$J(\delta, W) = n \cdot g_n(\delta)' W g_n(\delta)$

$\hat{\delta}(\hat{W}) = \underset{\tilde{\delta}}{\operatorname{argmin}}\, J(\tilde{\delta}, \hat{W}) = (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}S_{xy}$

$\bar{g} = \frac{1}{n}\sum x_i\varepsilon_i = g_n(\delta)$ [δ真值]

\delta true value

$S_{xz} = \frac{1}{n}\sum x_i z_i'$ $\quad S_{xy} = \frac{1}{n}\sum x_i y_i$

$\Sigma_{xz} = E(x_i z_i')$ $\quad \Sigma_{xy} = E(x_i y_i)$

3.4 GMM Defined

(1) Method of Moments

Basic Principle: Choose parameters so that the corresponding sample moments are also equal to zero.

① Population formula and Sample analogue

$$E(g(w_i;\delta)) \Rightarrow g_n(w_i;\tilde{\delta}) = g_n(\tilde{\delta}) \equiv \frac{1}{n}\sum g(w_i;\tilde{\delta})$$

$$g_n(\tilde{\delta}) = S_{xy} - S_{xz}\tilde{\delta} \qquad S_{xy} \equiv \frac{1}{n}\sum x_i y_i \qquad S_{xz} = \frac{1}{n}\sum x_i z_i'$$

② MM (When $K=L$) and IV Estimator

$$* \ \hat{\delta}_{IV} = S_{xz}^{-1}S_{xy} = \left(\frac{1}{n}\sum x_i z_i'\right)^{-1}\left(\frac{1}{n}\sum x_i y_i\right) \qquad [g_n(\hat{\delta}_{IV}) = 0]$$

When $z_i = x_i$, $\hat{\delta}_{IV} = \hat{\delta}_{OLS}$.

(2) Generalized Method of Moments (When $K > L$)

Problem: In overidentification case, system $g_n(\tilde{\delta}) = 0$ may have no solution.

Solution: Minimize distance in the sense of quadratic form.

① GMM Estimator ☆ Proof

Def. Let weighting matrix $\hat{W}$ be $K\times K$, symmetric, p.d. $\hat{W} \xrightarrow{P} W$, W also symmetric, p.d. The GMM Estimator of δ, $\hat{\delta}(\hat{W})$ is:

$$* \ \hat{\delta}(\hat{W}) \equiv \underset{\tilde{\delta}}{\operatorname{argmin}}\, J(\tilde{\delta}, \hat{W}) \qquad * \ J(\tilde{\delta}, \hat{W}) \equiv n \cdot g_n(\tilde{\delta})'\hat{W}g_n(\tilde{\delta})$$

$$\Rightarrow \hat{\delta}(\hat{W}) = (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}S_{xy}$$

② Sampling Error

$$\hat{\delta}(\hat{W}) - \delta = (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}\bar{g}. \qquad \bar{g} = \frac{1}{n}\sum x_i\varepsilon_i = g_n(\delta)$$

3.5 Large Sample Properties of GMM

(1) Asymptotic Distribution ☆ Proof

① Consistency $\hat{\delta}(\hat{W}) \xrightarrow{P} \delta$

② Asymptotic Normality $\sqrt{n}(\hat{\delta}(\hat{W}) - \delta) \xrightarrow{d} N(0, \text{Avar}(\hat{\delta}(\hat{W})))$

$$\text{Avar}(\hat{\delta}(\hat{W})) = (\Sigma_{xz}'W\Sigma_{xz})^{-1}\Sigma_{xz}'WSW\Sigma_{xz}(\Sigma_{xz}'W\Sigma_{xz})^{-1} \qquad S = E(g_i g_i') = E(\varepsilon_i^2 x_i x_i')$$

③ Consistent Estimate of $\text{Avar}(\hat{\delta}(\hat{W}))$

$$\widehat{\text{Avar}}(\hat{\delta}(\hat{W})) \equiv (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}\hat{S}\hat{W}S_{xz}(S_{xz}'\hat{W}S_{xz})^{-1}$$

(2) Error Variance

$$\hat{\varepsilon}_i = y_i - z_i'\hat{\delta} \qquad \frac{1}{n}\sum\hat{\varepsilon}_i^2 \xrightarrow{P} E(\varepsilon_i^2) \ \text{☆ Proof}$$

(3) Hypothesis Testing

① t-Statistics

$$H_0: \delta_\ell = \bar{\delta}_\ell \qquad t_\ell \equiv \frac{\sqrt{n}(\hat{\delta}_\ell(\hat{W}) - \bar{\delta}_\ell)}{\sqrt{\widehat{\text{Avar}}(\hat{\delta}(\hat{W}))_{\ell\ell}}} = \frac{\hat{\delta}_\ell(\hat{W}) - \bar{\delta}_\ell}{SE_\ell^*} \xrightarrow{d} N(0,1)$$

② Wald-Statistics

a) For linear transform

$H_0: R\delta = r$

$W \equiv n \cdot (R\hat{\delta}(\hat{W}) - r)' [R[\widehat{Avar}(\hat{\delta}(\hat{W}))]R']^{-1} (R\hat{\delta}(\hat{W}) - r) \xrightarrow{d} \chi^2(\#r)$

b) For general case

$H_0: a(\delta) = 0$ Let $A(\delta) = \frac{\partial a(\delta)}{\partial \delta'}$

$W \equiv n \cdot a(\hat{\delta}(\hat{W}))' \{A(\hat{\delta}(\hat{W}))[\widehat{Avar}(\hat{\delta}(\hat{W}))]A(\hat{\delta}(\hat{W}))'\}^{-1} a(\hat{\delta}(\hat{W})) \xrightarrow{d} \chi^2(\#a)$

(4) Estimation of S

$S = E(g_i g_i') = E(x_i x_i' \varepsilon_i^2)$ $\hat{S} = \frac{1}{n}\sum x_i x_i' \hat{\varepsilon}_i^2$ $\hat{S} \xrightarrow{P} S$

(5) Efficient GMM

① Proposition 3.5

A lower bound for asymptotic variance of GMM estimator is given by $(\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1}$. It's achieved if $\hat{W} = S^{-1}$. $\hat{\delta}(\hat{S}^{-1})$: Efficient GMM.

$\hat{\delta}(\hat{S}^{-1}) = (S_{xz}' \hat{S}^{-1} S_{xz})^{-1} (S_{xz}' \hat{S}^{-1} S_{xy})$

$Avar(\hat{\delta}(\hat{S}^{-1})) = (\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1} \Sigma_{xz}' S^{-1} S S^{-1} \Sigma_{xz} (\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1} = (\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1}$

② Two-step efficient GMM

Step 1: Choose $\hat{W} = S_{xx}^{-1}$ (or I), compute $\hat{\delta}(\hat{W})$, $\hat{\varepsilon}_i = y_i - z_i' \hat{\delta}(\hat{W})$

Step 2: Let $\hat{g}_i = x_i \hat{\varepsilon}_i$, $\bar{g} = \frac{1}{n}\sum x_i \hat{\varepsilon}_i$, $\hat{g}_i^* = \hat{g}_i - \bar{g}$. *

Then $\hat{W}^* = (\frac{1}{n}\sum \hat{g}_i^* \hat{g}_i^{*\prime})^{-1}$ $\hat{W}_u = (\frac{1}{n}\sum \hat{g}_i \hat{g}_i')^{-1}$ Two options, centered or uncentered

Choose $\hat{W} = \hat{W}^*$ or $\hat{W}_u$, compute $\tilde{\delta}(\hat{W})$.

③ Asymptotic Power

$t_l = \frac{\sqrt{n}(\hat{\delta}_l(\hat{W}) - \delta_l^{(n)})}{\sqrt{\widehat{Avar}(\hat{\delta}(\hat{W}))_{ll}}} + \underbrace{\frac{r}{\sqrt{\widehat{Avar}(\hat{\delta}(\hat{W}))_{ll}}}}_{\mu}$ $\delta_l^{(n)} = \bar{\delta}_l + \frac{r}{\sqrt{n}}$

$\xrightarrow{d} N(\mu, 1)$ ∴ Asymptotic power: Prob $(|x| > t_{\alpha/2})$ where $x \sim N(\mu, 1)$.

∴ $Avar(\hat{\delta}(\hat{W})) \downarrow \Rightarrow \mu \uparrow \Rightarrow$ power $\uparrow$. ∴ Power is maximized when we use efficient GMM estimator.

3.6 Testing Overidentifying Restrictions

$\bar{g} = \frac{1}{n}\sum g_i$ when $\tilde{\delta} = \delta$.

$J(\delta, \hat{S}^{-1}) = n \cdot \bar{g}' \hat{S}^{-1} \bar{g} = (\sqrt{n}\bar{g})' \hat{S}^{-1} (\sqrt{n}\bar{g}) \xrightarrow{d} \chi^2(K)$ ~~under H_0, $\tilde{\delta} = \delta$~~

$\sqrt{n}\bar{g} \xrightarrow{d} N(0, S)$

(1) Proposition 3.6

** $J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) = n \cdot g_n(\hat{\delta}(\hat{S}^{-1}))' \hat{S}^{-1} g_n(\hat{\delta}(\hat{S}^{-1})) \xrightarrow{d} \chi^2(K-L)$

Lemma: $v \xrightarrow{d} N(0, I)$ π is $L \times L$ idempotent matrix with rank q.

Then $v'\pi v \xrightarrow{d} \chi^2_q$

Tips: Specification test for all restrictions.

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \begin{array}{l} K_1 \text{ rows} \\ K-K_1 \text{ rows} \end{array}$$

(2) Testing Subsets of Orthogonality Conditions

Divide K instruments into two groups $x_i = [x_{i1}' \ x_{i2}']'$ x_{i2} are suspect.

We wish to test $E(x_{i2} \cdot \varepsilon_i) = 0$. Testable if $K_1 \geq L$.

Proposition 3.7

Assume $E(x_{i1} z_i')$ is of full column rank. Then for any consistent estimator $\hat{S}$ of S and $\hat{S}_{11}$ of S_{11}: $C \equiv J - J_1 \xrightarrow{d} \chi^2(K-K_1)$

where $J_1 = n \cdot g_{1n}(\bar{\delta})' (\hat{S}_{11})^{-1} g_{1n}(\bar{\delta})$

☆ Two-step procedure to get C-Statistics

* (3) Hypothesis Testing by Likelihood-Ratio Principle

Restricted Efficient GMM: $\bar{\delta}(\hat{S}^{-1}) \equiv \underset{\tilde{\delta}}{\text{argmin}}\, J(\tilde{\delta}, \hat{S}^{-1})$ subject to H_0

$LR \equiv J(\bar{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) - J(\hat{\delta}(\hat{S}^{-1}), \hat{S}^{-1}) \xrightarrow{d} \chi^2(a)$

Proposition 3.8

(a) W and LR ~~too~~ have same asymptotic distribution.

(b) $LR - W \xrightarrow{p} 0$

(c) If restrictions are linear, $R\delta = r$, $LR = W$ numerically.

- LR is invariance but W is not.
- If ~~$\hat{W} \neq \hat{S}^{-1}$~~ $\text{plim}\, \hat{W} \neq S^{-1}$, LR is not asymptotically Chi-Squared. But W will still be asymptotically Chi-Squared.
- Need to use same $\hat{S}^{-1}$ throughly.

3.8 Implication of Conditional Homoskedasticity

A 3.7 (Conditional homoskedasticity) $E(\varepsilon_i^2 | x_i) = \sigma^2$

$E(g_i g_i') = E(x_i x_i' \varepsilon_i^2) = \sigma^2 E(x_i x_i') = S$ $\quad \hat{S} = \hat{\sigma}^2 \frac{1}{n} \sum x_i x_i' = \hat{\sigma}^2 S_{xx}$

(1) Efficient GMM $\rightarrow$ 2SLS

$$\hat{\delta}(\hat{S}^{-1}) = [S_{xz}' \hat{S}^{-1} S_{xz}]^{-1} S_{xz}' \hat{S}^{-1} S_{xy} = [S_{xz}' (\hat{\sigma}^2 S_{xx})^{-1} S_{xz}]^{-1} S_{xz}' (\hat{\sigma}^2 S_{xx})^{-1} S_{xy} = [S_{xz}' S_{xx}^{-1} S_{xz}]^{-1} S_{xz}' S_{xx}^{-1} S_{xy}$$
$$= \hat{\delta}(S_{xx}^{-1}) = \hat{\delta}_{2SLS}$$

$\therefore$ No need to do first step.

$$\widehat{Avar}(\hat{\delta}_{2SLS}) = \hat{\sigma}^2 [S_{xz}' S_{xx}^{-1} S_{xz}]^{-1} S_{xz}' S_{xx}^{-1} S_{xx} S_{xx}^{-1} S_{xz} [S_{xz}' S_{xx}^{-1} S_{xz}]^{-1} = \hat{\sigma}^2 (S_{xz}' S_{xx}^{-1} S_{xz})^{-1}$$

$$Avar(\hat{\delta}_{2SLS}) = \sigma^2 (\Sigma_{xz}' \Sigma_{xx}^{-1} \Sigma_{xz})^{-1}$$

(2) Sargan's Statistic

$$J(\tilde{\delta}, (\hat{\sigma}^2 \cdot S_{xx})^{-1}) = n \cdot \frac{(S_{xy} - S_{xz}\tilde{\delta})' S_{xx}^{-1} (S_{xy} - S_{xz}\tilde{\delta})}{\hat{\sigma}^2} \xrightarrow{d} \chi^2(K-L)$$

★ Conclusion: Different estimators under GMM Framework

$$\text{GMM}\begin{cases}\xrightarrow{K=L} \text{IV estimator } \hat{\delta}_{IV}=S_{xz}^{-1}S_{xy} \xrightarrow{X=Z} \text{OLS estimator } \hat{\delta}_{OLS}=S_{xx}^{-1}S_{xy}\\ \xrightarrow{W=S^{-1}} \text{Efficient GMM } \hat{\delta}=(S_{xz}'\hat{S}^{-1}S_{xz})^{-1}S_{xz}'\hat{S}^{-1}S_{xy}\\ \xrightarrow{W=S^{-1},\text{ homo}} \text{2SLS estimator } \hat{\delta}_{2SLS}=(S_{xz}'S_{xx}^{-1}S_{xz})^{-1}S_{xz}'S_{xx}^{-1}S_{xy}\end{cases}$$

3.9 General GMM [Without assumption of linearity] (See Hansen)

(1) Settings

$E(g_i(\beta))=0 \qquad J(\beta)=n g_n'\hat{W}_n g_n \qquad g_n=\frac{1}{n}\sum g_i(\beta)$

F.O.C.: $\frac{\partial g_n'(\hat{\beta})}{\partial\beta}\cdot\hat{W}_n\cdot g_n(\hat{\beta})=0$

★★ (2) Asymptotic Distribution

$\frac{1}{\sqrt{n}}\sum g_i(\beta)\xrightarrow{d}N(0,S)$, $S=E(g_i(\beta)g_i'(\beta))$

$g_n(\hat{\beta})=g_n(\beta)+\underbrace{\frac{\partial g_n(\bar{\beta})}{\partial\beta}}_{G_n(\bar{\beta})}(\hat{\beta}-\beta)$, where $\bar{\beta}$ is between $\hat{\beta}$ and β.

$\therefore G_n(\hat{\beta})'\hat{W}[g_n(\beta)+\frac{\partial g_n(\bar{\beta})}{\partial\beta}(\hat{\beta}-\beta)]=0$

$\sqrt{n}(\hat{\beta}-\beta)=-[G_n(\hat{\beta})'\hat{W}G_n(\bar{\beta})]^{-1}G_n(\hat{\beta})'\hat{W}g_n(\beta)\cdot\sqrt{n}\xrightarrow{d}N(0,V)$

$V=Avar(\hat{\beta}(\hat{W}))=[G_n'(\beta)WG_n(\beta)]^{-1}G_n'(\beta)WSWG_n(\beta)[G_n'(\beta)WG_n(\beta)]^{-1}$

★★★ 3.10 Conditional Moment Restrictions (See Hansen)

(1) Settings

$E(e_i(\beta)|x_i)=0 \qquad$ stronger than $E(g_i(\beta))=0$

For any function $\phi(x_i)$, $g_i=\phi(x_i)e_i(\beta)$, $E(g_i(\beta))=0$.

$\therefore$ Any function of x_i, will define a moment condition, and consequently, a GMM estimator.

Question: What moment condition should we chose?

(2) Selection of best instrument

Optimal instrument approach (Chamberlain, 1987)

$R_i=E[\frac{\partial}{\partial\beta}e_i(\beta)|x_i] \qquad$ ~~$\sigma_i^2=E(e_i(\beta)^2|x_i)$~~ $\qquad \underset{r\times r}{\Sigma_i}=E[e_i(\beta)e_i'(\beta)|x_i]$

$F_i^{\circ}=-R_i^{\circ}\Sigma_i^{-1} \qquad g_n(\beta)=-R_i^{\circ}\Sigma_i^{-1}e_i(\beta)$

$\underset{\Sigma k_m \times 1}{g_i} = \begin{bmatrix} x_{i1}\cdot\varepsilon_{i1} \\ \vdots \\ x_{im}\cdot\varepsilon_{im}\end{bmatrix}$ $\underset{\Sigma L_m \times 1}{\delta} = \begin{bmatrix} \vdots \\ \delta_m \end{bmatrix}$ $\underset{\Sigma k_m \times \Sigma L_m}{\Sigma_{xz}} = \begin{bmatrix} \ddots & \\ & E(x_{im} z_{im}') \end{bmatrix}$

$g_i(\delta) = \begin{bmatrix} x_{i1}\cdot(y_{i1} - z_{i1}'\delta_1) \\ \vdots \\ x_{im}\cdot(y_{im} - z_{im}'\delta_m)\end{bmatrix}$ $\underset{\Sigma k_m \times 1}{\sigma_{xy}} = \begin{bmatrix} E(x_{i1}\cdot y_{i1}) \\ \vdots \\ E(x_{im}\cdot y_{im})\end{bmatrix}$ $\underset{\Sigma k_m \times \Sigma k_m}{S} = E(g_i g_i') = \begin{bmatrix} E(\varepsilon_{i1}\varepsilon_{i1} x_{i1} x_{i1}') & \cdots & E(\varepsilon_{i1}\varepsilon_{im} x_{i1} x_{im}') \\ \vdots & & \\ E(\varepsilon_{im}\varepsilon_{i1} x_{im} x_{i1}') & \cdots & E(\varepsilon_{im}\varepsilon_{im} x_{im} x_{im}')\end{bmatrix}$

Chapter 4 Multiple-Equation GMM

4.1 The Multiple-Equation Model

(1) Settings:

① m-th equation for observation i, denoted y_{im}.

A4.1 There are M linear equations $y_{im} = \underset{1\times L_m}{z_{im}'}\underset{L_m\times 1}{\delta_m} + \varepsilon_{im}$
[Assume no cross-equation restrictions]

② Let x_{im} be $k_m \times 1$ instruments vector for m-th equation. The set of instruments can differ across equations. [So k_m]

A4.2 Let w_i be unique and non-constant elements of $(y_{i1}, \cdots, y_{im}, z_{i1}, \cdots, z_{im}, x_{i1}, \cdots, x_{im})$. $\{w_i\}$ is jointly stationary and ergodic.
$\longrightarrow$ stronger than equation-by-equation

③ A 4.3 $E(x_{im}\cdot\varepsilon_{im}) = 0$ for each m.
$\therefore \underset{(\Sigma k_m \times 1)}{g_i} \equiv \begin{bmatrix} x_{i1}\cdot\varepsilon_{i1} \\ \vdots \\ x_{im}\cdot\varepsilon_{im}\end{bmatrix}$ Stack our moment conditions $E(g_i) = 0$

④ A 4.4 For each m, $\overset{k_m\times L_m}{E(x_{im} z_{im}')}$ is of full column rank.
Proof ✱ $\Sigma_{xz}\tilde{\delta} = \sigma_{xy}$ [From moment conditions] Σ_{xz} full column rank.

⑤ A4.5 $\{g_i\}$ is jointly m.d.s. $E(g_i g_i')$ is nonsingular.

4.2 Multiple-Equation GMM Defined

$\hat{W}$: weighting matrix $\Sigma_m k_m \times \Sigma_m k_m$

$\underset{\Sigma k_m \times 1}{g_n(\tilde{\delta})} = \begin{bmatrix} \frac{1}{n}\Sigma x_{i1}(y_{i1} - z_{i1}'\tilde{\delta}_1) \\ \vdots \\ \frac{1}{n}\Sigma x_{im}(y_{im} - z_{im}'\tilde{\delta}_m)\end{bmatrix} = S_{xy} - S_{xz}\tilde{\delta}$ $S_{xy} = \begin{bmatrix} \frac{1}{n}\Sigma x_{i1} y_{i1} \\ \vdots \\ \frac{1}{n}\Sigma x_{im} y_{im}\end{bmatrix}$ $S_{xz} = \begin{bmatrix} \frac{1}{n}\Sigma x_{i1} z_{i1}' & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n}\Sigma x_{im} z_{im}'\end{bmatrix}$

Denote:
$\hat{W} = \begin{bmatrix} \hat{W}_{11} & \cdots & \hat{W}_{1m} \\ \vdots & & \\ \hat{W}_{m1} & \cdots & \hat{W}_{mm}\end{bmatrix}$
$\hat{W}_{mh}$ is (m,h) block ($k_m \times k_h$) of $\hat{W}$.

$\tilde{\delta} = \begin{bmatrix} \tilde{\delta}_1 \\ \vdots \\ \tilde{\delta}_m \end{bmatrix}$ With same way: GMM estimator $\hat{\delta}(\hat{W}) = (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}S_{xy}$
Sampling error $\hat{\delta}(\hat{W}) - \delta = (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}\bar{g}$

$\therefore$ Stacked vector: g_i, $g_i(\delta)$, $\bar{g}$, δ, σ_{xy}, S_{xy}
Blocked ~~x~~ matrix: Σ_{xz}, S_{xz}, S, $\hat{W}$.

$$\hat{\delta}(\hat{W}) = \begin{bmatrix} \hat{\delta}_1(\hat{W}) \\ \vdots \\ \hat{\delta}_m(\hat{W})\end{bmatrix} = \begin{bmatrix} (\frac{1}{n}\Sigma z_{i1} x_{i1}')\hat{W}_{11}(\frac{1}{n}\Sigma x_{i1} z_{i1}') & \cdots & (\frac{1}{n}\Sigma z_{i1} x_{i1}')\hat{W}_{1m}(\frac{1}{n}\Sigma x_{im} z_{im}') \\ \vdots & & \vdots \\ (\frac{1}{n}\Sigma z_{im} x_{im}')\hat{W}_{m1}(\frac{1}{n}\Sigma x_{i1} z_{i1}') & \cdots & (\frac{1}{n}\Sigma z_{im} x_{im}')\hat{W}_{mm}(\frac{1}{n}\Sigma x_{im} z_{im}')\end{bmatrix}^{-1}$$

$$\begin{bmatrix} (\frac{1}{n}\Sigma z_{i1} x_{i1}')\hat{W}_{11}(\frac{1}{n}\Sigma x_{i1} y_{i1}) + \cdots + (\frac{1}{n}\Sigma z_{i1} x_{i1}')\hat{W}_{1m}(\frac{1}{n}\Sigma x_{im} y_{im}) \\ \vdots \\ (\frac{1}{n}\Sigma z_{im} x_{im}')\hat{W}_{m1}(\frac{1}{n}\Sigma x_{i1} y_{i1}) + \cdots + (\frac{1}{n}\Sigma z_{im} x_{im}')\hat{W}_{mm}(\frac{1}{n}\Sigma x_{im} y_{im})\end{bmatrix}$$

$$\hat{S} = \begin{bmatrix} \frac{1}{n}\sum \hat{\varepsilon}_{i1}\hat{\varepsilon}_{i1} x_{i1} x_{i1}' & \cdots & \frac{1}{n}\sum \hat{\varepsilon}_{i1}\hat{\varepsilon}_{im} x_{i1} x_{im}' \\ \vdots & & \vdots \\ \frac{1}{n}\sum \hat{\varepsilon}_{im}\hat{\varepsilon}_{i1} x_{im} x_{i1}' & \cdots & \frac{1}{n}\sum \hat{\varepsilon}_{im}\hat{\varepsilon}_{im} x_{im} x_{im}' \end{bmatrix}$$

$\hat{S} \xrightarrow{P} S$

4.3 Large-Sample Theory

(1) Test

$J \sim \chi^2(\sum_m k_m - \sum_m L_m)$ $\quad C \sim \chi^2(\sum k_m - \sum k_{1m})$

(2) Error moments

Let $\hat{\varepsilon}_{im} = y_{im} - z_{im}'\hat{\delta}_m$. $\hat{\delta}_m \xrightarrow{P} \delta_m$

Define $\hat{\sigma}_{mh} = \frac{1}{n}\sum \hat{\varepsilon}_{im}\hat{\varepsilon}_{ih}$, $\sigma_{mh} = E(\varepsilon_{im}\varepsilon_{ih})$ then $\hat{\sigma}_{mh} \xrightarrow{P} \sigma_{mh}$.

$\hat{S} \xrightarrow{P} S$ $\quad \therefore$ Using $\hat{S}^{-1}$, $\hat{\delta}(\hat{S}^{-1})$ is an efficient GMM Estimator.

(3) Asymptotic Variance

$$\text{Avar}(\hat{\delta}(\hat{S}^{-1})) = (\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1}$$

$$\widehat{\text{Avar}}(\hat{\delta}(\hat{S}^{-1})) = (S_{xz}' \hat{S}^{-1} S_{xz})^{-1}$$

4.4. Single Equation vs. Multiple-Equation

Estimate seperately $\overset{?}{\Leftrightarrow}$ Estimate jointly

Single GMM $\qquad$ Multiple GMM

(1) Equation-by-Equation Estimator

$$\hat{\delta}(\hat{W}) = (S_{xz}'\hat{W}S_{xz})^{-1}S_{xz}'\hat{W}S_{xy}$$

Only difference: $\hat{W}^* = \begin{bmatrix} \hat{W}_{11} & & 0 \\ & \ddots & \\ 0 & & \hat{W}_{mm} \end{bmatrix}$ choose a diagonal matrix extracted from $\hat{W}$ for joint estimation.

☆ Equation-by-Equation GMM is a particular multiple GMM.

(2) When are they equivalent?

Case 1: $L_m = k_m \ \forall m$. All the equations are just-identified. $\hat{\delta}(\hat{W}^*) = \hat{\delta}(\hat{S}$ numerically

Case 2: When equations are unrelated in the sense that

$$E(\varepsilon_{im}\varepsilon_{ih} x_{im} x_{ih}') = 0 \quad \forall m \neq h.$$

$$S^{-1} = \underset{n\to\infty}{\text{plim}}\, \hat{W} = \begin{bmatrix} E(\varepsilon_{i1}^2 x_{i1} x_{i1}')^{-1} & & \\ & \ddots & \\ & & E(\varepsilon_{i1}^2 x_{i1} x_{i1}')^{-1} \end{bmatrix} = \underset{n\to\infty}{\text{plim}}\, \hat{W}^*$$

$\therefore \hat{W} - \hat{S}^{-1} \xrightarrow{P} 0$, $\sqrt{n}\,\hat{\delta}(\hat{W}^*) - \sqrt{n}\,\hat{\delta}(\hat{S}^{-1}) \xrightarrow{P} 0$.

$\hat{\delta}(\hat{W}^*)$ and $\hat{\delta}(\hat{S}^{-1})$ are asymptotically equivalent.

(3) Weakness of joint estimation

① Seperate estimation has better finite-sample performance.

② The consistency of joint estimation presumes that the model is correctly specified. Biases due to a local misspecification contaminate the rest of the system.

4.5 Homoskedasticity and SUR

(1) Settings.

A4.7 $\quad E(\varepsilon_{im}\varepsilon_{ih} | x_{im}, x_{ih}) = \sigma_{mh}$

Kronecker product: $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1N}B \\ \vdots & & \vdots \\ a_{M1}B & \cdots & a_{MN}B \end{bmatrix}$

$$\therefore S = E(g_i g_i') = \begin{bmatrix} \sigma_{11} E(x_{i1} x_{i1}') & \cdots & \sigma_{1m} E(x_{i1} x_{im}') \\ \vdots & & \vdots \\ \sigma_{m1} E(x_{im} x_{i1}') & \cdots & \sigma_{mm} E(x_{im} x_{im}') \end{bmatrix}$$

$$\underset{m\times 1}{\varepsilon_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{im} \end{bmatrix} \quad \underset{m\times m}{\Sigma} \equiv \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{bmatrix} = E(\varepsilon_i \varepsilon_i') \quad \hat{\Sigma} \equiv \begin{bmatrix} \hat\sigma_{11} & \cdots & \hat\sigma_{1m} \\ \vdots & & \vdots \\ \hat\sigma_{m1} & \cdots & \hat\sigma_{mm} \end{bmatrix} = \frac{1}{n} \sum \hat\varepsilon_i \hat\varepsilon_i'$$

Let x_t be the common set of instruments. $x_i = x_{i1} = x_{i2} = \cdots = x_{im}$.

$$\underset{mk\times 1}{g_i} = \varepsilon_i \otimes \underset{k\times 1}{x_i} = \begin{bmatrix} \varepsilon_{i1} \cdot x_i \\ \vdots \\ \varepsilon_{im} \cdot x_i \end{bmatrix} \quad \underset{mk\times mk}{S} = \underset{m\times m}{\Sigma} \otimes \underset{k\times k}{E(x_i x_i')} \quad S^{-1} = \Sigma^{-1} \otimes [E(x_i x_i')]^{-1}$$

$$\hat S^{-1} = \hat\Sigma^{-1} \otimes \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1}$$

$\therefore$ For efficient GMM: $\hat W_{mh} = \hat\sigma^{mh} \cdot \left(\frac{1}{n} \sum x_i x_i'\right)^{-1}$

(2) SUR Regression

Assume: x_i = union of $(z_{i1}, \cdots, z_{im}) \Leftrightarrow E(z_{im} \cdot \varepsilon_{ih}) = 0$

We claim here that the predetermined regressors satisfy "cross" orthogonalities.

① 3SLS (More general than SUR)

When we only have A 4.7 but not $E(z_{im} \cdot \varepsilon_{ih}) = 0$, $\forall m, h$

$$\hat\delta_{3SLS} = \begin{bmatrix} \hat\sigma^{11} \hat A_{11} & \cdots & \hat\sigma^{1m} \hat A_{1m} \\ \vdots & & \vdots \\ \hat\sigma^{m1} \hat A_{m1} & \cdots & \hat\sigma^{mm} \hat A_{mm} \end{bmatrix}^{-1} \begin{bmatrix} \hat\sigma^{11} \hat c_{11} + \cdots + \hat\sigma^{1m} \hat c_{1m} \\ \vdots \\ \hat\sigma^{m1} \hat c_{m1} + \cdots + \hat\sigma^{mm} \hat c_{mm} \end{bmatrix}$$

where $\hat A_{mh} \equiv \left(\frac{1}{n} \sum_i z_{im} x_i'\right) \left(\frac{1}{n} \sum_i x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_i x_i z_{ih}'\right)$

$\hat c_{mh} \equiv \left(\frac{1}{n} \sum_i z_{im} x_i'\right) \left(\frac{1}{n} \sum_i x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_i x_i y_{ih}\right)$

$$\widehat{Avar}(\hat\delta_{3SLS}) = \begin{bmatrix} \hat\sigma^{11} \hat A_{11} & \cdots & \hat\sigma^{1m} \hat A_{1m} \\ \vdots & & \vdots \\ \hat\sigma^{m1} \hat A_{m1} & \cdots & \hat\sigma^{mm} \hat A_{mm} \end{bmatrix}^{-1}$$

$$Avar(\hat\delta_{3SLS}) = \begin{bmatrix} \sigma^{11} A_{11} & \cdots & \sigma^{1m} A_{1m} \\ \vdots & & \vdots \\ \sigma^{m1} A_{m1} & \cdots & \sigma^{mm} A_{mm} \end{bmatrix}^{-1} \quad A_{mh} = E(z_{im} x_i') [E(x_i x_i')]^{-1} E(x_i z_{ih}')$$

σ^{mh} is (m,h) element of Σ^{-1}.

② SUR Regression

In SUR setting: $x_{i1} = x_{i2} = \cdots = x_{im}$, A4.7, $E(z_{im} \cdot \varepsilon_{ih}) = 0$

$\hat A_{mh} = \frac{1}{n} \sum z_{im} z_{ih}'$ $\quad \hat c_{mh} = \frac{1}{n} \sum z_{im} y_{ih}$ $\quad A_{mh} = E(z_{im} z_{ih}')$

☆Sketch Proof of A_{mh}. Let $D(k \times L_m)$ be first L_m columns of I_k. $z_{im} = D' x_i$.

③ SUR and OLS

(a) When equations are all just identified, $z_{im} = x_i$ $\forall m$, then SUR is simply equation-by-equation OLS.

(b) SUR is more efficient than equation-by-equation OLS.

4.6 Common Coefficients

(1) Settings

A 4.1' $\quad y_{im} = z_{im}' \underset{L\times 1}{\delta} + \varepsilon_{im}$ $\quad \delta$ is common to all equations

When ~~SUR condition~~ homo + Same IV across equations:

$$\hat{S} = \hat{\Sigma} \otimes \left(\frac{1}{n}\sum x_i x_i'\right).$$

$$g_i(w_i;\tilde{\delta}) = \begin{bmatrix} x_{i1}\cdot(y_{i1} - z_{i1}'\tilde{\delta}) \\ \vdots \\ x_{im}\cdot(y_{im} - z_{im}'\tilde{\delta}) \end{bmatrix} \qquad E(g_i) = \begin{bmatrix} E(x_{i1}\cdot y_{i1}) \\ \vdots \\ E(x_{im}\cdot y_{im}) \end{bmatrix} - \begin{bmatrix} E(x_{i1}\cdot z_{i1}')\tilde{\delta} \\ \vdots \\ E(x_{im}\cdot z_{im}')\tilde{\delta} \end{bmatrix} = \sigma_{xy} - \Sigma_{xz}\tilde{\delta}$$

$$\underset{\Sigma K_m \times 1}{\sigma_{xy}} = \begin{bmatrix} E(x_{i1}\cdot y_{i1}) \\ \vdots \\ E(x_{im}\cdot y_{im}) \end{bmatrix} \qquad \Sigma_{xz} = \begin{bmatrix} E(x_{i1} z_{i1}') \\ \vdots \\ E(x_{im} z_{im}') \end{bmatrix}$$

Now Σ_{xz} is stacked.

<u>A 4.4'</u> Σ_{xz} is of full column rank.

(2) GMM Estimator

$$S_{xy} \equiv \begin{bmatrix} \frac{1}{n}\sum x_{i1} y_{i1} \\ \vdots \\ \frac{1}{n}\sum x_{im} y_{im} \end{bmatrix} \qquad S_{xz} \equiv \begin{bmatrix} \frac{1}{n}\sum x_{i1} z_{i1}' \\ \vdots \\ \frac{1}{n}\sum x_{im} z_{im}' \end{bmatrix}$$

$$\hat{\delta}(\hat{W}) = \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} z_{im} x_{im}'\right)\hat{W}_{mh}\left(\frac{1}{n}\sum_{i=1}^{n} x_{ih} z_{ih}'\right)\right\}\right]^{-1}\left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} z_{im} x_{im}'\right)\hat{W}_{mh}\left(\frac{1}{n}\sum_{i=1}^{n} x_{ih}\cdot y_{ih}\right)\right\}\right]$$

(3) Homoskedasticity and RE Estimator

Under homoskedasticity we have. $\hat{S} = \hat{\Sigma}\otimes\left(\frac{1}{n}\sum x_i x_i'\right)$

$$\therefore \hat{\delta}(\hat{S}^{-1}) = \left[\sum\sum\left\{\hat{\sigma}^{mh}\cdot\left(\frac{1}{n}\sum z_{im}x_i'\right)\left(\frac{1}{n}\sum x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum x_i z_{ih}'\right)\right\}\right]^{-1}$$
$$\cdot\left[\sum\sum\left\{\hat{\sigma}^{mh}\cdot\left(\frac{1}{n}\sum z_{im}x_i'\right)\left(\frac{1}{n}\sum x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum x_i y_{ih}\right)\right\}\right]$$

This is 3SLS with common coefficients.

(4) RE Estimator

Assume SUR condition and homo. Also the set of instruments is same across equations.

$$\hat{\delta}_{RE} = \left[\sum\sum\hat{\sigma}^{mh}\left(\frac{1}{n}\sum z_{im}z_{ih}'\right)\right]^{-1}\sum\sum\hat{\sigma}^{mh}\left(\frac{1}{n}\sum z_{im}\cdot y_{ih}\right)$$

$$\text{Avar}(\hat{\delta}_{RE}) = \left[\sum\sum\sigma^{mh}E(z_{im}z_{ih}')\right]^{-1} \qquad \hat{S} = \hat{\Sigma}\otimes\left(\frac{1}{n}\sum x_i x_i'\right)$$

(5) Pooled OLS

Using $\hat{W} = I_m\otimes\left(\frac{1}{n}\sum x_i x_i'\right)^{-1}$.

$$\hat{\delta}_{POLS} = \left(\sum_{i=1}^{n}\sum_{m=1}^{M} z_{im}z_{im}'\right)^{-1}\sum_{i=1}^{n}\sum_{m=1}^{M} z_{im}\cdot y_{im}$$

moment conditions exploited here: $E(z_{i1}\cdot\varepsilon_{i1} + \cdots + z_{im}\cdot\varepsilon_{im}) = 0$.

(6) Beautified formulas

$$\underset{M\times 1}{y_i} = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{im} \end{bmatrix} \qquad \underset{M\times L}{Z_i} = \begin{bmatrix} z_{i1}' \\ \vdots \\ z_{im}' \end{bmatrix} \qquad \underset{M\times 1}{\varepsilon_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{im} \end{bmatrix}$$

$$\Rightarrow \hat{\delta}_{RE} = \left(\frac{1}{n}\sum Z_i'\hat{\Sigma}^{-1}Z_i\right)^{-1}\frac{1}{n}\sum Z_i'\hat{\Sigma}^{-1}y_i \qquad [WLS]$$

$$\text{Avar}(\hat{\delta}_{RE}) = \left(E(Z_i'\Sigma^{-1}Z_i)\right)^{-1}$$

$$\hat{\delta}_{POLS} = \left(\frac{1}{n}\sum Z_i'Z_i\right)^{-1}\frac{1}{n}\sum Z_i'y_i \qquad \text{Avar}(\hat{\delta}_{POLS}) = [E(Z_i'Z_i)]^{-1}E(Z_i'\Sigma Z_i)[E(Z_i'Z_i)]^{-1}$$

☆ Conclusion different estimators under GMM Framework

Multiple-Equation GMM

- $\xrightarrow{\text{W diagonal}}$ Equation-by-Equation Estimator
- $\xrightarrow{\hat{W}=\hat{S}^{-1}}$ Efficient GMM
- $\xrightarrow[\text{Efficient, Same IV}]{\text{Homo}}$ 3SLS $\xrightarrow[\text{Same IV},\ x_i = \text{union of } z]{E(z_{im}\cdot\varepsilon_{ih})=0}$ SUR $\xrightarrow[\text{Same IV},\ \hat{W}=\hat{S}^{-1},\ \text{Same set of regressors}]{\text{Common Coef}}$ RE (WLS) $\xrightarrow{\hat{W}=I_m\otimes(\frac{1}{n}\sum x_i x_i')^{-1}}$ Pooled OLS

Q性质: $Q \cdot Q = Q$, sym, $Q 1_m = 0$, $Q\tilde{F}_i = \tilde{F}_i$, $Q d_i = 0$

Chapter 5 Panel Data

Assume i.i.d. A longitudinal or panel data set has multiple observations for a number of cross-section units.

5.1 The error-components model

(1) Assumptions

$y_i (M\times 1)$, $Z_i (M\times L)$, $\varepsilon_i (M\times 1)$

Same IV ① $y_i = Z_i\delta + \varepsilon_i$ ② $\{y_i, Z_i\}$ i.i.d. ③ SUR: $E(z_{im}\cdot\varepsilon_{ih}) = 0$ i.e. $E(\varepsilon_i \otimes x_i) = 0$

x_i = union of $(z_{i1}, \cdots, z_{iM})$ ④ $E(Z_i \otimes x_i)$ full column rank ⑤ $E(\varepsilon_i\varepsilon_i' | x_i) = E(\varepsilon_i\varepsilon_i') = \Sigma$

⑥ $E(g_i g_i')$ nonsingular.

(2) Error Components

① $\varepsilon_{im} = \alpha_i + \eta_{im}$ α_i: individual effect [Fixed effect] Common to all equations

Define $\underset{(M\times 1)}{\eta_i} = \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{iM} \end{bmatrix} \Rightarrow \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix} = \begin{bmatrix} z_{i1}'\delta \\ \vdots \\ z_{iM}'\delta \end{bmatrix} + \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{iM} \end{bmatrix}$ $\underset{(M\times 1)}{y_i} = Z_i\delta + 1_M \cdot \alpha_i + \eta_i$

Orthogonality conditions: $E(z_{im}\alpha_i) = 0$ + $E(z_{im}\cdot\eta_{ih}) = 0$

(3) Group Means

① Define Annihilator matrix of 1_M: $Q \equiv I_M - 1_M(1_M'1_M)^{-1}1_M' = I_M - \frac{1}{M}1_M 1_M' = I_M - \begin{bmatrix} \frac{1}{M} & \cdots & \frac{1}{M} \\ \vdots & & \vdots \\ \frac{1}{M} & \cdots & \frac{1}{M} \end{bmatrix}$ → extract deviations from group means

$\tilde{y}_i \equiv Q y_i = \begin{bmatrix} y_{i1} - \bar{y} \\ \vdots \\ y_{iM} - \bar{y} \end{bmatrix} = y_i - 1_M \cdot \bar{y}_i$ where $\bar{y}_i = \frac{1}{M}1_M' y_i = \frac{1}{M}\sum_{m=1}^{M} y_{im}$ group mean

(4) Reparameterization

To separate common regressors: $Z_i = (F_i \;\vdots\; 1_M b_i')$, $\delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$

One of the intercepts needs to be dropped from F_i and included in b_i.

☆ [FE] Identification condition: $E(QF_i \otimes x_i)$ is of full column rank.

$\therefore y_i = \underset{M\times L_1,\, L_1\times 1}{F_i\beta} + \underset{M\times 1}{1_M} \cdot \underset{1\times L_2,\, L_2\times 1}{b_i'\gamma} + 1_M\cdot\alpha_i + \eta_i$

$F_i = \begin{bmatrix} f_{i1}' \\ \vdots \\ f_{iM}' \end{bmatrix}$ $y_{im} = f_{im}'\beta + b_i'\gamma + \alpha_i + \eta_{im}$ f_{im}' is the m-th row of F_i.

5.2 Fixed-Effect Estimator

(1) Formula

$Q y_i = QF_i\beta + \underbrace{Q1_M b_i'\gamma}_{=0} + \underbrace{Q 1_M \alpha_i}_{=0} + Q\eta_i$ $[Q1_M = 0]$ Q is symmetric and $Q\cdot Q = Q$, $QF_i = \tilde{F}_i$

$\Rightarrow Q y_i = QF_i\beta + Q\eta_i$ $\tilde{y}_i = \tilde{F}_i\beta + \tilde{\eta}_i$ $\underset{(Mn\times 1)}{\tilde{y}} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_n \end{bmatrix}$ $\underset{(Mn\times L_1)}{\tilde{F}} \equiv \begin{bmatrix} \tilde{F}_1 \\ \vdots \\ \tilde{F}_n \end{bmatrix}$

$\hat{\beta}_{FE} = (\tilde{F}'\tilde{F})^{-1}(\tilde{F}'\tilde{y}) = \left(\frac{1}{n}\sum \tilde{F}_i'\tilde{F}_i\right)^{-1}\frac{1}{n}\sum\tilde{F}_i'\tilde{y}_i = \left(\frac{1}{n}\sum F_i' Q F_i\right)^{-1}\frac{1}{n}\sum F_i' Q y_i$

(2) Large-Sample Properties

☆ $\sqrt{n}(\hat\beta_{FE}-\beta) \xrightarrow{d} N(0, Avar(\hat\beta_{FE}))$ $\quad Avar(\hat\beta_{FE}) = [E(\tilde F_i'\tilde F_i)]^{-1} E[\tilde F_i' E(\tilde\eta_i\tilde\eta_i')\tilde F_i][E(\tilde F_i'\tilde F_i)]^{-1}$

(☆☆) Proof: ① $E(F_i' Q \eta_i) = E\left(\sum_m^M \sum_h^M q_{mh} \cdot f_{im}\cdot\eta_{ih}\right)$ 其 $q_{mh} \equiv (m,h)$ element of Q

$= \sum_m^M\sum_h^M q_{mh} E(f_{im}\cdot\eta_{ih}) = 0$

② $E(\tilde F_i'\tilde\eta_i\tilde\eta_i'\tilde F_i) = E[\tilde F_i' E(\tilde\eta_i\tilde\eta_i')\tilde F_i]$

(3) When η_i is spherical

Assumption: $E(\eta_i\eta_i') = \sigma_\eta^2 I_M \Rightarrow E(\tilde\eta_i\tilde\eta_i') = \sigma_\eta^2 Q$

Then $Avar(\hat\beta_{FE}) = \sigma_\eta^2\cdot[E(\tilde F_i'\tilde F_i)]^{-1}$

(4) FE vs. RE

① Comparison

FE doesn't use: $E(f_{im}\cdot\alpha_i)=0\ \forall m$, $E(b_i\cdot\alpha_i)=0$, $E(b_i\cdot\eta_{im})=0\ \forall m$.

Define $\hat\delta_{RE} \equiv \begin{bmatrix}\hat\beta_{RE}\\ \hat\gamma_{RE}\end{bmatrix}$. $\hat\beta_{RE}$ is efficient but $\hat\beta_{FE}$ is not.

② Hausman Specification Test

Let $\hat q \equiv \hat\beta_{FE}-\hat\beta_{RE}$. $\sqrt{n}\,\hat q$ is asymptotically normal and $Avar(\hat q) = Avar(\hat\beta_{FE}) - Avar(\hat\beta_{RE})$

∴ We have test:

$$H \equiv n\cdot\hat q'(\widehat{Avar(\hat q)})^{-1}\hat q \xrightarrow{d} \chi^2_{\#\beta}$$

For rows with missing values, we replace them with all zeroes

Unbalanced Panel

(1) Zeroing Out Missing Observations

Define $d_{im} = \begin{cases}1 & \text{if } m \text{ is in the sample}\\ 0 & \text{if not}\end{cases}$ $\quad \underset{(M\times1)}{d_i} = \begin{bmatrix}d_{i1}\\ \vdots\\ d_{iM}\end{bmatrix}$, $M_i = \sum_{m=1}^M d_{im} = \#$ obs from i

for simplicity

Tips: 无法观测的那行m，全取零

$\underset{(M\times1)}{y_i} = \begin{bmatrix}d_{i1}\cdot y_{i1}\\ \vdots\\ d_{iM}\cdot y_{iM}\end{bmatrix}$ $\quad \underset{(M\times\#\beta)}{F_i} = \begin{bmatrix}d_{i1}\cdot f_{i1}'\\ \vdots\\ d_{iM}\cdot f_{iM}'\end{bmatrix}$ $\quad \underset{(M\times1)}{\eta_i} = \begin{bmatrix}d_{i1}\cdot\eta_{i1}\\ \vdots\\ d_{iM}\cdot\eta_{iM}\end{bmatrix}$

[Just consider for each i and m, either all the elements of (y_{im}, f_{im}) are observable or none is observable] ↳ 简化

☆☆ Deviation of y_{im} from group mean over available obs.

∴ $y_i = F_i\beta + d_i\cdot b_i'\gamma + d_i\cdot\alpha_i + \eta_i$ In balanced case, $d_i = 1_M$.

Redefine $Q = I_M - d_i(d_i'd_i)^{-1}d_i' = I_M - \frac{1}{M_i}d_id_i'$ $\quad (M_i = d_i'd_i)$ $\quad Qd_i=0$

$\tilde y_i \equiv Q y_i = \begin{bmatrix}d_{i1}\cdot y_{i1} - d_{i1}\cdot\bar y_i\\ \vdots\\ d_{iM}\cdot y_{iM} - d_{iM}\cdot\bar y_i\end{bmatrix}$ where $\bar y_i \equiv \frac{1}{M_i}d_i'y_i = \frac{1}{M_i}\times$ (Sum of y_{im} for observable m)

Tips: 对不同i, Q(i)不同!!!

for different i, Q(i) is different

(2) Asymptotics

Assumption: $E(f_{im}\cdot\eta_{ih}|d_i)=0$ No selectivity bias.

Now: $F_i'Q\eta_i = \sum_m^M\sum_h^M q_{mh}^{(i)}\cdot d_{im}\cdot d_{ih}\cdot f_{im}\cdot\eta_{ih}$

$E(F_i'Q\eta_i) = \sum\sum E(q^{(i)}_{mh}\, d_{im}\, d_{ih}\, f_{im}\,\eta_{ih}) = \sum\sum E[q^{(i)}_{mh}\cdot d_{im}\cdot d_{ih}\cdot E(f_{im}\eta_{ih}|d_i)] = 0$